

Support for inferring user abilities for multimodal applications

Carlos Duarte Pedro Feiteira Daniel Costa David Costa

LaSIGE and Informatics Department

Faculty of Sciences of the University of Lisbon

Campo Grande, 1749-016 Lisboa, Portugal

{cad,pfeiteira}@di.fc.ul.pt, {thewisher,dcosta}@lasige.di.fc.ul.pt

Abstract

Multimodal interaction offers users the possibility to interact with computational systems using natural interaction modes. However, different users interact differently, integrating modalities in different patterns. By knowing how users exploit different modalities, we can try to build adaptive systems that will increase the effectiveness of multimodal interaction. For that we need to characterize user abilities. However, this is a time and effort consuming process, particularly in the early stages of design. We propose an approach that can be used in the early stages of the design process, by resorting to a Wizard of Oz based prototype of a multimodal system.

Keywords

Multimodal interaction, Adaptive interfaces, Adaptive multimodal systems, Design, Prototyping, Wizard of Oz.

1 INTRODUCTION

Multimodal interaction, by offering alternative input and output modalities, provides users with more natural interaction choices. By resorting to modalities like speech and gestures, they bring interaction closer to what users are accustomed from human-human interaction [Dumas 09]. Additionally, users can benefit from the flexibility of having different alternatives to interact. If the user can't point to an object to select it (physical impairment) speech recognition can replace the selection modality employed. Thus, natural interaction modes and the flexibility to allow users to select the most appropriate modes for them, are two of the main advantages of multimodal interfaces.

However, in order to take full advantage of this flexibility, and the associated variability in the context of use it brings, we need to assist the multimodal application. It is easy to envision various factors, sensory, physical and cognitive, that can impact the use of the different available modalities. Besides these, other factors, related to social or environmental characteristics, can also have such an impact. In order to tackle the added complexity of managing this number of variables, we can endow the system with capabilities to adapt to its user, thus becoming a multimodal adaptive system. As a result, not only users can choose the modalities they use to interact, but also the system can select the more appropriate modalities to present information to the users, and even improve the performance of recognition based modalities for input [Oviatt 08].

In order to support adaptation, the system will need to be able to identify its users and build a user model for each one. This requires a system that is capable of identifying

the user's abilities, both for its use and for the construction of the user model. To allow a multimodal adaptive system to have access to this information and to use it effectively from the beginning, some kind of initialization process is required. However, during the early stages of design, when there is not a clear definition of the user model attributes yet, to develop such an initialization procedure, using multimodal interaction, would require an effort disproportionate to the results that could be attained from it.

To support achieving the goal of characterizing the user, without developing a full scale multimodal application, we propose a *Wizard of Oz* based approach, allowing a human operator to replace some of the recognizers, while still supporting the goal of assisting in understanding what user characteristics are relevant for the application in development and how different users combine different modalities.

2 CONTEXT

Although the aforementioned motivation and concerns are universal, this approach is being developed in the scope of *GUIDE*, an EU funded project. *GUIDE* deals with multimodal interaction tailored for the elderly population in a home setting, with a TV as the central hub. However, the interaction means to be made available to the user are not limited to a TV screen, sound system and remote control.

Different modalities will be considered in *GUIDE*'s scope. Input will be provided through the remote control, speech, gestures, with and without the support of a surface, and recognition on a tablet based computing device. Output will be presented on the TV screen, and also through sound, speech and haptic feedback. Additionally, video and audio output will be combined in the form of avatars

to increase the personalization effect and try to create an emotional attachment between users and system.

3 THE PROTOTYPE

The *Wizard of Oz* based prototype aims at collecting enough information to build the initial version of the user profile. Although the profiling application will be used only once by each user, it has the responsibility of being the first contact users will have with the whole system, and thus contribute to their initial impressions of it. To acquire the required information, a set of different tests will be performed, involving several input and output modalities. The prototype's results will support two types of conclusions: first, how users behave in the face of what is requested from them; second, if the proposed tests are necessary and sufficient to collect all the required information.

3.1 User Abilities Characterizing Tests

As stated before, the aim of the application is to characterize the user's abilities. To work towards this aim, the prototype will have to allow the evaluation of tests in three main areas: sensory, physical and cognitive. The sensory tests will focus on visual and auditory related senses. To evaluate the user's vision impairments tests targeting visual acuity, and any impairments relating to text contrast and color [Ishihara 17] will be ministered.

Physical tests will enable the characterization of the user's physical abilities. In order to be able to support the most natural interaction possible, a full characterization of the user's abilities regarding gestures is the aim. To do that, we need to characterize the gesture making abilities regarding different attributes and different body parts. Tests will be made to evaluate the capability of making gestures with the head, the upper body and the lower body. The attainable gesture's amplitude, height and rotation will be evaluated.

The final set of tests aims at characterizing the cognitive traits that will be important when interacting with the set of applications considered in the scope of *GUIDE*. These traits have been initially identified as memory, working memory and recognition speed, and the set of tests used will be based on the Cambridge university's CANTAB tests [Sahakian 92].

3.2 Modalities of the Prototype

To be able to execute and gather data from all the tests described, the application will have to rely on video and speech recognition for input. As argued before, developing such an application to validate early stages of design would impose high effort demands. Our approach is to employ a *Wizard of Oz* based prototype that will replace the recognition based modalities with a human *Wizard*.

In what concerns the output modalities, text, video, speech and sound cover the requirements from the tests. These outputs are all controlled by the *Wizard* through the prototype's interface. The *Wizard* can also control the flow of the application's execution (what test to administer next), as well as the presentation parameters. The interface also

includes a set of fields to allow the *Wizard* to report and comment on the tests' results.

One feature of the prototype is the use of an avatar to interact with the user. The responsibility of asking the user to perform tasks and give feedback about the user's actions is placed on the avatar and not on a human experimenter. We opted for incorporating the avatar since the early stages of prototyping, since it is important to understand if, by using a humanized figure to interact, we are able to strengthen the engagement of elderly users with such a system.

3.3 Automated vs. Manual Tasks

As stated above, the interaction with the application is controlled by the *Wizard*. However, for some of the tests it is extremely hard for the *Wizard* to accurately perceive the user's input. This is particularly true for pointing tasks. It is virtually impossible for a human to accurately perceive the point that is being pointed at on a TV screen by other person. Our solution was to integrate a custom made application that uses a *WiiMote* as the input device, and allows for presenting on the screen a cursor showing what is being pointed at. The cursor can appear only on the *Wizard's* screen or on both the *Wizard's* and the user's screens.

4 CONCLUSION

In this paper we presented our approach for designing an application for the profiling of users of a multimodal adaptive system. Multimodal systems are dependent on different recognition technologies to support natural interaction modes. Given that, both the individual recognition technologies and their integration are still expensive, in terms of man-power to develop, and prone to errors in individual recognition modalities, our approach rests on a *Wizard of Oz* based prototype. Using a *Wizard* to replace the recognition technologies we eliminate possible recognition errors while postponing the integration of the multiple recognition technologies to a latter stage in the design and development of the system. For the user profiling task it is not relevant that the recognition is technologically based, neither that the recognition technologies are integrated.

References

- [Dumas 09] B. Dumas, D. Lalanne, e S. Oviatt. Multimodal interfaces: A survey of principles, models and frameworks. Em *Human Machine Interaction*, páginas 3–26. Springer Berlin / Heidelberg, 2009.
- [Ishihara 17] Shinobu Ishihara. *Tests for colour-blindness*. Hongo Harukicho, Tokyo, 1917.
- [Oviatt 08] S. Oviatt, C. Swindells, e A.M. Arthur. Implicit user-adaptive system engagement in speech and pen interfaces. Em *CHI*, páginas 969–978, 2008.
- [Sahakian 92] B.J. Sahakian e A.M. Owen. Computerized assessment in neuropsychiatry using cantab: discussion paper. *J. of the Royal Society of Medicine*, 85:399–402, 1992.