Producing multimodal output in the COMIC dialogue system

Mary Ellen Foster

School of Informatics, University of Edinburgh
http://www.iccs.informatics.ed.ac.uk/~mef/

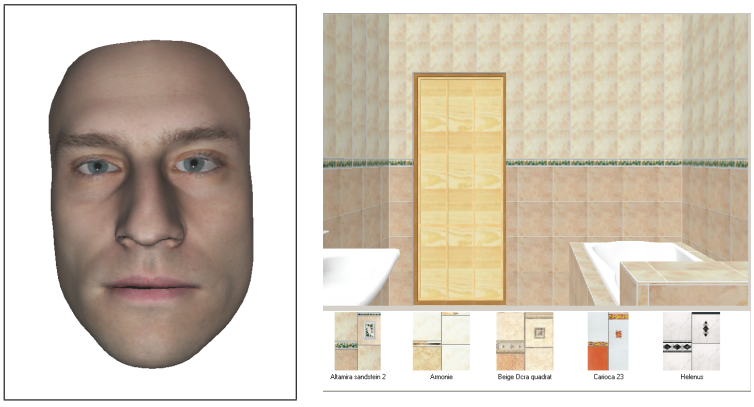
Research objective

COMIC[1] is an ongoing European project investigating multimodal dialogue systems. A main goal of the project is to use models and results from cognitive psychology in all parts of the system, in order to provide more intuitive interaction with the system.

For the output modules, a specific goal is to use an instance-based approach to plan presentation that are as similar as possible to known-good examples of multimodal output. This document describes how the multimodal output is planned and produced in the current project demonstrator, and outlines the plans for incorporating instance-based techniques into the final version of the demonstrator.

The COMIC demonstrator

As part of the COMIC project, a demonstrator system has been put together; this system is currently undergoing user evaluations as described in the following section. The demonstrator adds a multimodal dialogue interface to a CAD-like application used in bathroom sales situations to help clients redesign their rooms. The input to the system includes speech, handwriting, and pen gestures; the output combines synthesised speech, a "talking head" avatar, deictic gestures with a simulated mouse pointer, and control of the underlying application. Figure 1 shows the avatar and the bathroom-design application.



(a) Avatar (b) Bathroom-design application

Figure 1: Components of the COMIC demonstrator

[1] COnversational Multimodal Interaction with Computers; http://www.hcrc.ed.ac.uk/comic/.

The interaction with the demonstrator has four phases. First, the user specifies the shape of their bathroom and the location and size of the windows and doors, using a combination of pen and speech input. Next, the user is able to choose where the sanitary ware is to be located in their new room. After that, the system presents a number of options for tiles to use in the room, guiding the user as they browse through the (possibly very large) range of possibilities. Finally, the user is given a three-dimensional tour of the finished bathroom.

Multimodal fission in COMIC

In a multimodal system, the *fission* component is the one that chooses the output to be produced on each of the output channels, and then coordinates the output across the channels. The COMIC fission module processes a high-level specification from the dialogue manager, and creates and executes a concrete plan to realise that specification using the available channels. Figure 2 shows where fission fits into the overall system architecture; in this figure, "ViSoft" refers to the bathroom-design application.

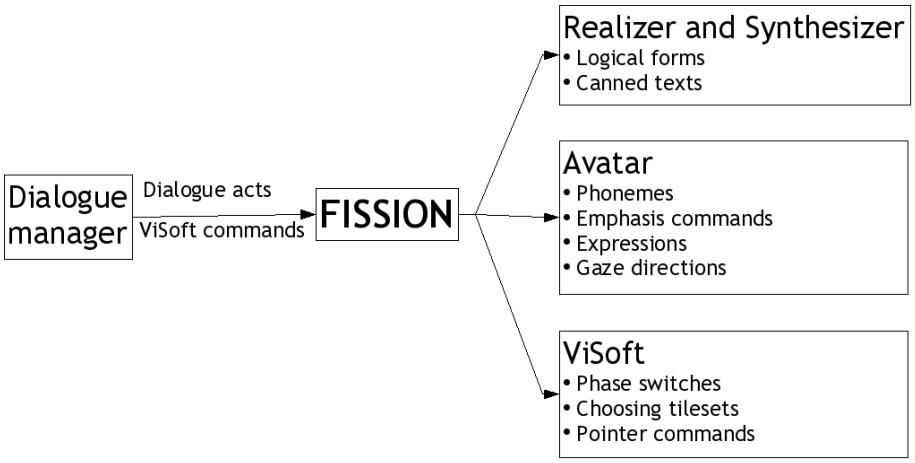


Figure 2: Input and output of the fission module

The fission module is implemented in Java, and uses XSLT template processing to perform most of the planning. The information about the available tiles is stored in a DAML+OIL ontology. Foster and White (2004) give a detailed description of the implementation of the text planner.

Figure 3 shows a sample plan generated by the fission module in response to an instruction from the dialogue manager to show and describe a particular set of tiles. There are three different types of nodes in this tree: standalone commands to be sent on a single output channel (the orange *nod* and *choose tileset* nodes); sentences, which may include coordinated facial expressions and mouse-pointer gestures along with the spoken text (the green leaf nodes containing text); and sequences of segments (the blue internal nodes).

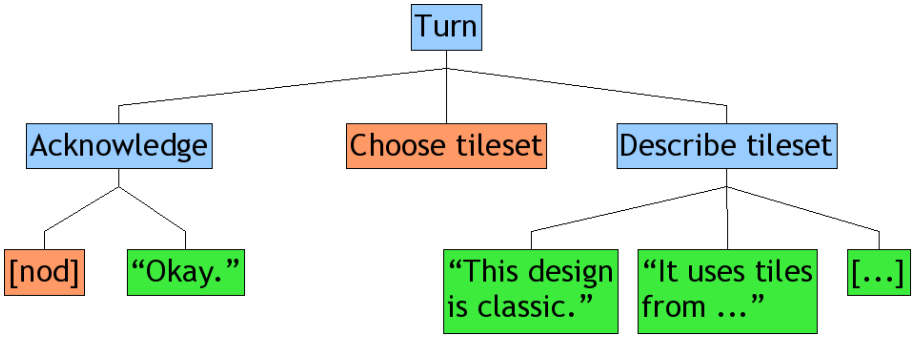


Figure 3: Output plan fragment

The current fission module addresses the goal of generating output based on human dialogues by planning the deictic mouse-pointer gestures using results from an annotated corpus (Foster, 2003) of human-human roleplaying dialogues in the same domain. Foster (2004) gives details of the findings from the corpus and how they are incorporated into the planning process. As described in the final section, we plan to incorporate instance-based selection of multimodal alternatives into future versions of the demonstrator.

Evaluating the demonstrator

Several user studies are currently under way to evaluate the current demonstrator, and to indicate the best ways to extend the demonstrator the final year of the project. The evaluations include the following:

- Comparing the success and satisfaction of users when they specify the shape of their bathroom using either the first phase of the COMIC demonstrator, or an existing, web-based system.
- Enabling and disabling the facial expressions of the avatar in the third (tile-selection) phase of the demonstrator, and examining the effect on user satisfaction and interaction smoothness.
- Comparing the naturalness and understandability of the corpus-based mouse-pointer gestures described in the preceding section, with that of the gestures that are generated in a rule-based manner.

Future development

In the final year of the COMIC project, we will focus on extending the capabilities of the system. In addition to expanding the range of output types in response to an increased range of tiles and increased capabilities of the output modules, there are two specific research directions for the fission module in particular.

We will investigate the use of instance-based generation techniques to provide a wider and more natural variety of multimodal output, by searching through a space of possible realisations of a message, and choosing the alternative that has the highest similarity to a corpus of "good" output examples. We will extend the fission-realiser interface to include explicit alternation in all output channels; the realiser will then choose the alternative based on n-gram comparisons against the target instance base.

We also plan to add an explicit model of the user's preferences to the system, in order to investigate how user modelling can improve guided browsing through a possibly very large set of alternatives. The user model will be used in the tile-selection phase to keep track of those features that the user expresses particular interest in. This will allow the system both to choose further tilesets to describe that they are likely to prefer, and to tailor the descriptions of those tilesets to the user's known preferences, using techniques similar to those used in the FLIGHTS system (Moore *et al.*, 2004).

References

FOSTER M E (2003). *Description of the "Wizard of Oz" recordings*. Deliverable 6.4, COMIC project.

FOSTER M E (2004). Corpus-based planning of deictic gestures in COMIC. In: *Proceedings of INLG-04 student session*. To appear.

FOSTER M E and WHITE M (2004). Techniques for text planning with XSLT. In: *Proceedings of NLPXML-2004*. To appear.

MOORE J, FOSTER M E, LEMON O, and WHITE M (2004). Generating tailored, comparative descriptions in spoken dialogue. In: *Proceedings of FLAIRS 2004*.

WHITE M (2004). Reining in CCG chart realization. In: *Proceedings of INLG 2004*. To appear.