A Gestural Approach to Presentation exploiting Motion Capture Metaphors

Stefania Cuccurullo
Rita Francese
Sharefa Murad
Ignazio Passero
Maurizio Tucci

Università degli Studi di Salerno,
Via Ponte don Melillo 1, Fisciano (SA), Italy
{scuccuru, francese, smurad, ipassero, mtucci}@unisa.it

ABSTRACT
Speaking in public may be a challenging task in terms of self-control and attention to the concepts to expose and to non-verbal communication. Presentation software, like Microsoft PowerPoint™ or OpenOffice, may support the speaker in organizing and controlling the flow of his/her discussion by commanding the slide change. In this paper we describe an approach exploiting the availability of the Microsoft Kinect™ advanced game controller to manage a presentation software through a Natural User Interface (NUI). The approach, named Kinect Presenter (KiP), adopts motion capture to recognize body gestures representing interaction metaphors. We perform a preliminary evaluation aiming at assessing the degree of support provided by the proposed interaction approach to the speaker activities. The assessment is based on the combined usage of two techniques: a questionnaire-based survey and an empirical analysis. The context of this study was constituted of Bachelor and PhD students in Computer Science at the University of Salerno, and teachers and employees from the same university. First results were adequate both in terms of satisfaction and performances, also when compared with a wireless mouse-based interaction approach.

Categories and Subject Descriptors
H.5.2 [Information Interfaces and Presentation]: User Interfaces; B.4.2 [Input/Output and Data Communication]: Input/Output Devices

General Terms

Keywords
Gesture-Based presentation; Kinect; Gesture-recognition;

1. INTRODUCTION
Speaking with the support of a presentation software is one of the most diffused practice in public communication. It is still one of the most widely adopted methods in teaching, conferencing, and, above all, in organizational context, where is largely used for communicating project results to team members or upper management, and during job meeting in general. PowerPoint™, provided by Microsoft as a toolkit of office, together with its open version OpenOffice, is the most popular tool to make a presentation. However, traditional keyboard and mouse based presentations prevent speakers from freely and closely interacting with the audiences, because the speaker continuously has to go to the computer to manage the Presentation. Nowadays, most of this problem has been overcome by the presence of wireless control that provides a good mobility. According to [28], also these devices have some drawbacks. Indeed, they offer a small touchpad difficult to use for controlling the mouse when the presenter is walking around, and does not allow the speaker to use his/her gestures to control the presentation. In addition, multiple interactions are not allowed: only one of these device can be used on the same computer.

Gesture-based interfaces can improve human-computer communication allowing users more natural and intuitive interaction modalities. In the last years, several research [1][6][12][18][19][21][28] and industrial works [3][9][16][17] have been devoted in designing efficient, robust, and inexpensive solutions to recognize hand or body gestures. Recently, the availability of greater processing power, wider memory, cameras, and sensors make possible to introduce this interaction modality in commonly used software, like a presentation software.

In this paper we describe an interaction approach that adopts one of the latest available advanced gaming controllers, Microsoft Kinect™, to allow a speaker to exploit natural human movements and gestures to control a presentation software. We present the results of a preliminary evaluation of the presentation system, named Kinect Presenter (KiP), and of the associated metaphors. The assessment was conducted considering both a questionnaire-based survey and an empirical analysis, aiming at evaluating the speaker satisfaction and performances when compared with a traditional way of conducting presentations, such as when using a wireless control.

The paper is organized as follows: Section 1 describes the state of the art related to gesture-based technologies and their use to support presentations. Section 3 presents the proposed system. Section 4 details the controlled experiment performed to evaluate the proposed approach, while Section 5 analyzes the assessment results. Finally, Section 6 concludes.

2. BACKGROUND
This section reports on the state of art concerning gesture-based interfaces and their relationship with presentation software control.
and describes the main characteristics of Kinect, the input device adopted in the proposed approach.

2.1 Gesture-based input devices

For the last four decades keyboard and mouse have been the primary means to interact with computers. Starting from 2006 with Nintendo Wii [17] and, successively, with Apple iPhone [9] in 2007, the consumer interest towards interfaces based on natural interaction modalities (speech, touch, gesture) is rapidly increasing. The term Natural User Interfaces (NUI) includes interaction modalities that "enable users to interact with computers in the way we interact with the world." [11]. As new devices that take advantage of easy and intuitive NUI appear on the market, users experiment unprecedented levels of control on the devices around them. Cameras and sensors pick up the movements of their bodies without the need of remotes or handheld tracking tools.

In medical systems and assistive technologies, gestures can be used to control the distribution of resources in hospitals, to interact with medical instrumentation, to control visualization displays, and to help handicapped users as part of their rehabilitation therapy. See [18][23] as an example. Some of these concepts have been exploited to improve medical procedures and systems; for example, Face MOUSE [18] satisfies the "come as you are" requirement, where surgeons control the motion of a laparoscope by making appropriate facial gestures without hand or foot switches or voice input. In [5] Gallo et al. use Microsoft XBox Kinect as an input device to develop a controller-free, highly interactive exploration of medical images. The system interface allows users to interact at a distance through hand and arm gestures. The description of other medical gesture-based applications is provided in [24].

The approaches to gesture-based input vary. The screens for iPhone, iPad, Android-based devices and the multi-touch Surface by Microsoft all react to pressure, motion, and the number of fingers used in touching the devices. Some devices react to shaking, rotating, tilting, or moving the device in space. See [22] as an example. The Wii controller along with similar gaming systems, works by combining a handheld accelerometer-based controller with a stationary infrared sensor to determine position, acceleration, and direction. Development in this area points on creating a minimal interface and on producing an experience of direct interaction such that, cognitively, hand and body become themselves input devices. The Sony PlayStation 3 Motion Controller also moves in this direction. Microsoft Kinect does not require the user to wear or hold anything while detects his/her motions.

Today, the technologies for gesture-based input continue to expand. As an example, EvoTox [3] is a unique hardware and software package that combines controller free gesture interaction, based on Kinect, with precise touch screen technology. It allows people to interact with Windows 7 through the Kinect system. In their system [26], Williamson et al. make use of video game console motion controllers including Microsoft Kinect, Playstation Move, and Nintendo Wiimote, combined with the Unity 3D game engine to support untethered interaction. The system makes use of a set of heuristic rules that recognize various actions taken from the Kinect's depth image 3D skeleton representation. These rules support seamless transitions between realistic physical interactions (e.g., actually walking and running) and proxied physical interactions (e.g. walking and running in place) that support locomotion in the larger Virtual Environment.

2.2 Gesture-based presentations

The idea to enable the user to perform public presentations in a more natural way is not new. As an example, Reifinger et al. in [21] proposed a gesture recognition system that is able to recognize static gestures, like pointing or grasping, as well as dynamic gestures, like drawing letters in the air. Based on a master-client structure, the gesture caption and recognition module receives tracking data from an infrared tracking system developed to support Augmented Reality applications. However, the system is not a controller-free interface as the user has to wears two light weighted infrared tracking targets at his thumb and index fingers.

SlideShow [2] is a gesture based intelligent user interface based on a remote stick, equipped by some inertial sensors designed specifically for lecturers. Operations are segmented from the movement sequence and divided into several automatically switched states. A Bayesian-based algorithm is addressed to segment the continuous gestures.

In [28] Yang and Li proposed to use Wiimote as a wireless mouse. They proposed to adopt this interaction approach in classrooms and conference rooms for presentations and interactive discussions. The main advantage is to allows multiple users to use multiple Wiimotes to operate on the same computer. The same result is reached by our approach, with the addition that no device has to be handheld by the speaker.

Alexander et al. presented Gestur [1], an open-source software framework developed in C# and mainly focused on real-time hand-gesture recognition. The framework produces an application that controls Microsoft PowerPoint presentations, whereby users indicate to the computer the direction to advance slides, terminate a presentation, or any other action initially configured. They do not performed any evaluation in their paper, neither the adopted metaphors have been detailed. Our approach differs from their in the adopted technologies and metaphors. In addition, our system supports multiple user interaction.

In their paper Zarraonandia et al. [29] tried to foresee some possible new technologies enriching the future of IT lecture scenarios, such as the adoption of gesture-based interaction. As in that work, also Fourney et al. studied the effects and implications of this kind of interaction on dynamic presentations [4].

2.3 Microsoft Kinect

In the proposed approach we adopted the Microsoft Kinect [14] motion sensing input device. It enables the user to naturally interact with software programs without the need of physically touching any object. This device supports facial and voice recognition, automatic player sign-in, 3D scene approximation and reconstruction, full-body motion capture, and tracking of four players simultaneously with 48 skeleton positions per player at 30 Hz [14][25].

The Kinect device, shown in Figure 1, is packed with state-of-the-art proprietary technologies. Its main hardware features are: a pair of depth-sensing range cameras, a system of infrared structured light sources, a multi-array microphone, and a regular RGB camera. The depth-sensing cameras can approximate distances of objects by continuously projecting and interpreting reflected results from the structured infrared light. The multi-array microphone assists in acoustic source localization and ambient noise suppression and provides support for voice recognition and headset-free live chats [13].
Recently, range sensors have been largely adopted to capture human motions thanks to the support they offer to a non-invasive system setup. In particular, Time-of-flight (TOF) sensors provide, at high frame rates, dense depth measurements in each point in the scene. TOF cameras capture an ordinary RGB image and create a distance map of the scene using the light and ranging (LIDAR) detection schema: modulated light is emitted by LEDs or lasers and the depth is estimated by measuring the delay between emitted and reflected light. LIDAR makes the TOF cameras insensitive to shadows and changes in lighting, allowing a disambiguation of poses with a similar appearance. More recently, a less expensive solution to obtain 3D information from video, with respect to the one implemented in the TOF cameras, has emerged. This solution projects structured IR light patterns on the scene and retrieves depth information from how the structured light interferes with the objects in the scene. This is the mechanism used in the Microsoft Xbox Kinect™.

Starting from this information, Kinect creates a depth map in real time, where each pixel represents an estimation of the distance between the Kinect sensor and the nearest object in the scene corresponding to the pixel location. Based on this map, the Kinect system software supports applications such as Kip in the accurate and efficient tracking of the skeleton of a human body in three dimensions.

3. THE KIP APPROACH

The Kinect Presenter (KiP) system is a PowerPoint controller that adopts motion capture for creating a gestural interface for managing public presentations. In particular, by using Microsoft Kinect™ as the only input device, the system user interface allows a speaker to interact at a distance through hand and arm gestures. The system provides gesture commands to run the presentation, go to the next slide and go backwards as minimal actions during the talk.

KiP is a C# application that is connected to the Kinect device via the official Kinect SDK beta provided by Microsoft [13]. The system tracks the body skeleton generated by Kinect and maps it into the configured set of gestures representing the various commands that can act on the running presentation. When a user position is similar to a predefined one, the corresponding command is enacted.

Let us observe that the KiP allows to multiple users to operate on the same presentation and interactively discuss during a meeting.

The next subsections summarize the usability goal we identified and describe the proposed interaction metaphors.

3.1 Usability requirements of the KiP interface

The requirements of a gesture interface vary depending on the application type [24]. For example, an entertainment system does not need the same accuracy in gesture-recognition than a surgical system.

Speakers often use hand gestures when talking. The use of gesture-recognition methodology in presenting or speaking imposes the right selection of the gestures among as set of predefined ones. This is related to recognition, a component of accuracy together with detection and tracking, which indicates that command gestures should not confused with other movements. Another main factor is intuitiveness. The gestures selected to command the presenter interface should have a clear relationship with the functionalities they execute. This is correlated to the need of having a reduced mental load. The user should naturally drive the interaction and easily remember which movement he has to perform. A heavy mental load due to the need of thinking to the gesture to perform risks to distract him/her from the discussion he/she is conducting. Also Comfort should be taken into account. Indeed, gestures should not require particular effort. The gesture recognition should be performed in real-time (Responsiveness). If it does not happen the interaction is impracticable.

3.2 The KiP interface

The interaction metaphors adopted to control the slide presenter has to be natural and intuitive, but at the same time they have to prevent conflicts with the user movements. In addition, the gestures should be meaningful when managing a presentation. To this aim we consider the interaction modality offered by PowerPoint to manage a presentation and try to translate these commands in simple gestures. The command gestures proposed in this approach are depicted on Figure 2.
To start a PowerPoint presentation in traditional modality the user clicks on the presentation button in the lower right part of the application screen. As shown in Figure 2(a), to start the presentation in KiP modality the speaker raises both the hands, forming a right angle with each arm. To go to the next slide, the PowerPoint user presses the right arrow. KiP does the same action when the user raises the right hand, forming a right angle also in this case, as shown in Figure 2 (b). Similarly, to go to the previous slide the PowerPoint user presses the left arrow and the KiP user raises the left hand (Figure 2 (c)). Let us note that a natural speaking movement like the one shown in Figure 2 (d) is not recognized by the system as a command.

4. EVALUATION

In this section we describe the data set and then the techniques we have adopted to evaluate the system. These techniques are based on both a questionnaire-based survey and an empirical analysis [27], aiming at assessing the tool usability when KiP performances and user satisfaction are compared to the usage of a wireless remote control presentation, named WiP in the rest of the paper.

4.1 The data set

The study was conducted in a research laboratory at the University of Salerno. Data for the study have been gathered considering a group of eighteen volunteers. Seven of them were Bachelor students in Computer Science at the University of Salerno, five were PhD students, four were teachers and the remaining were employees of the same university. Before performing the experiment, the subjects were asked to answer to a pre-experiment questionnaire evaluating the users skills that can influence the evaluation, aggregating three factors: PPT, the Power-Point experience, GB-Devices, highlighting the previous experience in gesture-base gaming controllers, and, finally, PA, the Presenting Attitude of the users, as the overall general experience in public speaking. The answers to the survey questionnaire have been evaluated on the seven-point Likert scale [20]: from 1 (very low) to 7 (very high).

The results in Figure 3 report a general high experience in both the use of Power-Point and in personal attitude to present in general. There is one outlier, one of the two employees that has a very low previous PPT experience. The gesture-base devices experience remains low, but the existence of two outliers revealed that two students are practice in the selected devices and technologies.

4.2 Experiment Design

In order to properly design the experiment and analyze the results, the following independent variable needs to be considered:

- **Method**: this variable indicates the factor on which the study is focused, i.e. KiP and WiP.

The considered dependent variable are:

- **Time**: the time required to perform the task.
- **Mistakes**: the mistakes made by the subjects while performing the task, such as go backwards one more time.

4.3 Material and execution

The study has been performed in one-to-one session (i.e. a supervisor for each subject). First, all the subjects have been introduced the KiP tool and its main functionalities. Similarly, they were introduced to the Wireless Presenter (WiP) functionalities too. The adopted remote control exposes several buttons, including next slide button and previous slide button and a small-sized rectangle touchpad for moving the mouse cursor. Successively, they have been asked to use each of the tools for 5 minutes, without invoking any kind of tutor support and on a presentation different from the ones adopted in Task 1 and 2. The subjects were asked to perform two tasks. Both the tasks required the participants to give a short presentation of 9 slides on different simple topics, as described in the previous section. At the end of
each task, the subjects filled in a post task survey questionnaire to achieve information on their satisfaction. During the experiment, the supervisor did not provide any help to the subjects to avoid biasing the experiment. He only wrote the comments and problems of the subjects. For each subject the needed time to accomplish the experiment was annotated as well.

Because the comparison was performed between a gesture-based interface and a traditional one, the survey questionnaire adopted in this evaluation is a standard usability questionnaire presented in [15] and named USE. It proposes thirty questions that are grouped to evaluate a software product considering four dimensions: Usefulness, Ease of Learning, Ease of Use and Satisfaction.

The answers to the questions of the survey questionnaires have been scored on the seven-point Likert scale: from -3 (strongly disagree) to 3 (strongly agree).

5. RESULTS
In this section we report the results of the proposed study, examining, in particular, the subjective evaluation related to the survey questionnaire and the objective empirical evaluation.

5.1 Survey results
The subjective evaluation statistics of the experiment are given in Table 2, where for each method, task and usability factor the minimum and the maximum value are shown, together with the median, the mean and the standard deviation. We also report the questionnaire results using the BoxPlot diagrams in Figure 4 and 5. These figures show the subjective results collected after using the WiP and KiP methods, respectively. In this way, it is possible to highlight the dispersion and the skewness of the sample.

Table 2. Survey statistics

<table>
<thead>
<tr>
<th>Task</th>
<th>Met.</th>
<th>Factor</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>KiP</td>
<td>Usefulness</td>
<td>0</td>
<td>3</td>
<td>1.49</td>
<td>1.68</td>
<td>0.64</td>
</tr>
<tr>
<td>2</td>
<td>WiP</td>
<td>Usefulness</td>
<td>-1</td>
<td>3</td>
<td>1.40</td>
<td>1.31</td>
<td>0.51</td>
</tr>
<tr>
<td>1</td>
<td>KiP</td>
<td>Ease of Use</td>
<td>-1</td>
<td>3</td>
<td>1.48</td>
<td>1.40</td>
<td>0.68</td>
</tr>
<tr>
<td>2</td>
<td>WiP</td>
<td>Ease of Use</td>
<td>-3</td>
<td>3</td>
<td>1.52</td>
<td>1.68</td>
<td>0.57</td>
</tr>
<tr>
<td>1</td>
<td>KiP</td>
<td>Ease of Learning</td>
<td>0</td>
<td>3</td>
<td>1.97</td>
<td>2</td>
<td>0.70</td>
</tr>
<tr>
<td>2</td>
<td>WiP</td>
<td>Ease of Learning</td>
<td>1</td>
<td>3</td>
<td>2.16</td>
<td>2</td>
<td>0.38</td>
</tr>
<tr>
<td>1</td>
<td>KiP</td>
<td>Satisfaction</td>
<td>-1</td>
<td>3</td>
<td>1.63</td>
<td>1.85</td>
<td>0.63</td>
</tr>
<tr>
<td>2</td>
<td>WiP</td>
<td>Satisfaction</td>
<td>-1</td>
<td>3</td>
<td>1.13</td>
<td>1</td>
<td>0.80</td>
</tr>
</tbody>
</table>

As Table 2 depicts, the subjects found KiP a bit more useful with $\mu_\text{KiP}=1.49$, versus $\mu_\text{WiP}=1.40$ in case of WiP. The mean of the participants found the easiness of the two tools very similar ($\mu_\text{KiP}=1.48$ and $\mu_\text{WiP}=1.52$) and this is a very positive results obtained by a new interface, never used before by the subjects, when compared with a well known interaction modality, such as a mouse. Concerning the Learnability, this is a very critical aspect for the success of a gesture-based, interface because it denotes that the proposed gesture patterns used to control applications are easy to perform and remember [24]. Thus, considering the novelty of the interface and the previous knowledge of PPT usage, the results of KiP ($\mu_\text{KiP}=1.97$ and $\mu_\text{WiP}=2.16$) denotes that the gesture are enough intuitive and natural. The opinion on the Easy of learning is not homogeneous, as Figure 5 revealed. In addition, we had two outliers that scored 3 for WiP. Examining the subjects' profile, we discovered that one of them was a teacher very expert in PowerPoint usage, while the other was a Bachelor student.

From Figure 4 and 5, it is clear that the KiP reaches a higher consensus for the overall satisfaction dimension ($\mu_\text{KiP}=1.63$ and $\mu_\text{WiP}=1.13$), except for one outlier.
Another interesting aspect that makes a tool attractive is the user perception of fun. Indeed, according to Igbaria et al. [8], the perceived fun had a stronger effect on user satisfaction than perceived usefulness. For this reason, we examine in detail the opinions concerning Q26, "It is fun to use." Let us observe that, on this aspect, the participants largely prefer KiP ($\mu_{KiP} = 2$ and $\sigma_{KiP} = 1.19$, $\mu_{WiP} = 0.72$ and $\sigma_{WiP} = 0.96$). When examining the detailed opinions concerning Q27, "It works the way I want it to work.", the users appreciate KiP functionalities as the ones of WiP ($\mu_{KiP} = 1.67$ and $\sigma_{KiP} = 0.97$, $\mu_{WiP} = 1.61$ and $\sigma_{WiP} = 0.77$), but the overall satisfaction, question Q24, "I am satisfied with it", shows positive results for KiP ($\mu_{KiP} = 2$ and $\sigma_{KiP} = 0.69$, $\mu_{WiP} = 1.5$ and $\sigma_{WiP} = 0.92$).

Concerning Q30, "It is pleasant to use", the majority of the subjects found the KiP interface pleasant, also with respect to WiP ($\mu_{KiP} = 1.67$ and $\sigma_{KiP} = 0.84$, $\mu_{WiP} = 1.27$ and $\sigma_{WiP} = 0.90$).

In Figure 6 we detailed the satisfaction factor results grouped by subject category (Students, PhD Students, Teacher, Employee). Let us note that Bachelor and PhD students are very satisfied. Quite good opinion are also reached by the teachers, but this opinion is not uniform. Quite neutral are the employee.

5.2 Empirical analysis results
The empirical analysis evaluates the performance of the two tools in terms of the accomplishment time of the tasks and the numbers of mistakes. The measures related to the time for WiP and KiP are shown in Table 3, respectively, while the related BoxPlot is depicted in Figure 7.

**Table 3 - WiP Time statistics (in sec.)**

<table>
<thead>
<tr>
<th>Task</th>
<th>Method</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>KiP</td>
<td>55</td>
<td>94</td>
<td>71.16</td>
<td>70</td>
<td>39</td>
</tr>
<tr>
<td>2</td>
<td>WiP</td>
<td>46</td>
<td>88</td>
<td>66.90</td>
<td>42</td>
<td>13.56</td>
</tr>
</tbody>
</table>

Considering the results shown in Table 3, it is evident that globally WiP performed better than KiP. Indeed, each participant using the KiP modality employed a mean of 14.23 sec. for slide, while this time is reduced to 13.38 sec. in case of WiP.

The experiment sample was variously composed. Thus, as in the survey case, we better investigated the performances in terms of the kind of users. As Table 4 revealed, PhD students perform a little better than Bachelor students, probably because they have a better attitude to speak in public and they are more practice in PowerPoint usage. However, the former get a good mean result ($\mu = 61.80$) and the Bachelor students ($\mu = 67.57$). This observation can also be deducted from the BoxPlot in Figure 8, where the time performances for user group are resumed. Let us note that the teacher performances are concentrated near the median. This means that they employed about the same time to accomplish the task.

**Table 4 - KiP time statistics for user groups**

<table>
<thead>
<tr>
<th>Users</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students</td>
<td>63</td>
<td>66</td>
<td>67.57</td>
<td>60</td>
<td>5.22</td>
</tr>
<tr>
<td>PhD students</td>
<td>55</td>
<td>77</td>
<td>61.80</td>
<td>64</td>
<td>8.75</td>
</tr>
<tr>
<td>Teachers</td>
<td>78</td>
<td>82</td>
<td>80</td>
<td>80</td>
<td>1.63</td>
</tr>
<tr>
<td>Employees</td>
<td>85</td>
<td>94</td>
<td>89.50</td>
<td>89.5</td>
<td>6.36</td>
</tr>
</tbody>
</table>

Concerning the user mistakes, they were very few: in the KiP case, two users made 1 mistakes, because they confused between right and left when they has to go next or back, while in the WiP method only a user made 1 mistake, pressing on time more the next button. He was an employee non expert in PowerPoint usage that also made an error in the KiP modality.

**Figure 7. The Time Performance results.**

It is important to point out that the supervisor noted that no natural movement has been erroneously considered by KiP as a user command. Indeed, one of the important aspect of this experimentation is the verification that the system does not make false positive, that is that it erroneously detects slide change commands (Recognition requirement). This is also confirmed by the results related to question Q12, "I don't notice any inconsistencies as I use it", that are very similar for the two approaches: $\mu_{KiP} = 2.28$ and $\sigma_{KiP} = 1.01$, $\mu_{WiP} = 2.5$ and $\sigma_{WiP} = 1.15$.

5.3 Discussion
The survey questionnaire results revealed that the general judgment of the KiP interaction modality is considered appropriate, also when compared with a wireless mouse-based interaction approach. It is important to point out that we compared an interaction modality (WiP), which most participants well known, to a new one (KiP), where specific body movement are required. As a result, the proposed experience had been capable of positively influencing the subject perception of this new interaction approach. The sample of the experiment was variously...
composed and this positive result is mainly true for the student subjects that are accustomed to game-like interaction. Also PhD students appreciate the new approach. This is interesting, because this category of subjects will become the speakers of the near future. The satisfaction degree of teachers is also good, but some of them had not too positive performances. Indeed, the controlled experiments shown that teachers performed better for the WiP method in accomplishment time. This is probably due to the practice that they had in using a traditional mouse. Concerning the simplicity of the system usage, the supervisor registered very few mistakes. Thus, there is no statistical difference on the number of mistakes made by the subjects to accomplish the task, when using the WiP and the KiP methods. This result is a clear indication of intuitiveness, that is that the gesture types have a clear cognitive association with the functions they perform [24]. A positive perception on Learnability denotes that the proposed gestures are enough natural and intuitive, and this is a key factor for the success of a gesture-based interface.

The need of recalling the gesture trajectories and associated actions can add mental load to a user. Also physical effort can reduce the system acceptance. Opinion concerning the effort was not particularly positive, especially for less young people. Even if the WiP modality also does not reach far more better results. A prolonged successive study should better verify this aspect.

In order to comprehend the strengths and limitations of this study threats that could affect its internal, construct and external validity need to be discussed. The internal validity threats are relevant for our study as we aimed at concluding that the proposed system effectively supports speakers during a presentation. There is the risk that subjects might have learned how to improve their performances in the second laboratory session. This validity threats is mitigated by the experiment design. Indeed, each group worked, over the two Labs, on different tasks and with two different methods with reversed order.

The construct validity threats could be present in this study. Let us note that the construct validity could be affected by the true and false positive defects that experts manually identified. Social threats (e.g. evaluation apprehension) could also affect the observed results. The subjects were volunteer and the students were not evaluated on the results they obtained. All the participants were not aware of the experiment aim. Finally, the survey questionnaire was standard.

External validity refers to the approximate truth of conclusions involving generalizations. This kind of threat is always present when students are used as subjects [7]. Generally, Bachelor Computer Science Students at the University of Salerno take presentations during several project works. In addition, other kinds of subjects participated to the study, more and less involved in the usage of this kind of software, and we examined in detail their opinions, performances and behavior. Moreover, none of the subjects abandoned the study. To confirm or contradict the achieved results, replications using a larger dataset will be conducted.

It is worth noting that conclusion validity threats are not present in this study, as statistical tests have not been performed to reject null hypotheses.

6. CONCLUSION

Gesture-based interfaces involve significant usability challenges, including fast response time, high recognition accuracy, learnability and user satisfaction. Probably, for these reasons few vision-based gesture systems have matured beyond prototypes to reach the commercial market. Nevertheless, there is strong evidence that gesture-based interactive applications will become important players in next-generation interface systems, due to their ease of access and naturalness of control. In this work we exploited specific interaction metaphors to control a presentation software. In particular, we adopt Microsoft Kinect as input device to control the presentation process. The proposed interaction metaphors are simple, temporally short, and natural: to control the presentation a speaker have to remember only four postures. The approach was evaluated both with subjective measurement and with a controlled experiment, measuring performance and satisfaction of the users. Results are encouraging in terms of satisfaction and simplicity. Indeed, the gestures selected by the interface developers resulted easy to perform and to remember, also when compared to a classical interaction modality, such as wireless remote controller. Positive results in terms of performances were mainly reached for the youngest component of the sample, Bachelor and PhD students.

We plan to complete the assessment phase by encouraging the AVI speakers to use the KiP system during their talks and collecting their opinions and performances. In addition, to increase the body of knowledge about the efficacy and effectiveness of the proposed approach we will also replicate the study in different contexts with subject with different background, such as in an industrial context, and will make observation for a longer time, i.e. during a whole university course, also to evaluate as the learning effect reduces the mental load and improves performances. We also aim at comparing the proposed approach with the features offered by Nintendo Wiimote, following the directions proposed in [28].

As future work, we tend to use the Kinect voice recognition capability to examine the response and effectiveness of both gesture and voice commands together.

7. REFERENCES


