Multimodal Dialogue for Ambient Intelligence and Smart Environments

Ramón López-Cózar and Zoraida Callejas

1 Introduction

Ambient Intelligence (AmI) and Smart Environments (SmE) are based on three foundations: ubiquitous computing, ubiquitous communication and intelligent adaptive interfaces [41]. This type of systems consists of a series of interconnected computing and sensing devices which surround the user pervasively in his environment and are invisible to him, providing a service that is dynamically adapted to the interaction context, so that users can naturally interact with the system and thus perceive it as intelligent.

To ensure such a natural and intelligent interaction, it is necessary to provide an effective, easy, safe and transparent interaction between the user and the system. With this objective, as an attempt to enhance and ease human-to-computer interaction, in the last years there has been an increasing interest in simulating human-to-human communication, employing the so-called multimodal dialogue systems [46]. These systems go beyond both the desktop metaphor and the traditional speech-only interfaces by incorporating several communication modalities, such as speech, gaze, gestures or facial expressions.

Multimodal dialogue systems offer several advantages. Firstly, they can make use of automatic recognition techniques to sense the environment allowing the user to employ different input modalities, some of these technologies are automatic speech recognition [62], natural language processing [12], face location and tracking [77], gaze tracking [58], lipreading recognition [13], gesture recognition [39], and handwriting recognition [78].
Secondly, these systems typically employ several output modalities to interact with the user, which allows to stimulate several of his senses simultaneously, and thus enhance the understanding of the messages generated by the system. These modalities are implemented using technologies such as graphic generation, natural language generation [33], speech synthesis [17], sound generation [24], and tactile/haptic generation [70].

Thirdly, the combination of modalities in the input and output allows to obtain more meaningful and reliable interpretations of the interaction context. This is, on the one hand, because complementary input modalities provide with non-redundant information which helps creating a richer model of the interaction. On the other hand, redundant input modalities increase the accuracy and reduce the uncertainty of the information [11]. Besides, both the system and the user can choose the adequate interaction modalities to carry out the communication, thus enabling a better adaptation to environmental conditions such as light/acoustic conditions or privacy. Furthermore, the possibility to choose alternative ways of providing and receiving information allows disabled people to communicate with this type of system using the interaction modalities that best suit their needs.

Researchers have developed multimodal dialogue systems for a number of applications, for example, interaction with mobile robots [44] and information retrieval [26]. These systems have also been applied to the main topic of this book, for example, to enhance the user-system interaction in homes [54, 23], academic centers [47], hospitals [8] and theme parks [56].

After this brief introduction, the remaining of this chapter is organised as follows. Section 2 deals with context awareness and user modelling. Section 3 addresses the handling of input and contextual information and how to build abstractions on the environment to acquire it. Section 4 centres on dialogue management, i.e. on how the system responds to inputs and to changes in the environment, focusing on interaction and confirmation strategies. Section 5 addresses response generation, discussing the fission of multimodal information by means of synthesised speech, sounds, graphical objects and animated agents. Section 6 addresses system evaluation, describing the peculiarities of the evaluation of multimodal interfaces for AmI and SmE applications. Section 7 presents the conclusions.

2 Context Awareness

Although there is not a complete agreement on the definition of context, the most widely accepted is the one proposed by [15]: “Any information that can be used to characterize the situation of an entity (...) relevant to the interaction between a user and an application, including the user and the application themselves”. As can be observed from this definition, any information source can be considered context as long as it provides knowledge relevant to handle the communication between the user and the system. In addition, the user is also considered to be part of the contextual information.