Position Paper for the W3C MMI Workshop
Graham Wilcock and Kristiina Jokinen
University of Helsinki and University of Tampere

SCXML, Multimodal Dialogue Systems and MMI Architecture

We are interested in the workshop because the topic is important for our research: we wish to learn more about how the MMI architecture supports

a) fusion of modalities
b) incremental presentation
c) design of cooperative interaction.

The first author has worked on XML-based language processing (natural language analysis, generation, and annotation) and is interested in practical applications of emerging technology, while the second author has played a leading role in several research and development projects on spoken dialogue systems, and is interested in prinicpled-based representation and architectures for implementing fundamental aspects of human communication. Our current purpose is to combine our previous work. We approach MMI from two different but related perspectives: SCXML as a basis for voice interfaces and cooperative multi modal route navigation.

SCXML as a basis for voice interfaces is described in [Wilcock 2007]. The basic ideas of statecharts are illustrated by means of a simple "stopwatch" demo with an open source Java implementation of SCXML. The basic version of the demo comes from Jakarta Commons SCXML [Apache Software Foundation 2006], an open source Java implementation of SCXML, and has a graphical user interface (GUI), which displays the time and enables the stopwatch to be started, paused, stopped and reset by mouse clicks. The voice user interface (VUI) is added to the demo using the Sphinx-4 open source Java speech recognizer [Carnegie Mellon University 2004] and the FreeTTS open source Java speech synthesizer [Sun Microsystems 2005]. In a "speaking stopwatch" version of the demo, when the user stops or pauses the stopwatch by a mouse click on the appropriate button, the time is read out aloud by the speech synthesizer. In addition, a brief prompt is spoken when the user starts, un-pauses or resets the stopwatch. In a "listening stopwatch" version of the demo, the user starts, stops, pauses, un -pauses, and resets the stopwatch either by voice commands or by mouse clicks. The speech recognizer uses a small JSGF grammar for the voice commands. All three versions (basic, speaking, listening) follow the same state transitions which are defined in the SCXML statechart file.

Cooperative multimodal route navigation is the basis of the MUMS system, a PDA-based route navigation system which allows the user to query public transportation information using spoken language commands and pen-pointing gestures on a map. It also provides route information in speech and graphical output. The system is described in more detail in [Hurtig & Jokinen 2005, 2006; Jokinen 2007], and its evaluation is reported in [Jokinen & Hurtig 2006]. However, although the MUMS system is based on the SOA (Service Oriented Architecture approach, and XML is used as a general interface langauge between the different layers of technology components, it is still rather application specific in its representation and processing of information. We are thus looking for a more systematic approach to develop applications, so that the infrastructure would allow easy experimentation with different technology components and their internal functioning.

Since SCXML supports a clean separation of data, logic, and user interface, based on the data-flow-presentation (DFP) architectural pattern, we believe it has benefits for our research. Our ultimate goal is to experiment with different possibilities to develop natural user interfaces; natural in the sense of allowing the use of natural language, and also in the sense of providing intuitive functionality for the user. Also, we are interested in investigating how the design of spoken dialogue systems and multimodal route navigation system with an emphasis on the human cooperation aspects will be enabled in SCXML and MMI. In dialogue research, SCXML has been proposed by [Kronlid and Lager 2007] as a basis for implementing the information-state update (ISU) framework for dialogue management, which is a promising approach we would like to pursue further.

The route navigation application resembles the “Driving Directions” user case of W3C Multimodal Interaction Use Cases since in both cases the system gives instructions of how to go forward. In both cases, the user needs to understand the instructions given by the system, and the system should “listen” to the user and observe if the message has gone across.

We are especially keen on finding good solutions for the interaction problems that are generally considered “natural” (in the above two senses) but which are also generally considered difficult due to lack of descriptive research and available technology:

(1) incremental representation of information and allowing the user to zoom in and out both verbally and on the map
(2) allowing users to give feedback concerning their understanding in different ways: providing an answer to an explicit question (“Did you say the Opera stop?”), continuing the interaction with an appropriate next step (“Give me the next piece of information”), and by subtle signalling in their speech (variation of pronunciation together with the length of the following pause can signal wish to continue rather than the end of one’s turn).

Concerning the topics of the workshop listed in the CFP, we would in particular like to address the following questions:

- Requirements for extensions to the MMI Architecture to improve the support of speech, GUI and Ink interfaces on portable handheld multimodal devices.
- How to dynamically select appropriate modalities.
- Use of scripts to enable the customization of the user interface based upon previous user input.
- Support for effective user interfaces for various modes of interaction, in terms of contextual prompts, constrained text input, and declarative event handlers, taking account of uncertainties in user input.

We are also interested in the solution to the following questions:

- How to process early and late information fusion.
- Plans to support multimodal applications and what standards are needed.
- Re-use of existing markup languages for prompts and constraints on user input.

References

Apache Software Foundation: The Apache Jakarta Project, Commons SCXML. http://jakarta.apache.org/commons/scxml/ (2006)

Carnegie Mellon University: Sphinx-4: A speech recognizer written entirely in the Java programming language. http://cmusphinx.sourceforge.net/sphinx4/ (2004)

Harel, D.: Statecharts: A Visual Formalism for Complex Systems. Science of Computer Programming 8, North Holland (1987).

Hurtig, T. and Jokinen, K.: On Multimodal Route Navigation in PDAs, 2nd Baltic Conference on Human Language Technologies, pp. 261−266, Tallinn, Estonia, 2005.

Hurtig, T. and Jokinen, K.: Modality fusion in a route navigation system. Proceedings of the IUI 2006 Workshop on Effective Multimoda Dialoguel Interfaces, pp. 19-24, 2006.

Jokinen, K.: Interaction and Mobile Route Navigation Application. In Meng, L., A. Zipf, and S. Winter (eds.) Map-based mobile services - usage context, interaction and application, Springer series on Geoinformatics, 2007.

Jokinen, K. and Hurtig, T.: User Expectations and Real Experience on a Multimodal Interactive System. Proceedings of Interspeech, Pittsburgh, US, 2006.

Kronlid, F. and Lager, T.: Implementing the Information-State Update Approach to Dialogue Management in a Slightly Extended SCXML. 11th International Workshop on the Semantics and Pragmatics of Dialogue, Trento, Italy. pp. 99-106, 2007.

Sun Microsystems: FreeTTS 1.2: A speech synthesizer written entirely in the Java programming language. http://freetts.sourceforge.net/ (2005)

Wilcock, G.: SCXML and Voice Interfaces. 3rd Baltic Conference on Human Language Technologies, Kaunas, Lithuania, 2007. http://conference.vdu.lt/viewabstract.php?id=158&cf=7

W3C: State Chart XML (SCXML): State Machine Notation for Control Abstraction. http://www.w3.org/TR/2007/WD-scxml-20070221/ (2007)