
Minimizing Modality Bias When Exploring Input Preferences for Multimodal Systems in New Domains: the Archivus Case Study

Agnes Lisowska

ISSCO/TIM/ETI

University of Geneva

40 Blvd du Pont-d'Arve

1211, Geneva, Switzerland

Agnes.Lisowska@issco.unige.ch

Mireille Betrancourt

TECFA/FPSE

University of Geneva

Rue des Acacias 54

1227, Geneva, Switzerland

Mireille.Betrancourt@tecfa.unige.ch

Susan Armstrong

ISSCO/TIM/ETI

University of Geneva

40 Blvd du Pont-d'Arve

1211, Geneva, Switzerland

Susan.Armstrong@issco.unige.ch

Martin Rajman

Center for Global Computing

Ecole Polytechnique Fédérale de

Lausanne (EPFL)

CH-1015 Lausanne, Switzerland

Martin.Rajman@epfl.ch

Abstract

In this paper we discuss the problems faced when trying to design an evaluation protocol for a multimodal system using novel input modalities and in a new domain. In particular, we focus on the problem of trying to minimize bias towards certain modalities and interaction patterns that might be introduced by experimenters in the instructions given to users which explain how the system can be used.

Keywords

Multimodal interface, input, evaluation

ACM Classification Keywords

H.5.1 Multimedia Information Systems, H.5.2 User Interfaces

Introduction

The design and implementation of user interfaces for novel domains involves a particular set of challenges. Traditional methods for user requirements gathering can be insufficient since users are unlikely to have performed the proposed task before, or because the

Copyright is held by the author/owner(s).

CHI 2007, April 28 – May 3, 2007, San Jose, USA

ACM 1-xxxxxx

new interface provides users with access to an information space that they are unfamiliar with. In both of these cases users have a hard time imagining how they might want to use or interact with such an interface. The situation becomes even more complicated when novel input mechanisms such as voice or pen are introduced into the same system because designers not only have to foresee optimal functionalities and layouts, but also how users might use the novel modalities to interact with the system.

One possible solution to this problem is to create a high-fidelity prototype and then perform a series of evaluations that serve a two-fold purpose - to evaluate the design hypotheses, and to see how users react to and use the novel input modalities. However, the implantation of such exploratory systems can be quite costly and their evaluation rendered difficult by the fact that several interconnected factors must be evaluated simultaneously (in particular the choice of input modalities is strictly coupled with the overall usability of the system) and that in addition to learning how to interact with a new interface and in a new domain, the user needs to be convinced of the fidelity and utility of the novel input devices.

In our work with the Archivus system we opted for the high-fidelity prototype, reducing development overhead by using the Wizard of Oz technique during evaluations to replace some of the key functionalities that are difficult and costly to implement robustly with little initial data, such as natural language processing components and dialogue management strategies. However, we were still faced with the problem of how to evaluate the different elements of the system without influencing the user as to input modality

choices and interaction patterns. This paper presents the solution we adopted and discusses its benefits and drawbacks.

Background on the Archivus system

The new domain we are working in is multimedia meeting browsing and retrieval. Within the IM2 project [4], meetings are recorded in specially equipped SmartRooms which are capable of recording everything that happened in a meeting (video, audio, slide capture and electronic whiteboard use). Consequently, meetings that take place in these rooms leave an extensive series of browsable and searchable artifacts in the form of audio and video files, electronic copies of documents (physical and electronic) used during the meeting, and a text transcript of what was said. Moreover, this multimedia data is annotated (topics, dialogue acts, argumentative classes, etc) and all of this information, along with meta-data about the meeting such as the names of participants and the date and location of the meeting is stored in a multimedia database. As a result, users have access to a much richer resource of information about meetings than they ever had before, and ways to explore this data at levels beyond standard keyword search on text. However, due to the novelty of this type of information in this domain, users have few preconceptions about how best to exploit the data.

Archivus is an interface that acts as a portal to the multimedia meeting database. It represents meetings as books, and the entire database as a library. The interface is shown in Figure 1 below and more details about the functionalities available in the system can be found in [3].

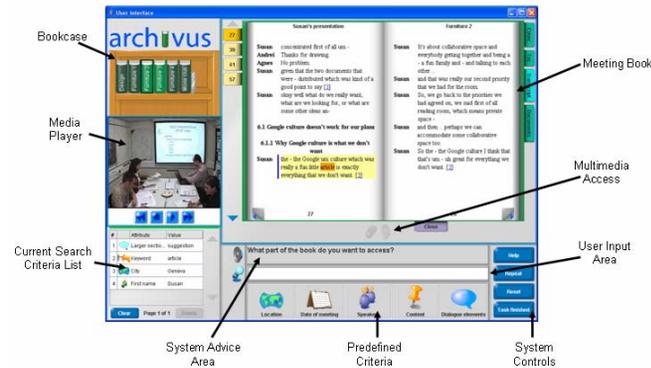


Figure 1. The Archivus interface

As has already been mentioned, it is difficult to gather user requirements for a new domain. We chose the questionnaire approach, where we explained to users the type of data and annotations that would be available and asked them to list natural language queries relevant to the types of information that they might want to find in the system. These queries gave us some insight into the key functionalities that might be necessary to include in the design [2].

However, our goal in the design of the system was primarily to make it as user friendly as possible, so while we based our design on indications from the study, we also wanted to include as few *a priori* assumptions about how users would interact with the system as we could. This meant that Archivus should allow for a variety of input modalities (pen, mouse, voice and keyboard) which we felt would be most suited to the domain and environment of use for such a system. Moreover, in order not to limit the user as to when and how a particular functionality might be used,

we made the system flexibly multimodal, which means that any of the modalities can be used at any time to access any part of the interface. Finally, as a basis for the dialogue strategies foreseen for linguistically driven interactions, we used the Rapid Dialogue Prototyping Methodology developed at the EPFL [1].

Choosing the evaluation protocol

We were faced with several challenges when selecting and developing the evaluation protocol. The first was that the results had to meet the needs of several interdependent research goals. One of these was looking at the types of dialogue strategies that would be needed. Another was to examine the nature of the natural language use (vocabulary, grammar etc) in order to develop robust natural language processing capabilities. Finally, we wanted to examine which of the proposed input modalities users would find most useful, both overall and for specific tasks.

We also had several practical constraints to take into consideration – specifically, limited time and manpower for running the experiments with large numbers of users. This meant that the experiments had to be fairly short (2 hours per user), and would not require users to come back for several sessions over a period of time.

Finally, and perhaps most importantly, since the system was developed to have as few *a priori* assumptions as possible about interaction, we wanted to leverage that freedom in the evaluation. Consequently, we wanted to minimize the amount of influence that we had over how users would use the input modalities available and the interaction patterns that they would use to find information in the system. This turned out to be the most difficult task to

accomplish and we considered and eliminated several possible protocols, some of which are discussed in the following paragraphs, before settling on the one that we believed best suited our needs.

Free exploration, where the user 'plays' with the system on their own and without any instruction for a fixed amount of time, was eliminated since we believed that users would not discover all of the functionalities of the system, nor would they try sufficient degrees of freedom with the various novel input modalities to recognize the benefits that they could provide.

Structured exploration, where the user would be given a manual for the system and would be able to interact with it at the same time was also rejected. While in this case we could ensure that the user was made aware of all of the possible functionalities of the system, we were still left with the risk of them not being willing to explore using the novel modalities on their own.

A protocol that seemed plausible and would provide a solution to the two problems presented by the preceding protocols was to have the experimenter do a guided tutorial with the user, where they show the user both the different functionalities of the system that are available and how to access those functionalities using different modalities. This would ensure that the user sees all of the functionalities of the system that are available, and also a variety of ways in which to access them. Moreover, there would be no paper trace of the interaction patterns presented by the experimenter, so the user would find it more difficult to emulate the patterns in their own interactions. However, we did find one significant drawback to this protocol – it seemed to be impractical for large-scale evaluations. Since we

were planning to run an evaluation with over 80 users, it would be very difficult for the experimenter to ensure that each user was given the same amount of information in the same manner. It would be very difficult to control for variables such as social factors and experimenter fatigue.

Finally for the pilot experiments we opted to give the user an explanation of the system on paper, as we might have with the structured exploration protocol, but in this case, the user was not allowed to interact with the system until it came time to do the actual evaluations. This was a conscious choice motivated by the fact that this way, there would be no 'priming' or favoritism for specific modalities before the experiments began. All users would have the exact same level of knowledge about the system, and there would be no established interaction patterns that the user could follow. The detailed protocol is given in the following section.

The pilot experiment

The pilot experiment involved 24 users who were first given a demographic questionnaire to fill in and a consent form to sign. Then they were given an explanation of what the evaluation was about – specifically, they were asked to pretend that they had just been hired by a company and had been asked by their manager to do some fact finding and checking using Archivus. Finally, they were given a document which described the system and all of its functionalities in detail. At the time however, users were not allowed to interact with the system itself.

Once they had filled out the forms and read the documents, they began the two evaluation sessions,

where users had 20 minutes per session to answer a series of true/false or short answer questions using Archivus. In the first session, users had access to a subset of all available modalities. This was so that the experimenters could compare the performance of specific modalities and modality combinations to one another. In the second session, users had access to all available modalities, which enabled the experimenters to examine whether users had a preference for certain modalities over others. The final part of the protocol was a paper-based questionnaire that was meant to elicit their subjective impressions of using the system. The experimenter was not in the room during the sessions except when the user was filling out the second questionnaire.

Unfortunately, we found that the explanation of the system alone was insufficient for the user to be able to understand the system functionalities and be comfortable enough with the system to perform the evaluation tasks. A significant number of users said that they would have liked to have had a tutorial for using the system.

The challenge of an unbiased tutorial

We had been initially reluctant to provide users with a tutorial since we thought that it would unduly bias both the interaction paths that they chose and the modalities used for specific interactions. However, the obvious difficulties that the users had in using the system to answer the questions made it evident that one would be necessary.

Ideally, users would have been trained in the use of the novel modalities such as voice and pen input on other interfaces so that when the time came to evaluate the

Archivus system, they would be comfortable with them and would only have to learn the functionalities of the system. However, both time and manpower constraints made this impossible. Consequently, the tutorial had to not only teach users about the different functionalities and data types available in the system, but also to prove to users that the novel technologies worked well and could be beneficial to the users' interaction with the interface.

We created a separate tutorial for each of the possible modality combinations under examination (ten in all). The example questions, which took users step-by-step through how to find the answers, were the same in each of the ten tutorials, as were the interaction paths used to reach the answers, except where the nature of a given modality enforced a change. For example, when using mouse and pen the path taken to find a keyword involves more steps than in the case of using voice, where the user would immediately be able to specify the keyword. Moreover, we tried to keep the language as modality-neutral as possible, using words such as 'select' instead of 'click'. The new protocol using the tutorials is described in the following sections.

The evaluation protocol with a tutorial

As in the case of the pilot experiments, users were given the questionnaire and consent form, as well as the explanation of the evaluation. But, in this case, instead of the written explanation of the functionality, users were given the tutorial. The first page of the tutorial briefly explained in general terms the Archivus system and the parts of the interface. Users were also given a manual (a refined version of the explanation document from the pilot experiments), which explained in more detail the various parts of the system. The

users were told that the manual was there as a reference and that they did not have to read it. As in the pilot, the next 2 steps were the actual evaluation sessions, followed by the questionnaire.

Discussion and conclusions

We found that with the tutorials, users were much more comfortable with the system, exhibited by a larger success rate in answering questions, and were also more open to using the novel modalities such voice and pen.

We did however, notice that while users were now more willing to try out and continue to use novel input modalities, and in particular voice, they still tended to follow interaction patterns established in the tutorial to various degrees. With certain input modalities such as pen and mouse there were limited ways of reaching the information, so it was unclear whether their choice of pattern was driven by the tutorial or whether users simply believed that this was the optimal path to the information. However, in cases such as voice, or voice use in combination with other modalities where there are a larger number of possible paths to a solution, it became much more evident that users were following tutorial interaction patterns rather finding their own. This was of course not the case for all users – some quickly established their own ways of using the system.

We believe that in cases such as ours, it would be virtually impossible not to establish some interaction patterns in a tutorial. In the end, it appears to be a question of selecting the appropriate tradeoffs for the problems that one is trying to address, and to be careful to take any bias that might result from these tradeoffs into account when doing the data analysis.

Acknowledgements

The authors would like to thank Mirek Melichar, Marita Ailomaa, Pavel Cenek, Jiri Mikulasek, and Lukas Matena for their help in the design and implementation of the system and in executing the evaluation experiments, as well as Marcin Bogobowicz for the Archivus graphics. We would also like to thank the Swiss National Science Foundation for funding this work.

References

- [1] Bui, T.H., Rajman, M., Melichar M. Rapid Dialogue Prototyping Methodology. In *Proc. of TSD 2004*, Springer-Verlag (2004), 579-586.
- [2] Lisowska, A., Multimodal Interface Design for the Multimodal Meeting Domain: Preliminary Indications from a Query Analysis Study. 2003, University of Geneva: Geneva. p. 30.
- [3] Lisowska, A., Rajman, M. Bui, T.H. Archivus: A system for accessing the content of recorded multimodal meetings. In *Proc of MLMI'04*, Springer-Verlag (2004).
- [4] www.im2.ch