Journal of Visual Languages and Computing 31 (2015) 150–159

Contents lists available at ScienceDirect

Journal of Visual Languages and Computing

journal homepage: www.elsevier.com/locate/jvlc

Discovery and registration of components in multimodal systems distributed on the IoT☆

B. Helena Rodriguez [a], Jean-Claude Moissinac [b]

[a] W3C's MMI Working Group Editor for the Discovery and Registration Activity, France
[b] Institut Mines-Telecom, Telecom ParisTech CNRS LTCI 46, rue Barrault 75634, Paris Cedex 13, France

ARTICLE INFO

Available online 24 October 2015

Keywords:
W3C
Multimodal interaction
Semantic interaction
Interface services
MMI architecture and interfaces
Pervasive computing
Context awareness

ABSTRACT

One of the major gaps in the current HTML5 web platform, is the lack of interoperable means for a multimodal application to discover services and applications available in a given space and network for example, in smart houses or in applications for the Internet of Things. To address this gap, we produced a SOA approach that was submitted to the W3C's Multimodal Working Group and validated as a draft recommendation that aims to allow the discovery and registration of components used in multimodal systems for the web of things. In this approach, the components are described and virtualized in a module communicating through two dedicated events, they are registered in a Resources Manager to facilitate the fine management of concurrent multimodal interactions, and the system supports an interoperable discovery, registration and filtering of the features provided by heterogeneous and dynamic components in the web of things. In this paper we will present the W3C's Multimodal Architecture and Interfaces standard and explain the Discovery and Registration mechanism adopted to address the need of finding and integrating components into distributed and dynamic multimodal systems.

1. Introduction

The Multimodal Architecture and Interfaces (MMI-Arch) is a current Recommendation of the World Wide Consortium [1] introducing a generic structure and a communication protocol to allow the components in a multimodal system to communicate with each other. It proposes a generic event-driven architecture and a general frame of reference focused in the control of the flow of data messages [2].

This web-oriented frame of reference has been proposed due a growing evolution of distributed approaches for multimodal systems "on the cloud". These approaches are mostly produced in ad-hoc solutions, as shown by a state of the art produced on 2012 and covering 100 relevant multimodal systems where it was observed that more than 97% of the systems had little or no discovery and registration support [3].

At the time and for historical reasons, the W3C's MMI Architecture and Interfaces (MMI-Arch) and its runtime framework [4] were not addressing: (1) the component's discovery and registration to support the fusion (integration) and the fission (composition) mechanisms, (2) the modality component's data model needed to build this registry and (3) the modality component's annotation to facilitate the orchestration (and even the turn-taking) mechanism in a dynamical system.

These three issues are covered and to some extent resolved by a new proposal – mostly SOA oriented, which is now adopted as a W3C's public draft recommendation. [5]

This proposal is based on an interoperable extension for the MMI-Arch's model, designed to support the automation of the discovery, registration and composition of

☆ This paper has been recommended for acceptance by Shi Kho Chang.

E-mail addresses: helena.rodriguez@shopedia.fr (B.H. Rodriguez), moissinac@telecom-paristech.fr (J.-C. Moissinac).

http://dx.doi.org/10.1016/j.jvlc.2015.10.007
1045-926X/

multimodal semantic services. It is also designed to fulfill the requirements of a high-level Quality of Service (QoS) like: the accurate selection of components when these are not available anymore, do not meet the expected functionality or disrupt the context of use.

With these goals in mind, the contribution was structured on three parts: (1) a new addressing method needed for the component's announcement at bootstrapping; (2) an architectural extension in order to support the handling of the state of the multimodal system using a virtual component approach for registration and (3) two new events for the messaging mechanism, to address the requirements of discovery and registration in distributed systems.

These three parts will be completed by the creation of a common and interoperable vocabulary of states and generic features to allow the gross and generic discovery of modalities in large-networks over a concrete networking layer.

In the following sections we will present this standardization work as follows: In Section 2 we will give an overview of the problem, followed by a study of the related work in Section 3 we describe the work on Discovery and Registration and finally, In Section 4 we present a conclusion and some perspectives.

2. Problem statement

Historically, multimodal systems were implemented in stable and well-known environments. Their complexity demanded laboratory-like implementations and very few experiences were developed for real-time contexts or component distribution. But this situation has evolved.

The web developer's community is progressively confronted with the need of modality integration and this challenge is expected to be huge in the years to come, when large-scale networks like the Internet of Things will attain a state of maturity.

The increasing amount of user-produced and collected data will also require a more dynamic software behavior with a more dynamic and adequate approach (1) to handle the user's technical environment in a context where the demand for energy supply is getting higher and higher, and (2) to encourage and improve the efficiency in consumption boosting the creation of systems compatible with smart-grid technologies.

For example, in Japan [6] (as in European countries) the distributed applications will play a very strategic role in the reduction of energy consumption, helping to evolve to an on-demand model. With this goal, the sustainable consumption in houses must be handled and analyzed distantly, using data collected by multimodal applications installed on multiple kinds of devices of the Internet of Things. These applications must interact in a coordinated manner in order to improve the energetic efficiency of the application behavior, to collaborate in the home automation management and in some cases, even the user profiling; the whole in a distributed platform.

Thus, we face the raising of multiple issues, concerning the multimodal user interaction with very heterogeneous types of devices (some of them with low resources), protocols and messaging mechanisms to be synchronized in an interoperable way.

In this context, on one side, modality discovery and selection for distributed applications becomes a new working horizon giving new challenges for multimodal systems, the user-centric design and the web research. And in the other side, generic and interoperable approaches, using web technologies but capable of going beyond the browser model, will be unavoidable.

2.1. Multimodal discovery and registration

Multimodal systems are computer systems endowed with rich capabilities for human–machine interaction and able to interpret information from various communication modes. According to [7] the three principal features of multimodal systems are:

(1) the fusion of different types of data; (2) real-time processing and temporal constraints imposed on information processing; and (3) the fission of restituted data: a process for realizing an abstract message through output on some combination of the available channels.

On these systems, modality management is mostly of the time hard-coded, leaving aside the problem of a generic architecture respond to extensibility issues and the need of discovery, monitoring and coordination of modalities in real-time with context-awareness. Consequently, multimodal applications are manually composed by developers and shared via web APIs and embedded web technologies, in an ad-hoc and proprietary way.

To address this lack of a generic approach, the MMI Architecture proposes an architectural pattern for any system communicating with the user through different modalities simultaneously instantiated in the same interaction cycle. In this unique context of interaction the final user can dynamically switch modalities.

This kind of bi-directional system combines inputs and outputs of multiple sensorial modes and modalities (e.g. voice, gesture, handwriting, biometrics capture, temperature sensing, etc.) and can be used to identify the meaning of the user's behavior or to compose intelligently a more adapted, relevant and pertinent message. Other important characteristic of the Multimodal Architecture and Interfaces specification is that it uses the MVC design pattern generalizing the View to the broader context of the multimodal interaction presentation, where the information can be rendered in a combination of various modalities.

Thus, the MMI recommendation distinguishes (Fig. 1) three types of components: the Interaction Manager, the Data Component and the Modality Components, covering the forms of representing information in a known and recognizable rendered structure.

For example, acoustic data can be expressed as a musical sound modality (e.g. a human singing) or as a speech modality (e.g. a human talking). The component representing the presentation layer in the MMI Architecture is indeed, a Modality Component. This is a logic entity that handles the input and output of different hardware devices (e.g. microphone, graphic tablet, keyboard) or software services (e.g. motion detection, biometrics sensing).