

Adaptivity for Multimodal User Interfaces in Mobile Situations

R. Kernchen, K. Mößner, R. Tafazolli
CCSR – The University of Surrey, Guildford, GU2 7XH, UK
{r.kernchen, k.moessner, r.tafazolli}@surrey.ac.uk

Abstract

Multimodality is a fact of human communication, increasingly our ways to communicate change and humans do interact with machines (be it a mundane ATM transaction, the calling of an automatised call center, or the setting/disarming of a residential alarm system). However, these interactions are mostly limited to single input and output schemes, thus loosing a lot of additional information a human communication partner would sense. Multimodality was perceived to exactly tackle this point. This paper describes a framework and approach to operate multimodal interaction mechanisms in both the fixed as well as the mobile environments. The paper describes a scheme that facilitates the dynamic binding and release of user-interface devices (such as screens, keyboards, etc.) to support multimodal interactions in mobile environments and to enable the user to ‘make use’ of any possible user interface device available (and allowed), thus supporting the individuals changing communication environment. The principles and basic functionality of an Adaptive Multimodal Human Interface-Device Binding Engine are outlined.

1. Introduction

Communication between individuals (as well as within groups) takes place using more than one level of interaction/information exchange. Use of “voice only” or “text only” communication facilities has removed a lot of additional information from a communication. This was true until the introduction of multimedia communication systems, where more than one communication means (e.g. voice plus video) facilitated a more multi layered communication between users. Yet, communication is not only restricted to take place between humans only, but information exchange increasingly happens between humans and machines/systems. In the early nineties, the concept of Multimodal interactions came into the spotlight of research. The term was then defined as: “*Multimodality is the capacity of the system to*

communicate with a user along different types of communications channels and to extract and convey meaning automatically” [1].

While initial multimodal communications was restricted to fixed terminals only, examples for first approaches to multimodal functionality was the introduction of a pointing device (i.e. the mouse), the current tendency indicates that multimodality is even more useful in wireless and mobile environments (e.g. speech output of text messages in in-car situations, translation of text into sign language for hearing impaired users, etc). Many approaches and designs have already been published and some products have already reached the wireless market place. Yet, one of the main limitations of current multimodal devices is their rigid functionality; once designed, the number and type of modalities that a device can serve remains fixed. Multimodal systems can be extended e.g. by extending the application (i.e. to accept and interpret information from additional modality types) and binding additional user interface devices into the systems, e.g. user interfaces like cameras to track movements, gestures and mimic were added to desktop machines.

However, such extensions usually require a complete system redesign and are not suitable to cater for the flexibility required in ever changing mobile environments. And taking into account that Multimodality becomes even more important in an increasingly sensorised environment, where it is expected to become a main facilitator for interactions between the communication environment/system and user.

This process of slow changes to the user face will be insufficient in a mobile environment where the number and type of possible available user-interface devices changes over time (i.e. when moving in a room with a large display or audio system). To really make use of the advantages multimodal interactions can introduce to mobile environments, it has to be made possible that any available user-interface devices can be bound into a system as they become accessible. There are many technical challenges to be overcome,

including the definition of adaptive fusion or fission mechanisms or the interface device binding mechanisms, or the definition of a mobile-multimodal usage cycle, these and related issues are raised in this paper.

The following section will highlight the problem area, followed by an overview on the principles and functions necessary for the binding mechanism and an initial description of the multimodal cycle/procedures. The final section will draw some conclusions.

2. Challenges for User Interfaces in Mobile Environments

Mobile environments together with applications for handicapped users are the areas where multimodality and the capability for multimodal interactions will find their biggest application. But before this can be achieved a number of hurdles have to be overcome, the question of how Multimodal Interfaces can and need to be supported in mobile environments has to be answered and the necessary technologies and approaches have to be investigated and developed. Concurrent use of multiple modalities in mobile and wireless environments will provide a more natural interaction mechanism to the user (much more than a single modality could provide), however, even when in combination (i.e. multiple different interfaces are being used) there are some basic requirements towards any human-machine interface, they can be summarized as:

- The support platform for the (multimodal) user interface needs to facilitate combinational/ adaptive device and modality usage (e.g. gesture, speech, pointing, etc. [1],[2]).
- Support for intelligent intuitive interaction through several input modalities (i.e. this also reduces the Human Computer Interaction (HCI) communication overhead [1]).
- Multimodal interfaces in mobile environments need to take, not only the features of the application but also information about the whole environmental context (e.g. location, time, device information, situation, etc.) into account (reduces possible user interaction regarding the environment).
- The support mechanisms need to be able to 'predict/guess' user preferences and user (interface usage) behaviour, to efficiently make use of the available interfaces etc.

Mobility provides the opportunity (for the user) to access those (user) interface devices that may be available in a local area. Yet, beside the advantages

such possibility offers, there are also associated (technical as well as economical) problems. One of the main difficulties when using multimodality in mobile environments is the high volatility of the configurations (due to terminal movements as well as to communication channel effects). This paper describes an approach implementing an architectural framework that facilitates environment dependent adaptation of multimodal user interfaces. The approach considers the movement of a central communication device within an environment where one to many user interface devices can be wirelessly (on the fly) be connected to extend the devices (and applications) user interface capabilities. In this way, our approach provides a solution which enables the always best suited human machine interaction.

3. Enabling User Interface Adaptivity

Considering the requirements for a system that supports multimodality in mobile environments and situations, an architecture has been conceived which supports the dynamic discovery, negotiation and binding of wirelessly accessible user interface devices. Furthermore, the scheme is foreseen to enable, based on user interface context and application capabilities and requirements, to facilitate the adaptation of the application to make best possible use of the available user interfaces.

The framework developed is called "Adaptive Multimodal Human Interface Device Binding Engine (AMHI-DBE)". The engine offers the support functionality needed to implement the aforementioned adaptivity for multimodal applications.

One of the main parts of the architecture is the Multimodal Human Interface (MHI), it consists of one portal device (PD – each user is equipped with one of these) and one to several human-interface devices (HID), as indicatively shown in Figure 1. The PD is always available to the user and provides the core-platform for multimodal applications. Furthermore, it has the task to discover potential HIDs in the local neighborhood. After completion of the discovery phase, available HI Devices and the PD may define the possible multimodal interface that potentially can be implemented to facilitate more convenient user interaction.

All available (location dependent) interface devices may be applied in any combination, as required by the multimodal application. This can include that a device may provide a number of human in- and output modalities, depending on the number of provided service interfaces of current available HIDs.

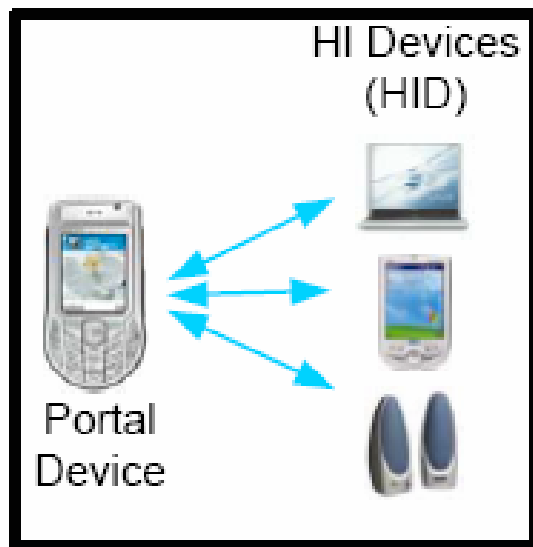


Figure 1. The Multimodal Human Interface (MHI) in Mobile Situations

These dynamic mechanisms to bind and use user interface devices and their (interface) services offered is expected to facilitate the adaptivity needed for multimodality in mobile situations.

4. Mechanisms and Sequences

User interface devices can (and should) be used in combination to build a multimodal user interface. Based on this are our model of required system modules (modalities) and their interactions, together implementing the MHI, illustrated in Figure 2. The portal device executes the multimodal application as well as dynamically assigned components of the Adaptive MHI Device Binding Engine (AMHI-DBE). Within the application server, multimodal applications use a platform to send application requests to the AMHI-DBE (MM Application requests) or receive information about available HI Devices (HI Device/Service Information). “Heavy” components of the AMHI-DBE are executed in the MUI Support Server in order to take some of the load of the Portal Device, since most state of the art mobile devices do not provide sufficient processing power.

Figure 2 – provides an overview of the system, including the main modules and main features. Within this architecture, the AMHI-DBE is responsible for all tasks regarding interaction between the modalities. The basic interactions with the HI Devices include HI Device/Service Discovery and Multimodal Content

Delivery (media streams). Beyond these basic tasks, the AMHI-DBE has to provide a number of additional functions, they are distributed over a number of functional blocks, detailed descriptions can be found in [3], the focus of this paper is on the overall procedure, the device and service discovery and the device binding mechanism.

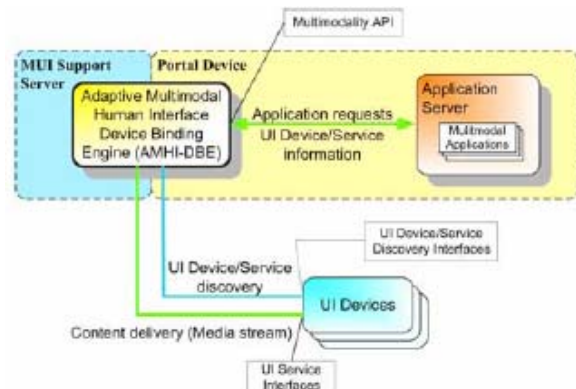


Figure 2. System Modules - Multimodal Interfaces Support in Mobile Situations

4.1. Interface Usage Sequence

In contrast to fixed or predefined Multimodal user interfaces, the availability of user interface devices can (and will change) during a service/application session (when in mobile situations). Hence the discovery, dynamic binding, usage and release of interface devices will be a continuous cycle of events (as depicted in Figure 3). This cycle starts with the discovery of modality bearing devices within the ‘usable zone’ (i.e. the environment in which a device is useful for user and application). Once the modalities are known and availability of the device is ensured, the Modality Fusion/Fission features are to be established. Based on this and some additional context information (i.e. connectivity, mobility type, predicted duration of service/application session, etc), the portal device will need to establish the context which in turn feeds into the decision making process of whether or not a device binding should take place.

After making a decision to bind a discovered device and to use the modality offered as part of the application, the device becomes ‘bound’ into the overall Multimodal user interface.. The AMHI-DBE then has to implement any content adaptation/transformation functions necessary (i.e. transcoding from one to another video format, etc.). Finally, when all steps are completed, the actual content delivery using the extended Multimodal interface can take place.

As aforementioned, since the portal devices environment may change (due to mobility) the portal device will need to keep on evaluating the availability of additional interface devices. And whenever necessary, will have to repeat the cycle to bind new interfaces (i.e. the procedure for termination of a binding is currently being developed and will be documented later).

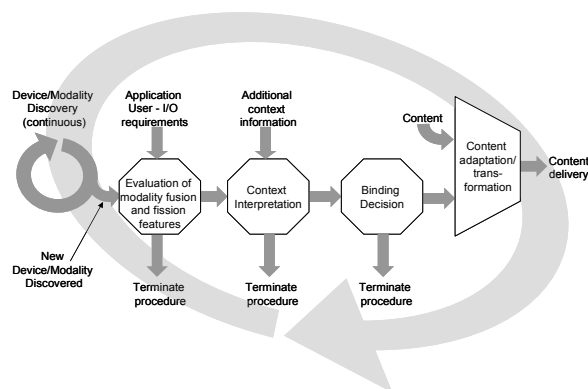


Figure 3. Multimodality Cycle in Mobile Situations

4.2. Device and Modality Service Discovery

User Interface Devices which are used during Multimodal communication sessions can be assumed to be subject to fluctuation of HI Device configurations. To minimise the configuration overhead, automatic device and service discovery [4] processes can be employed to ensure that the possible user in their current location can interact with all available HI Devices. First, the discovery procedures acquire information about the device specific capabilities (e.g. screen size, audio card), this is followed by an exchange of information about the available service interfaces for support of different modalities (e.g. audio or video output, speech input, etc.). The device/service information obtained is stored in corresponding 'device and service profiles' ready to be used by other functional blocks or directly by the application.

4.3. Device and Modality Binding

The AMHI-DBE's foundation is a rule based multimodal device binding decision block. Any device binding decisions will be made considering the outcomes of the semantic modality fusion model, the context interpretation model and device capabilities (see [3] for more detail). Evaluation of these three

models leads to a comprehensive content-to-device delivery model. This delivery model describes the path and modality via which interface device the content should be delivered to the user (and vice versa, which device should be used to collect content that was originated by the user). If an interface-device does not provide the required capabilities or the necessary service, then the content adaptation/ transformation block can transform the content into required content format.

5. Conclusions

The paper outlined and initially described the dynamic formation of user interfaces to provide multi-device based multimodal user interfaces in mobile situations. The concepts and mechanisms introduced will facilitate the dynamic discovery, binding and usage of user interface devices in situations where a user may roam through environments where different interface devices (e.g. screen, speaker, camera) are available and can be (temporarily) used by the user.

Although there are considerable threats and issues to be tackled, the mechanisms have been described concentrating on the feasibility of the concept rather than the security and privacy implications.

Acknowledgement

This work has been performed in the framework of the EU funded project (MobiLife). The authors would like to acknowledge the contributions of their colleagues from MobiLife consortium.

References

- [1] Nigay L., Coutaz J., "A Design Space for Multimodal Systems: Concurrent Processing and Data Fusion", *INTERCHI'93 Proceedings*, ACM Press, Amsterdam, May, 1993, pp. 172-178.
- [2] Oviatt S.L., "Multimodal interfaces", In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, J. JACKO AND A. SEARS, Eds. Lawrence Erlbaum Assoc., Mahwah, NJ, 2003, chap.14, 286-304
- [3] Moessner K., Kernchen R., Tafazolli R., Adaptive Multimodal Human Interface-Device Binding Engine (AMHI-DBE), *Proceedings of the 12th Meeting of the ireless World Research Forum*, Toronto, Canada, 4/5. November 2004.
- [4] Golden G. Richard III, "Service and Device Discovery", New York: McGraw-Hill, 2002.