A USER DRIVEN EXPERIENCE IN THE DESIGN OF A MULTIMODAL INTERFACE FOR INDUSTRIAL DESIGN REVIEW

G. CONTI, M. WITZEL AND R. DE AMICIS
Fondazione Graphitech, 38050 Povo (TN), *Italy*
{giuseppe.conti,martin.witzel, raffaele.de.amicis}@graphitech.it

Abstract. In this work we present a novel approach to author multimodal dialogues through a graph-based user-centric approach tailored to virtual design reviews. Distinct modalities and the interaction dialogue itself need to be highly configurable according to each user's specific preferences and specific scenario. We will show how modalities can be bound together via a bidirectional graph in an authoring tool to allow the specification of application-specific domain keywords and user-centered composition of modalities. Modalities thus shall not be hard-wired to the application itself. Instead, we provide a complete persistent definition of all used modalities outside the application, including gestures, voice and the graphical user interface. Through incremental navigation within the graph, the user is enabled to specify precise modalities for interactions. Specifically, nodes and edges of the graph represent the dialogue of the user with the system. Furthermore, they reflect the application scenario through their names. Using such a flexible approach to interaction configurability has required to resolve the application functionalities in a flexible manner because of the complete separation of user interaction and application response.

1. Introduction

The research work presented in this paper shows the result of the development of a novel mechanism for the definition of a multi-modal interaction dialogue within a collaborative Virtual Reality environment for car design review, developed in the context of the EU project IMPROVE.

The entire system has been developed following a requirement analysis and use case definition which has helped capturing the user experiences and expectations from the system. The result of the development has brought to authoring an interface capable to create a user-tailored interaction dialogue based on manipulation of a so-called "interaction graph". This way the interaction level is completely de-coupled from the inner logic of the main application. Therefore this approach can be extended to any domain.

The results of the research have been extensively assessed through two test sessions run at the premises of a real world major car producer which was partner of the project. During the test several groups of two users were asked to assess the interface and its usability within a scenario representative of their real working environment. The final results have demonstrated the improvement in usability following the suggestion by the users.

This paper illustrates the development process, the test sessions, how these have been run and how the user's experience has played a major role in the design of the system. The experience has brought to substantial improvements in terms of system usability and has shown how a user-driven design can ameliorate the quality and effectiveness of multimodal interfaces. As a result of addressing the users' experience a two hand interaction interface has been developed within a personalised hardware configuration.

2. State of the Art

New interfaces have gone beyond the mere decoding of users' pointing actions by taking advantage of the information encoded through voice, gestures or gaze. This has led to multimodal VR interfaces where multiple communication channels (Latoschik, 2001) are used. The resulting interfaces make use of technologies which allow the user to interact through voice recognition and text-to-speech synthesis or gestures. Such interfaces differ fundamentally from traditional GUIs since they adopt a novel, probabilistic approach rather than a simple event-driven command mechanism.

Research works have also brought to the creation of portable multimodal VR/AR environments based on PDAs (Goose et. al., 2003). Voice is used to navigate, annotate, and communicate (through voice-over-IP) with other users while a context sensitive interface shows the available speech commands. Such a multimodal approach, although very promising, however lacks in standardized technologies, interaction paradigm and technologies.

The authors in (Kaiser, et. al., 2003) developed an integration of gestures and speech by recognizing signals in parallel. Unimodal recognizers were used to output lists of speech and 3D gesture hypotheses which were then routed to the time-aware multimodal integrator. The work of (Forbus et. al.,. 2001) proposed an open agent architecture to adapt to available input and output resources in order to provide distributed access to multimodal services. Another author in (Martin, 1998) introduced graphs for binding modalities. However, this was done on a non-semantical level and provided means of customization only via a specification language.

3. The Design of User-Centered Multimodal Dialogues

3.1. GRAPH-BASED APPROACH

During design reviews, several people with different professional backgrounds are working collaboratively and each user has potentially his/her preferred way of working with software systems. A big challenge of collaborative design reviews, in terms of interaction, is then to give users access to the same functionalities through customized user-centered modalities. For this reason it is essential to tackle the problem of distinct modality configurations according to each user's needs. This includes customization of gestures, voice commands as well as the Graphical User Interface (GUI).

The whole interaction infrastructure in fact needs to be highly customizable according to the user' s specific needs, to the design review scenario itself as well as ro his/her aesthetic taste in order to improve the users' efficiency and perception of the application. For example, the interface should allow an architect to define his personal interaction process when loading a 3D model regardless of whether he/she wants to use a circular gesture, to select a specific icon on the GUI or to use a speech command in his native language.

The problem becomes more complex when applications need to address distinct industrial products and scenarios. In fact the heterogeneity of design review participants requires distinct modalities and in fact the interaction dialogue itself to be highly configurable according to each user's specific preferences and scenario definitions.

The system described in this paper integrates voice commands, gestures and traditional dialog elements which can be tailored to the users' interaction preferences and scenario requirements. This is achieved by binding modalities together via a bidirectional graph. The nodes and edges of the graph represent the dialogue of the user with the system. These way users are enabled to specify precise modalities for interactions by navigating through the graphs.

The dialogue of the user with the application and in fact the behavior and functionalities of the application are custimisabe by using a graphical authoring tool. For this purpose we have developed a command interpreter which provides the end user with intuitive but powerful means to control the application.

Our two-tier model decouples the application's functionalities from the actual user interactions and it offers an exceptional degree of freedom for customization with respect to gestural and voice input. The interaction can be designed according to each user's needs by using a visual editor. Further we have offered the user multimodal sequential input for controlling the applications behavior. It has to be highlighted that user profiles can be realized in configurations stored outside the application. For this reason we have been able to regionalize the program to use different languages with respect to speech input and dialog content.

3.2. MULTIMODAL AUTHORING

Editing interactions directly in the aforementioned persistent graph definition text file yields satisfactory results for small graphs and with users with a deep insight into the application and its functionalities. However it is rather inconvenient to the targeted group of design reviewers. Interaction nodes and edges have to be creatable, editable and understandable with ease. For this we have decided to use a commercial software distribution to facilitate modifications of graphs like to allow functionalities such as creation and removal of vertices, altering vertex names, etc.

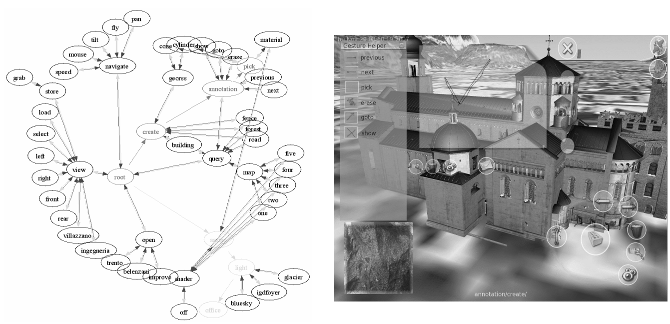

*Figure 1.*An Interaction Graph (left), the Personalized Graph-based User Interface (right).

4. Evaluation and Assessment

Specifically when dealing with user interaction, it is of utmost importance to perform a validation of the achieved results. The research work illustrated in this paper was validated by relevant users at the ELASIS research center who provided the authors with inspiring and constructive feedback.

Specifically we undertook two extensive test sessions during which advantages and shortcomings of our approaches became evident. The first user test, run at an early stage of the project, assessed the usability of the first collaborative user interface, which was hardwired to the application. At this stage only a touchscreen interaction modality was allowed. Therefore during this first test there were no means available for any customization with respect to devices and their scenario and use.

The results of the first test brought to some major changes as the user underlined the importance of an elevated usability with respect to collaboration, customizable interface design, application feedback, personalized interaction schemes as well as scenario specific navigation metaphors. Additionally after the first test it was decided to deliver a high degree of suggestive support for available interactions because of their high degree of flexibility, offering support for a training phase.

As illustrated in the following sections the resulting final system was eventually assessed once more by a representative user group. The final results were positive confirming the effectiveness of the proposed approach within a real industrial scenario.

4.1. USERS

The test session was run throughout a working day at the FIAT Research Centre, ELASIS, Italy with users being asked to assess the system in groups of two during their working activities. None of the users had been previously informed of the event in order to avoid cross influencing. The test involved 9 users from ELASIS, 7 men 2 women, all from Italy, aged between 25-34, with different profiles and with an average good experience with CAD/CAS. Each user was asked to specify his/her previous experience with VR systems. As a result the majority of users had very good experience with VR while 30% of users had no experience at all. Four users had used the previous version of the system during the first test session.

4.2 ASSESSMENT PROCEDURE

At the beginning of the test, users were invited to enter the VR lab and were given a short 5 minutes introduction to the system and to the hardware configuration. Users were explained the concept of the interaction graph, being used by the system, and how it was possible to change the entire interaction architecture via a simple configuration, by following the motto "Configure once, interact in any way". For this, it was shown to each group how to customize the interaction metaphor by creating the most appropriate combination of gestures/spoken commands/actions. Specifically, it was shown to each user how to use the interaction graph viewer and how to change the interaction dialogue. Eventually users were briefly taught that:

- Nodes of the graph define the actions.
- Connections define commands.
- Edges are used to define the interaction mechanism.
- Edges have properties.

- Actions/handlers are identified by path/order of user interactions.
- User interactions can be fully re-arranged and customized according to the application context and user-tailored.
- The separation between application and interaction definition.

An example of change in the configuration was carried on together with the user group. In particular, it was shown to users that it was possible to have seamless integrations of modalities by using nodes as definitions of actions/domains. Users were shown how edge attributes could be used to specify how to access nodes and how to change attributes to define gestures and speech available to each action (how to advance in the graph).

4.3. ISOMETRIC QUESTIONNAIRE

An ISOMETRIC questionnaire was prepared according to International Standard ISO 9241. The assessment has been done according to ISO guidelines based on the following seven categories:

- Suitability for the task.
- Self descriptiveness.
- Controllability.
- Error tolerance.
- Suitability for individualization.
- Suitability for learning.
- Conformity with user expectations.

4.3.1. Suitability for the Task

The largest difference was found in the fact that users felt that the software forced them to perform tasks that were not related to their actual work. However it was positively perceived by the users that the software allowed them to perform their tasks completely and that the arrangement of the elements was sensible with respect to the work to be carried out. Additionally, it was perceived that the way data was outputted reflected the tasks users needed to perform with the software. Positively a higher share of users considered that it was easy to adapt the software to perform new tasks and that the majority of the commands were easy to find.

4.3.2 Self Descriptiveness

The results in this category reflected a better tendency if compared with results emerged in the first test session. Results, albeit always close to average never got very high ranking. According to the users' feedback during the de-briefing, this was due to the fact that, all communication between the interface and the users, entirely lied on the graphical language with virtually no text. Although this made the entire interface much more compact, and once accustomed, much faster to use, at the beginning this required interpretation by the user. A positive point was that the user felt that it was possible to understand immediately what was meant by the information displayed by the software.

4.3.3. Controllability

The results in this group showed an improvement when compared to results from the first test. However results did not achieve a high ranking. According to the users feedback during the debriefing, this was due to the fact that all communication between the interface and the users entirely lied on the graphical indicators with virtually no text. Although this made the entire interface much more compact, and once accustomed, much faster to use, this required interpretation by the uninitiated user. A positive point is that the user felt that it was possible to understand immediately what was meant from the information displayed in the browser.

4.3.4. Error Tolerance

Users perceived that if they made a mistake while completing an action, they could not easily roll-back to a previous state, that when they made entries these were not first checked for correctness before further processing is initiated. A good share of this general feeling, according to the comments, was caused by the lack of an "undo" feature. This generated a sense of anxiety as users feared that potentially wrong action could not be corrected.

4.3.5. Suitability for Individualization

Extremely good results were collected in terms of suitability for individualization. The user perceived that: the software let them adapt forms, screens and menus to suit their individual preferences. The software could be easily adapted to suit their own level of knowledge and skill. They were able to adjust the amount of information (data, text, graphics, etc) displayed on-screen to their needs. The software let them change the names of commands, objects and actions to suit their personal vocabulary. They could adjust the attributes (e.g. speed) of the input devices (e.g. mouse, keyboard, tracked hand) to suit their individual needs. These very positive results stemmed from the high degree of customization provided by the authoring of the interface, which allowed every user, if required, to define a customized interaction multimodal dialogue.

4.3.6. Suitability for Learning

The results in terms of suitability showed that a short time was required to learn the software and that it was easy to re-learn and to use. The guidance through the novel graph-based approach via the helping functionalities was acknowledged very positively.

4.3.7. Conformity with User Expectations

The results showed a familiar pattern to previous results. The software was considered to be consistently designed and its behaviour easy to predict. Designations usage was considered less consistent with standard software the users were familiar with, although this was not necessarily a negative feeling as it represented a departure from the traditional way of working.

4.4. VIDEO ANALYSIS

The most important benefit highlighted by the users was the responsiveness of the system. The first version was slow for instance when rendering notes. This caused a sense of unease since the interaction process became slower. The current version allowed a more fluent interaction. Further, the use of the tablet was not positively considered as it did not allowed direct feedback and it obliged the user to interact with a device (the tablet) while receiving visual feedback from another (the screen). The touchscreen instead was very well accepted. Most users had noticed how the resolution was adequate and furthermore it allowed the use of fingers, in place of the pen. Users who had experienced the previous version of the system had also underlined that the touchscreen was preferred to the TabletPC also for the heat the latter produced, which made it very unpleasant to use after a few minutes.

References

Latoschik, M. E.: 2001, A gesture processing framwork for multmodal interaction in virtual reality, *Proc. of the 1st international conference on computer graphics, virtual reality and visualization.*

Forbus, K. D., Ferguson, R. W. and Usher, J. M.: 2001, Towards a Computational Model of Sketching, *Proc. of the 6th international conference on intelligent user interfaces.*

Goose, S. and Schneider, G.: 2003, Augmented Reality in the Palm of your Hand: A PDA-Based Framework Offering a Location-based, 3D and Speech-Driven User Interface, *in Proc. of Workshop on "Wearable Computing".*

Kaiser, E. et al.: 2003, Mutual disambiguation of 3D Multimodal Interaction in Augmented and Virtual Reality, *5th International Conference on Multimodal Interfaces (ICMI 2003),* pp. 12-19.

Martin, J. C.: 1998, TYCOON: Theoretical Framework and Software Tools for Multimodal Interfaces*, Intelligence and Multimodality in Multimedia Interfaces.*