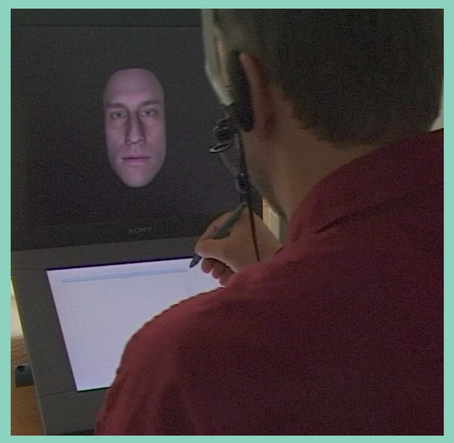
Multimodal Generation in the COMIC Dialogue System

Mary Ellen Foster Michael White
School of Informatics, University of Edinburgh

Andrea Setzer Roberta Catizone
Department of Computer Science, University of Sheffield

The COMIC System



Multimodal dialogue interface to a CAD-like application for bathroom redesign

Phases
- Specify blueprint of room
- Place sanitary ware
- Browse tiling options
- 3D walkthrough of room

Input
- Speech recognition
- Pen gestures
- Handwriting

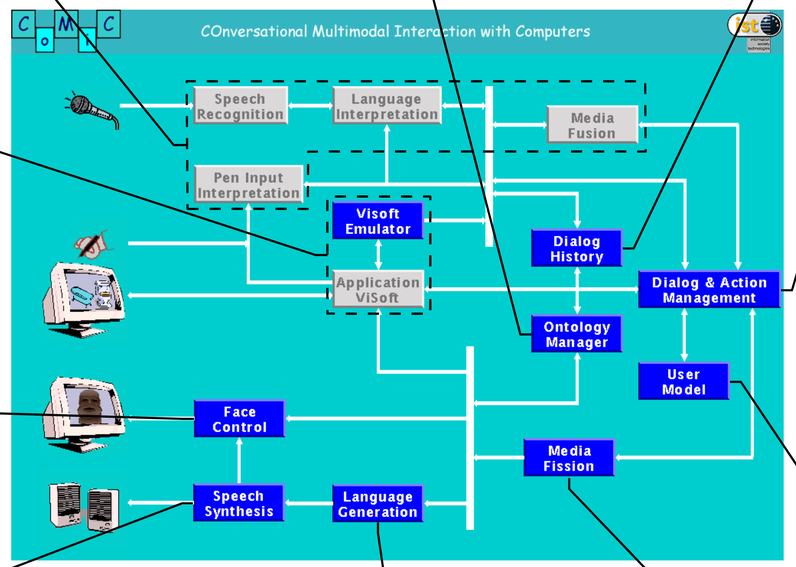
Output
- Speech synthesis
- Talking head behaviours
- Control of application
- Deictic pointer gestures

Details and publications
http://www.hcrc.ed.ac.uk/comic/

Input processing and fusion – not in the demo

Design application – emulated for the demo

Talking head – demo uses RUTH (DeCarlo *et al.*, 2004)



- Represented in DAML+OIL
- Stores information about all tile designs

- Which designs have been described
- What has been said about them
- Used by DAM and Fission

- Decides what to show and say in response to user input
- Input: multiple scored hypotheses
- Output: high-level output specification
- Domain-specific information stored in *Dialogue Action Forms* (DAFs)
- Dialogue state stored on stack
- Details: (Catizone *et al.*, 2003)

- Based on Multi-Attribute Utility Theory
- Stores features user likes and dislikes
- Used by DAM and Fission

- Client to Festival synthesiser (Clark *et al.*, 2004)
- Uses custom-build Festival 2 voice with support for prosodic mark-up

- Creates prosodically annotated text for synthesis
- Works in “anytime” mode
- Uses statistical language models to select highest-scoring output
- Based on OpenCCG (White, 2005a;b)

- Selects content to meet specification
- Coordinates output across channels
- Prepares and executes plans in parallel
- (Foster, 2005; Foster and White, 2005)

References

- R. Catizone, A. Setzer, and Y. Wilks. (2003) Multimodal dialogue management in the COMIC project. *Proc. EACL 2003 Workshop on Dialogue Systems*.
- R.A.J. Clark, K. Richmond, and S. King. (2004) Festival 2 – build your own general purpose unit selection speech synthesiser. *Proc. ISCA Workshop on Speech Synthesis*.
- D. DeCarlo, M. Stone, C. Revilla, and J.J. Venditti. (2004) Specifying and animating facial signals for discourse in embodied conversational agents. *Computer Animation and Virtual Worlds* 15(1):27-38.
- M.E. Foster. (2005) Interleaved planning and output in the COMIC fission module. *Proc. ACL 2005 Workshop on Software*.
- M.E. Foster and M. White. (2005) Assessing the impact of adaptive generation in the COMIC dialogue system. *Proc. IJCAI 2005 Dialogue Workshop*.
- M. White. (2005a) Designing an extensible API for integrating language modeling and realization. *Proc. ACL 2005 Workshop on Software*.
- M. White. (2005b) Efficient realization of coordinate structures in Combinatory Categorial Grammar. *Research on Language and Computation*, in press.