Chapter 8
Studying Natural Interaction in Multimodal, Multi-Surface, Multiuser Scenarios

Carlos Duarte
University of Lisbon, Portugal

Andreia Ribeiro
University of Lisbon, Portugal

Rafael Nunes
University of Lisbon, Portugal

ABSTRACT

Current technological apparati have made it possible for natural input systems to reach our homes, businesses, and learning sites. However, and despite some of these systems being already commercialized, there is still a pressing need to better understand how people interact with these apparati, given the whole array of intervening contextual factors. This chapter presents two studies of how people interact with systems supporting gesture and speech on different interaction surfaces: one supporting touch, the other pointing. The naturally occurring commands for both modalities and both surfaces have been identified in these studies. Furthermore, the studies show how surfaces are used, and which modalities are employed based on factors such as the number of people collaborating in the tasks and the placement of appearing objects in the system, thus contributing to the future design of such systems.

INTRODUCTION

Natural interaction technologies are becoming widely available nowadays. Gesture recognition is available in tablets and smartphones, but also in entertainment systems like Microsoft's Kinect. Speech recognition, available in some desktop operating systems for some time now, is also becoming mainstream in more recent smartphones. Traditionally, the major driving force for the development work in these systems has been the field of entertainment (Silva & Bowman, 2009). However, these technologies are reaching a maturing point that makes them available and desirable for other fields (Hinrichs & Carpendale, 2011; Morris, Wobbrock, & Wilson, 2010; Voida, Tobiasz, Stromer, Isenberg, & Carpendale, 2009).

To further strengthen this adoption process, a continuous process of understanding user interac-

DOI: 10.4018/978-1-4666-4623-0.ch008

tion in these scenarios is required. This chapter addresses this need by adopting the following complementary approaches. First, by exposing users to an interactive prototype, without imposed command languages or other technical restrictions, it is possible to elicit the language users naturally employ in different modalities when interacting with the system. Second, by deploying a prototype based on the elicited commands, it is possible to study natural interaction in scenarios characterized by diverse surfaces and number of users.

In this chapter, we present two studies aiming to increase the knowledge regarding gestural and speech interaction in scenarios with interactive surfaces. In the first study, a single surface out of the participants' reach is used. In the second study, we added an interactive touch surface to the initial set-up.

These studies focused initially on understanding how the characteristics of the performed action influenced the way modalities were used. This was followed by analyzing what is the impact on the usage of different surfaces and different modalities of the number of people interacting with the system, and of the application behavior when it comes to rendering new objects.

This chapter is organized in the following way. It starts by reviewing relevant work related to gesture and speech interaction, and to previous efforts on finding how users interact with this technology. It then briefly motivates to the use of the aforementioned two stages approach. The following section presents the first study, conducted with only a single surface. The ensuing section describes the second study, with two surfaces with different characteristics. In each of these sections, the findings and results from the two stages are presented and discussed. The last two sections consider future directions for this technology, and conclude the chapter proposing some design recommendations on the use of multiple modalities and multiple surfaces.

BACKGROUND

Gestural interaction is becoming pervasive. It can be found in tablets and smartphones, who offer their users touch based interfaces, supporting direct manipulation and semaphoric gestures (Quek et al., 2002). Microsoft Kinect and other entertainment systems support deictic and semaphoric gestures also. While people interact naturally with each other through gestures, gesture dictionaries are still required for HCI. This has been acknowledged in several works that tried to understand how people interact with computers through gestures (Kurdyukova, Redlin, & André, 2012; Miki, Miyajima, Nishino, Kitaoka, & Takeda, 2008; Wobbrock, Morris, & Wilson, 2009; Yin & Davis, 2010).

While some gestures have become standard for performing actions (e.g. pinch for zooming), there is still a need to characterize the way people perform general actions in a computing environment (Dang, Straub, & André, 2009; Epps, Lichman, & Wu, 2006; Neca & Duarte, 2011). These studies show that people not only present variability in the gestures they make for each command, but also in how they make it (e.g. by using different hand posture). This impacts the interaction design of applications that want to make use of gestures, but also the way gesture recognizers need to perform. Two of the major problems identified are: (1) people perform the same gesture for different actions; and (2) people find it very difficult to come up with gestures for actions that can not be addressed through direct manipulation (e.g. deleting an object in an interactive space without a recycle bin to drop the object into).

One way to address these problems is to combine gestural with speech interaction (Bourguet & Ando, 1998; Neca & Duarte, 2011; Tse, Greenberg, & Shen, 2006). When speech is available as an input modality, the way gestures are used changes, assuming more of a supporting role, providing arguments for the action specified in the speech command (e.g. pointing at an object