Hands Free Presentations: Multimodal Interaction with PowerPoint

Dzianis Bartashevich[1] [0000−0001−9631−3623]
Leonardo Oliveira[1] [0000−0002−6940−6223]
António Teixeira[1,2] [0000−0002−7675−1236]
Samuel Silva[1,2] [0000−0002−9858−8249]

[1] DETI, University of Aveiro, Campus Universitário de Santiago, Portugal
[2] Institute of Electronics and Informatics Engineering of Aveiro (IEETA), Portugal

Abstract. In several situations, such as when we are eating or giving a presentation in a conference, it is unnatural and awkward to be forced to have direct contact with a computer keyboard or mouse. New interaction modalities can be very useful in providing novel and more natural solutions. In this work, we explore new ways of interacting with a presentation given by someone to an audience. Making use of multimodality (speech input and output, body gestures and their fusion), the speaker can move between slides, point to the screen, highlight contents, play videos, control the volume, or ask a virtual assistant to read the contents of a slide. This approach potentially makes presentations more natural, without forcing the user to directly interact with a computer, allowing the desirable focus on the message and audience.

Keywords: Multimodal Interaction · Presentations · PowerPoint · Voice Control · Gesture Control · Usability.

1 Introduction

Imagining a presentation supported by slides in a conference, typically the speaker has the need to be close to a computer or to use a remote control to move between slides or pointing out something. This is unnatural and uncomfortable in a scenario where the speaker should be focused on the audience and the message being transmitted.

New interaction modalities can be very useful in providing novel and more natural solutions. Providing multiple modes of interacting with a system is an important way to adapt it to the user. From the more conventional ways, like text, to the increasingly explored ways, like speech or gestures, combining different modalities potentially allows the user to interact more efficiently and naturally. This is called Multimodal Interaction, and has been a subject with increasing focus in recent years.

Aiming at demonstrating the potential of multimodal interaction, including speech and gestures to free users from being close and in direct contact with a computer - as it is the case in a public presentation in a conference - a widely used system - Microsoft PowerPoint - was selected for a proof of concept.

This work started from this perspective, aiming to explore new ways of interacting with PowerPoint through different modalities, making it more pleasant and natural. After this introduction, in Section 2 we describe the approach taken to address these issues, with the results being presented in Section 3. We finalize with some final considerations, in Section 4.

2 Proof of Concept System

We can think of a scenario where the speaker (the user) is interacting with the audience, for example in INForum. The main ways of communication are speech and gestures. So, while interacting with people, he should also have the opportunity to interact with the system by speech and gestures. All of that without touching in anything, not even in a remote control. The needed functionalities include navigating between slides or pointing to the screen while away from the computer.

We designed a system where the presentation is supported by speech and gestures. Fig. 1 shows the system architecture. It comprises modules for the two modalities, fusion, management of the interaction, and the application controlling PowerPoint.

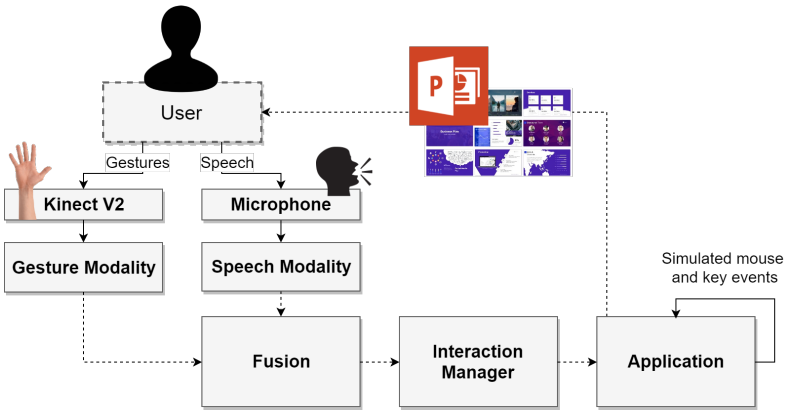

Fig. 1. The architecture of the system.

The gestures are captured by a Kinect (v. 2.0) [1] and the speech by a microphone. These are processed by two different nodes that run independently (Gesture Modality and Speech Modality). They are posteriorly merged in a Fusion node, responsible for combining them and forward the result to an Interaction Manager. For instance, the Fusion can merge redundant actions (like moving to the next slide by voice OR by gesture) or complementary actions (that are done by using voice AND gestures simultaneously). The Fusion and Interaction Manager modules are part of a Multimodal Framework [2] previously developed at IEETA, which allowed us to accelerate the implementation process.

The Interaction Manager provides an unique API to the application, that is in charge of controlling the PowerPoint presentation. This application simulates mouse and keyboards events [3, 4] to allow controlling several features. This includes, for instance, activating the pen tool to underline something. There is also voice feedback when it is necessary, through a virtual assistant.

As for the speech, the system, has a short grammar that recognizes sentences, that act as input commands. The range of recognized sentences was written to support several options, to get a more natural interaction. the gestures were chosen to remind the actions they are used for. Each gesture was trained with Visual Gesture Builder, recording 10-20 samples from each of two people.

The application has a small GUI at the bottom right of the screen to provide feedback to the speaker about the success of his actions (see Fig. 2). This can be specially useful when learning how to use the system. Otherwise, it can be hidden.

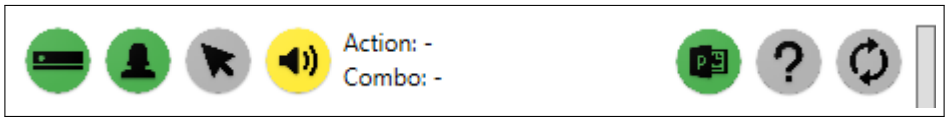

Fig. 2. The GUI. The first four icons show the status of the Kinect, the virtual assistant, the mouse control and the volume control. In the middle, the detected actions are shown. The last three icons are buttons to select a PowerPoint presentation, show an help dialog or refresh a presentation already selected (in case it is edited). The bar on the right is a switch with three different levels to collapse the interface.

3 Results

With a system built as depicted in Section 2, it was possible to cover a wide range of use cases. In what follows we illustrate some of the possibilities of using the system. For a complete view of what can be done please watch this video[3] with a demonstration.

The user can move between slides using gestures (see Fig. 3) or throughout the speech, saying something like "Please, move to the next slide"[4].

The user can also open his hand and point to something in the presentation. A red dot simulating the effect of a laser pointer is shown, on screen, moving with the hand's movement. If desired, he can use his forefinger to activate the pen tool, which will allow highlighting particular content in a slide. If the slide contains a video or any other clickable object, the user can close the hand to click it. In the case of a video, it will start playing or pause.

The user can control the sound of the system. For that, he places the left hand close to the ear and rises or lowers the right hand according to the intended variation of sound. The variation follows the progressive movement of the arm.

[3] https://youtu.be/pcbRcsfxFX4

[4] The system is designed in Portuguese. This and the following discourse examples are direct translations from the original interactions.

Some other examples of interaction by speech may include asking to read the content of a slide [5] - "Read this slide." - or asking the time - "What time is it?". Maria, the virtual assistant, will answer in accordance.

Additionally, the proposed system can be used by several users simultaneously. For instance, while one speaker starts playing a video, the other adjusts the sound.

Fig. 3. Some gestures. From left to right: change a slide; use a pointer; underline; control the volume.

4 Conclusion

The designed system tackled the issue of controlling presentations directly with a computer or using a remote control. Instead, having free hands allows the speaker to focus on the audience. We explored new ways of interaction that we expect to be more natural and comfortable to the speaker.

The implementation consists of an ad hoc solution, limited to what we can control externally (simulating keyboard and mouse events) and using a Kinect, that is costly and might not be easy to carry and setup. Even so, in an ideal scenario, an auditorium could have an equipment to track the speaker's gestures, as it already has projectors and microphones. We noticed some issues with noise while using speech as input, being that an aspect to improve in the future.

We consider the developed solution a good proof of concept of what can be done in the specific scenario of giving presentations, using multimodality to make systems more adapted to the user instead of forcing the user to adapt to the systems.

References

1. Rahman, M.: Beginning Microsoft Kinect for Windows SDK 2.0: Motion and Depth Sensing for Natural User Interfaces. 1st edn. Apress, CA, USA (2017)
2. Almeida, N.: Multimodal Interaction - Contributions to Simplify Application Development, PhD thesis, Universidade de Aveiro, (2017)
3. Use keyboard shortcuts to deliver PowerPoint presentations, https://support.office.com/en-us/article/use-keyboard-shortcuts-to-deliver-powerpoint-presentations-1524ffce-bd2a-45f4-9a7f-f18b992b93a0. Last accessed 8 Jul 2018
4. SendKeys Class, https://msdn.microsoft.com/en-us/library/system.windows.forms.sendkeys.aspx. Last accessed 8 Jul 2018
5. Save Text from PowerPiont to .txt/.doc in C#, https://code.msdn.microsoft.com/office/Save-Text-form-PowerPiont-a5744d3c. Last accessed 8 Jul 2018