University of Fribourg, Switzerland

Mini-Projet : Kemote

Final report

Multimodal Interfaces

Thi Thu Hang Nguyen – Thibaud Chardonnens

Contents

1. Objective

Our team was assigned a project whose objective is to develop Kemote which stands for R**emote** by **K**inect. This software is a remote control application running under Windows operating systems.

2. Motivation

The interest of Kemote is to allow us to gain experience in the domain of voice and gesture recognition.

With the commercial products currently available in this domain, each Windows application requires a specific single-purpose remote. For example, Kensington Presenter Pro Remote for PowerPoint, HP remote control (picasso) for Windows Media Center, etc. We would like to prototype a multi-purpose remote.

The Kemote project allows us to exploit the Microsoft Kinect SDK and other open source SDKs to implement Kemote functionalities.

Last but not least, by implementing Kemote successfully, we fulfill the requirements of the Interface Multimodal course, namely, each student group must develop its own application by applying the knowledge acquired in this course.

3. Functionalities

Kemote is designed to remotely control PowerPoint, Microsoft Word, Windows Media Player, Internet browser, etc. through voice and gesture recognition. Within our limited time frame in the spring semester 2012, we select a limited set of functionalities to implement.

For PowerPoint Presentation, the following remote controls were developed:

- launch PowerPoint application
- change the slides: next, back

For Windows Media Player (WMP), the selected remote controls are:

- launch WMP
- start, stop, next, back

The Internet browser controls are:

- launch navigator
- roll up, roll down

4. Implementation

4.1. Hardware

To run Kemote, a PC running Windows and a Kinect Xbox 360 are required.

1. Kinect

Kinect is a motion sensing input device developed by Microsoft for the Xbox 360 video game console and Windows PCs. It permits users to control and interact with the Xbox 360 without manipulating a game controller. In our project with Kinect, instead of using the PC keyboard and mouse, the users issue commands through natural gestures and voice.

Kinect can be simultaneously used with:

- the Wii Remote Plus via the Wii console
- and PlayStation Move with PlayStation Eye motion controllers via PlayStation 3 home console.

To obtain excellent precision of gesture and voice recognition, up to 4 Kinect sensors should be plugged into the same computer. Within our limited budget, we use only one Kinect sensor.

2. PC running Windows

In its current version, Kemote runs on Windows platform. We will continue to improve Kemote in the future versions, to include other platforms such as Linux and Mac OS by using open source SDKs.

4.2. Software

4.2.1. Programming language and libraries

We developed Kemote using Microsoft Kinect SDK v1.0 on Windows platform and C#.

4.2.2. Modalities

The input modalities provided in Kemote are:

- Speech, captured through Kinect Xbox 360
- Gesture, captured also by Kinect Xbox 360.

The users can choose either "speech" or "gesture" or both to issue remote commands, with the restriction that the first command to launch an application (PowerPoint, Word, WMP, Internet browser, etc.) must be vocal.

4.2.3. CASE & CARE models

1. The CASE properties

Among the four properties of the CASE model, shown in figure 1 (**C**oncurrent, **A**lternate, **S**ynergistic and **E**xclusive), our Kemote software supports the Exclusive property, meaning that Kemote executes one task after the other using one modality at a time, no co-reference.

Figure 1: CASE Models

2. The CARE Properties

Among the four properties of the CARE model (**C**omplementarity, **A**ssignment, **R**edundancy and **E**quivalence), Kemote supports complementarity, assignment and equivalence.

- Complementarity: to play a song on WMP, the user loudly pronounces "windows media" to launch WMP, and simultaneously waves his right hand.

- Assignment: voice modality must be used as the first command to open a windows application. The user has no other choice.

- Equivalence: waving the right hand and pronouncing "play" are equivalent modalities to play a song on WMP.

5. User evaluation

In order to evaluate our Kemote, we conducted a market survey with ten selected users.

a. The survey contains the following questions:

1. Do you ever have experience with wireless presentation remote designed for PowerPoint?

☐Yes   ☐No

2. Do you need a multi-purpose remote designed for PowerPoint, Internet browser, Windows Media, etc.?

☐Yes   ☐No

If yes: ☐ As soon as possible

                    ☐ Next 3 months

                    ☐ 4 months to 6 months

                    ☐ 6 months to a year

                    ☐ > Year

3. Can you estimate the price that you are willing to pay for such remote?

☐ 10 CHF   ☐ 15 CHF   ☐ 20 CHF   ☐ 25 CHF

4. After testing Kemote, please indicate below your degree of satisfaction.

☐ excellent   ☐ good   ☐ satisfactory   ☐ poor

b. Survey results

1. 80% users answered yes, 20% no.

2. 5 users need a multi-purpose remote.

Out of these 5 users:

- 1 needs it immediately
- 2 from 4 to 6 months
- 1 more than one year
- 1 doesn't know

3. 100% interviewees agree to pay the remote at 10 CHF.

4. The high percentage of "good" and "satisfactory" shows that Kemote needs to be improved by adding more remote functionalities.

6. Problems and limitations

During the development process, we encountered the following technical problems:

- The function keys are defined differently depending on manufacturers, operating systems and applications. For example, the counterpart of the *PageUp* key on the HP keyboard is the "*fn+pg up*" key combination on the keyboard of a Pavilion dm1 desktop. The current version of our software Kemote can only interpret keys on HP keyboard under Windows operating systems (XP, Windows 7).

- In order to guarantee a good image quality, users must stand in front of the Kinect sensor within 1.2 to 3.5 m.

- The voice recognition sensor has low performance. For example, it takes time for the Kemote to interpret and execute a vocal "play" command.

- The voice sensor recognized precisely vocal commands issued in native American or British. Non-native English accents are rarely interpreted without errors.

7. Conclusion

The Kemote project allows our team to put into practice the theoretical knowledge in the course "Multimodal Interfaces": human – computer interaction, CASE & CARE models, user evaluation of application software through open-close questions, etc.

As students in management who attend a computer-oriented course, this project is a great opportunity for us to learn and practice the programming language C# from scratch. At the beginning, C# requires tremendous effort and time from us. As the project advances, we become more and more comfortable and skilled with the hardware and software environment of Kemote.

8. Bibliography

Course. (n.d.). Retrieved from http://diuf.unifr.ch/diva/courses/multimodal/projects.html

Denis Lalanne, Jürg Kohlas. (2009). *Human Machine Interaction: Research Results of the MMI Program.* Springer.

Kinect for windows. (n.d.). Retrieved from http://www.microsoft.com

Tzovaras, D. (2008). *Multimodal User Interfaces: From Signals to Interaction* (1 ed.). Springer.