Chapter XXII
Standards for Multimodal Interaction

Deborah A. Dahl
Conversational Technologies, USA

ABSTRACT

This chapter discusses a wide variety of current and emerging standards that support multimodal applications, including standards for architecture and communication, application definition, the user interface, and certifications. It focuses on standards for voice and GUI interaction. Some of the major standards discussed include the W3C multimodal architecture, VoiceXML, SCXML, EMMA, and speech grammar standards. The chapter concludes with a description of how the standards participate in a multimodal application and some future directions.

INTRODUCTION

This chapter will discuss the topic of standards for multimodal interaction. We begin by briefly defining several different categories of standards and talk about the reasons that standards are especially helpful in multimodal systems. There will also be a short discussion of the role of tools in standards and a discussion of requirements for multimodal standards, such as support for distributed systems, extensibility to new modalities, and the ability to support a variety of platforms, such as different types of mobile devices.

Most of the chapter will be focused on specific standards for multimodal interaction and their relationship to each other. We conclude with a few examples of present and future standards-based multimodal systems.
BACKGROUND

Essentially a standard represents an agreement among a community about the meaning of a term or on a way of doing things. In some cases standards are arbitrary, for example, which side of the road you drive on, and in some cases standards represent an agreement on best practices. A standard might be enforced legally, such as building codes, or food safety, or it might just be an agreement within an industry on how to do things. In this section we present a classification of the different types of standards which will lay the groundwork for the discussion of specific standards in the later sections.

Architecture: Components and Communication

Architectural standards define the overall organization of a system, its components, their functions, and how they communicate. In the context of multimodality, we will discuss the World Wide Web Consortium’s Multimodal Architecture and Interfaces standard (Barnett, Dahl et al., 2008) and the older DARPA Communicator standard (Bayer, 2005) and describe their commonalities and differences.

Architectural standards describe how functions are allocated among specific hardware/software components and how they communicate. The goal of architectural standards is to ensure interoperability of components, even if they are developed completely independently. The World Wide Web is an excellent example of an architecture which supports independent servers, clients and applications with a high level of interoperability.

Carefully-defined communication standards are critically important if components developed by different organizations are to interoperate. Communication takes place at several levels. The underlying protocols, such as TCP/IP and HTTP, will not be discussed here since they are not specific to multimodal systems, but do need to be referenced in multimodal standards in order to insure interoperability. Higher level communication protocols specific to multimodal systems, which we will discuss here, include the high level multimodal interaction life cycle events defined in (Barnett, Dahl et al., 2008) as well as standards defining the format of data payloads for the representation of user input. These include the Extensible MultiModal Annotation (EMMA) specification (Johnston et al., 2007) and InkML (Chee, Froumentin, & Watt, 2006) for representing stylus traces. We will also discuss the Media Resources Control Protocol (MRCP) (Shanmugham & Burnett, 2008). MRCP controls speech media servers which perform the functions of speech recognition, speech synthesis, and speaker recognition. Finally, we discuss some biometric standards, such as the BioAPI, being developed by the BioAPI Consortium.

Application Definition

The next type of standard is a standard language for defining an application, especially through the use of standardized markup. The goal of markup standards is to make applications easier to build by:

- Abstracting away as many procedural details of the application as possible so that the developer can concentrate on functionality
- Supporting reuse

The best-known standardized markup, which defines GUI web applications, is HTML. XML (W3C, 2000), which might be considered a metamarkup, is used as a tool to define other modality-specific markups. This chapter assumes some basic knowledge of XML. We will not discuss these well-known formats in detail. In the specific area of multimodal interaction, we will discuss the following: