Chapter X
Speech and Gaze Control for Desktop Environments

Emiliano Castellina
Politecnico di Torino, Italy

Fulvio Corno
Politecnico di Torino, Italy

Paolo Pellegrino
Politecnico di Torino, Italy

ABSTRACT
This chapter illustrates a multimodal system based on the integration of speech- and gaze-based inputs for interaction with a real desktop environment. In this system, multimodal interactions aim at overcoming the intrinsic limit of each input channel taken alone. The chapter introduces the main eye tracking and speech recognition technologies, and describes a multimodal system that integrates the two input channels by generating a real-time vocal grammar based on gaze-driven contextual information. The proposed approach shows how the combined used of auditive and visual clues actually permits to achieve mutual disambiguation in the interaction with a real desktop environment. As a result, the system enables the use of low cost audio-visual devices for every day tasks even when traditional pointing devices, such as a keyboard or a mouse, are unsuitable for use with a personal computer.

INTRODUCTION
In various situations people need to interact with a Personal Computer without having the possibility to use traditional pointing devices, such as a keyboard or a mouse. People may need both hands free to fulfill other tasks (such as driving or operating some equipment), or may be sub-
Speech and Gaze Control for Desktop Environments

In general, most interaction channels, taken alone, are inherently ambiguous, and far-from-intuitive interfaces are usually necessary to deal with this issue (Oviatt, 1999). To keep the interaction simple and efficient, multimodal interfaces have been proposed, which try to exploit the peculiar advantages of each input technique, while compensating for their disadvantages. Among these, gaze-, gesture- and speech-based approaches are considered the most natural, especially for people with disabilities.

Particularly, unobtrusive techniques are the most preferred, as they aim at enhancing the interaction experience in the most transparent way, avoiding the introduction of wearable “gadgets” which often make the user uncomfortable. Unfortunately, this is still a strong technical constraint, which must be softened by some necessary trade-offs. For instance, while speech-recognition may simply require wearing a microphone, eye tracking is usually constrained to using some fixed reference point (e.g., either a head mounted or wall mounted camera), making it suitable only for applications in limited areas. Additionally, the environmental conditions render eye tracking unusable with current mobile devices, which are instead more appropriate for multimodal gesture- and speech-based interaction.

Indeed, the ambient conditions play a major role when choosing the technologies to use and the strategies to adopt, always taking into account the final cost of the proposed solution.

In this context, the chapter discusses a gaze- and speech-based approach for the interaction with the existing GUI widgets provided by the Operating System. While various studies (see the Related Works section) already explored the possibility of integrating gaze and speech information in laboratory experiments, we aim at extending those results to realistic desktop environments. The statistical characteristics (size, labels, mutual distance, overlapping, commands, …) of the widgets in a modern GUI are extremely different from physical impairment, either temporary or permanent.

In recent years, various alternatives to classical input devices like keyboard and mouse and novel interaction paradigms have been proposed and experimented. Haptic devices, head-mounted devices, and open-space virtual environments are just a few examples. With these futurist technologies, although still far from being perfect, people may receive immediate feedback from a remote computing unit while manipulating common objects. Special cameras, usually mounted on special glasses, allow to track either the eye or the environment, and provide visual hints and remote control of the objects in the surrounding space. More intrusive approaches exploit contact lenses with micro sensors to grasp information about the user’s gaze (Reulen & Bakker, 1982). In other approaches special gloves are used to interact with the environment through gestures in the space (Thomas & Piekarski, 2002).

In parallel to computer vision-based techniques, voice interaction is also adopted as an alternative or complementary channel for natural human-computer interaction, allowing the user to issue voice commands. Speech recognition engines of different complexity are used to identify words from a vocabulary and to interpret users intentions. These functionalities are often integrated by context knowledge to reduce recognition errors or command ambiguity. For instance, several mobile phones currently provide speech-driven composition of a number in the contact list, which is by itself a reduced contextual vocabulary for this application. Information about the possible “valid” commands in the current context is essential for trimming down the vocabulary size and enhance recognition rate. At the same time, the current vocabulary might be inherently ambiguous, as the same command might apply to different objects or the same object might support different commands: also in this case, contextual information may be used to infer user intentions.
Related Content

A Festival-Wide Social Network Using 2D Barcodes, Mobile Phones and Situated Displays
[www.igi-global.com/chapter/festival-wide-social-network/69638?camid=4v1a](www.igi-global.com/chapter/festival-wide-social-network/69638?camid=4v1a)

Information Security and Virtual Teams
[www.igi-global.com/article/information-security-and-virtual-teams/193638?camid=4v1a](www.igi-global.com/article/information-security-and-virtual-teams/193638?camid=4v1a)

Earcons Versus Auditory Icons in Communicating Computing Events: Learning and User Preference

Online Interactions as a Terror Management Mechanism: How Death Anxiety Affects Facebook Use
[www.igi-global.com/article/online-interactions-as-a-terror-management-mechanism/162754?camid=4v1a](www.igi-global.com/article/online-interactions-as-a-terror-management-mechanism/162754?camid=4v1a)