TOWARDS MULTIMODAL WEB INTERACTION

Web pages you can speak to and gesture at

Dave Raggett (W3C/Canon), Max Froumentin (W3C), Philipp Hoschka (W3C)

1. INTRODUCTION

W3C is developing standards for a new class of devices that support multiple modes of interaction.

1.1 The Dream

The Multimodal Interaction Activity is focused on developing open standards that enable the following vision:

Extending the Web to allow multiple modes of interaction: GUI, Speech, Vision, Pen, Gestures, Haptic interfaces, ...

Augmenting human to computer and human to human interaction: Communication services involving multiple devices and multiple people

Anywhere, Any device, Any time: Services that adapt to the device, user preferences and environmental conditions

Accessible to all

The Multimodal Interaction Activity is extending the Web user interface to allow multiple modes of interaction—aural, visual and tactile—offering users the means to provide input using their voice or their hands via a key pad, keyboard, mouse, or stylus. For output, users will be able to listen to spoken prompts and audio, and to view information on graphical displays. The specifications developed by the Multimodal Interaction Working Group should be implementable on a royalty-free basis.

1.2 Application Areas

The Multimodal Interaction Working Group should be of interest to a range of organizations in different industry sectors.

1.2.1 Mobile

Multimodal applications are of particular interest for mobile devices. Speech offers a welcome means to interact with smaller devices, allowing one-handed and hands-free operation. Users benefit from being able to choose which modalities they find convenient in any situation. The Working Group should be of interest to companies developing smart phones and personal digital assistants or who are interested in providing tools and technology to support the delivery of multimodal services to such devices.

1.2.2 Automotive and Telematics

With the emergence of dashboard integrated high resolution color displays for navigation, communication and entertainment services, W3C's work on open standards for multimodal interaction should be of interest to companies working on developing the next generation of in-car systems.

1.2.3 Multimodal interfaces in the office

Multimodal has benefits for desktops and wall mounted interactive displays, offering a richer user experience and the chance to use speech and pens as alternatives to the mouse and keyboard. W3C's standardization work in this area should be of interest to companies developing browsers and authoring technologies, and who wish to ensure that the resulting standards live up to their needs.

1.2.4 Multimodal interfaces in the home

In addition to desktop access to the Web, multimodal interfaces are expected to add value to remote control of home entertainment systems, as well as finding a role for other systems around the home. Companies involved in developing embedded systems and consumer electronics should be interested in W3C's work on multimodal interaction.

2. CURRENT SITUATION

The Multimodal Interaction Working Group was launched in 2002 following a joint workshop between the W3C and the WAP Forum. Relevant W3C Member contributions have been received on SALT and X+V. The Working Group's initial focus was on use cases and requirements. This led to the publication of the W3C Multimodal Interaction Framework, and in turn to work on extensible multi-modal annotations (EMMA), and InkML, an XML language for ink traces. The Working Group has also worked on integration of composite multimodal input; dynamic adaptation to device configurations, user preferences and environmental conditions; modality component interfaces; and a study of current approaches to interaction management. The Working Group is now in the process of being re-chartered for a further two years. The following organizations are currently participating in the Working Group:

Access, Alcatel, Apple, Aspect, AT&T, Avaya, BeVocal, Canon, Cisco, Comverse, EDS, Ericsson, France Telecom, Fraunhofer Institute, HP, IBM, INRIA, Intel, IWA/HWG, Kirusa, Loquendo, Microsoft, Mitsubishi Electric, NEC, Nokia, Nortel Networks, Nuance Communications, OnMobile Systems, Openstream, Opera Software, Oracle, Panasonic, ScanSoft, Siemens, SnowShore Networks, Sun Microsystems, Telera, Tellme Networks, T-Online International, Toyohashi University of Technology, V-Enable, Vocalocity, VoiceGenie Technologies, Voxeo

All participating organizations are required to make a patent disclosure statement as set out in the W3C's Current Patent Practice (CPP) Note. A separate page is being maintained for patent disclosures for the Multimodal Interaction Activity. The Working Group is obliged by its charter to produce a specification which relies only on intellectual property available on a royalty-free basis.

3. WORK IN PROGRESS

This is intended to give you a brief summary of each of the major work items under development by the Multimodal Interaction Working Group. The suite of specifications is known as the W3C Multimodal Interaction Framework.

Introduction, 6 May 2003. The Multimodal Interaction Framework introduces a general framework for multimodal interaction, and the kinds of markup languages being considered.

Use cases, 4 December 2002. Multimodal Interaction Use Cases describes several use cases that are helping us to better understand the requirements for multimodal interaction.

Core requirements, 8 January 2003. Multimodal Interaction Requirements describes fundamental requirements for the specifications under development in the W3C Multimodal Interaction Activity.

3.1 Extensible Multi-Modal Annotations (EMMA)

Requirements, 13 January 2003
Working Draft, 18 December 2003
Last Call Working Draft, Spring 2004

EMMA is being developed as a data exchange format for the interface between input processors and interaction management systems. It will define the means for recognizers to annotate application specific data with information such as confidence scores, time stamps, input mode (e.g. key strokes, speech or pen), alternative recognition hypotheses, and partial recognition results etc. EMMA is a target data format for the semantic interpretation specification being developed in the Voice Browser Activity, and which describes annotations to speech grammars for extracting application specific data as a result of speech recognition. EMMA supercedes earlier work on the natural language semantics markup language in the Voice Browser Activity.

3.2 Modality Interfaces

W3C Note expected June 2004

A common framework based upon the W3C Document Object Model (DOM) for the abstract software interfaces between user interface components for different modalities and the host environment provided by the interaction manager. The Voice Browser Working Group is expected to develop modality interfaces for Speech and DTMF. The Multimodal Interaction Working Group is tasked with defining interfaces for ink and keystrokes, enabling the use of grammars for constrained input, and the context sensitive binding of gestures to semantics. To facilitate secure end-user authentication, the framework should support the integration of bio-metric interfaces for voice, fingerprint and handwriting, etc.

3.3 Building a shared understanding of interaction management

W3C Note expected September 2004
A study of approaches to interaction management for multimodal applications from a practical and theoretical perspective, looking at standalone and distributed solutions, and at different levels of abstraction. The study will identify promising approaches for further work on standardization in collaboration with other W3C working groups.

3.4 Integration of Composite Multimodal Input

W3C Note expected June 2004
Defining the basis for an interoperable treatment of composite multimodal input, for instance, a combination of speech and pen gestures. A report is in preparation on uses cases and a range of potential approaches.

3.5 System and Environment

First Working Draft expected June 2004
A framework for enabling applications to dynamically adapt to match the current device capabilities, device configuration, user preferences and environmental conditions, such as low battery alerts or loss of network connectivity. Other possible changes include muting the microphone and disabling audio output. Dynamic configurations include snapping a camera attachment onto a cell phone or bringing devices together with Bluetooth, e.g. a camera phone and a color printer. This work is being done in collaboration with the W3C Device Independence activity.

3.6 Sessions

This work is being integrated into other deliverables.
Dynamic configurations and distributed multimodal applications present new challenges to Web developers. Sessions provide the basis for subscribing to events, synchronizing data, and hiding details of protocols and addressing mechanisms.

3.7 InkML - an XML language for ink traces

Requirements, 22 January 2003
Working Draft, 23 February 2004
Last Call Working Draft, TBD

This work item sets out to define an XML data exchange format for ink entered with an electronic pen or stylus as part of a multimodal system. This will enable the capture and server-side processing of handwriting, gestures, drawings, and specific notations for mathematics, music, chemistry and other fields, as well as supporting further research on this processing. The Ink subgroup maintains a separate public page devoted to W3C's work on pen and stylus input.

4. CONCLUSION

W3C has been active in the area of multimodal Web access for several years. Very likely, multimodal Web access will be the first widespread practical use of multimodal technology, and will have a similar impact on the adoption multimodal technology as the original Web had on the adoption of Internet technology. Industry interest in use of multimodal Web technology is increasing, and the first key specifications are being put in place at the W3C.

Acknowledgement: This work was supported by the EU FP6 project MWeb. The views are those of the authors, and do not represent the views of the funding agency.