Chapter 1.1
An Overview of Multimodal Interaction Techniques and Applications

Marie-Luce Bourguet
Queen Mary University of London, UK

INTRODUCTION

Desktop multimedia (multimedia personal computers) dates from the early 1970s. At that time, the enabling force behind multimedia was the emergence of the new *digital technologies* in the form of digital text, sound, animation, photography, and, more recently, video. Nowadays, multimedia systems mostly are concerned with the compression and transmission of data over networks, large capacity and miniaturized storage devices, and quality of services; however, what fundamentally characterizes a multimedia application is that it does not understand the data (sound, graphics, video, etc.) that it manipulates. In contrast, intelligent multimedia systems at the crossing of the artificial intelligence and multimedia disciplines gradually have gained the ability to understand, interpret, and generate data with respect to content.

Multimodal interfaces are a class of intelligent multimedia systems that make use of multiple and natural means of communication (modalities), such as speech, handwriting, gestures, and gaze, to support human-machine interaction. More specifically, the term *modality* describes human perception on one of the three following perception channels: visual, auditive, and tactile. Multimodality qualifies interactions that comprise more than one modality on either the input (from the human to the machine) or the output (from the machine to the human) and the use of more than one device on either side (e.g., microphone, camera, display, keyboard, mouse, pen, track ball, data glove). Some of the technologies used for implementing multimodal interaction come from speech processing and computer vision; for example, speech recognition, gaze tracking, recognition of facial expressions and gestures, perception of sounds for localization purposes, lip movement analysis (to improve speech recognition), and integration of speech and gesture information.

In 1980, the put-that-there system (Bolt, 1980) was developed at the Massachusetts Institute of Technology and was one of the first multimodal systems. In this system, users simultaneously could speak and point at a large-screen graphics

Copyright © 2008, IGI Global, distributing in print or electronic forms without written permission of IGI Global is prohibited.

display surface in order to manipulate simple shapes. In the 1990s, multimodal interfaces started to depart from the rather simple speech-and-point paradigm to integrate more powerful modalities such as pen gestures and handwriting input (Vo, 1996) or haptic output. Currently, multimodal interfaces have started to understand 3D hand gestures, body postures, and facial expressions (Ko, 2003), thanks to recent progress in computer vision techniques.

BACKGROUND

In this section, we briefly review the different types of modality combinations, the user benefits brought by multimodality, and multimodal software architectures.

Combinations of Modalities

Multimodality does not consist in the mere juxtaposition of several modalities in the user interface; it enables the synergistic use of different combinations of modalities. Modality combinations can take several forms (e.g., redundancy and complementarity) and fulfill several roles (e.g., disambiguation, support, and modulation).

Two modalities are said to be redundant when they convey the same information. Redundancy is well illustrated by speech and lip movements. The redundancy of signals can be used to increase the accuracy of signal recognition and the overall robustness of the interaction (Duchnowski, 1994).

Two modalities are said to be complementary when each of them conveys only part of a message but their integration results in a complete message. Complementarity allows for increased flexibility and efficiency, because a user can select the modality of communication that is the most appropriate for a given type of information.

Mutual disambiguation occurs when the integration of ambiguous messages results in the resolution of the ambiguity. Let us imagine a user pointing at two overlapped figures on a screen, a circle and a square, while saying "the square." The gesture is ambiguous because of the overlap of the figures, and the speech also may be ambiguous if there is more than one square visible on the screen. However, the integration of these two signals yields a perfectly unambiguous message.

Support describes the role taken by one modality to enhance another modality that is said to be dominant; for example, speech often is accompanied by hand gestures that simply support the speech production and help to smooth the communication process.

Finally, modulation occurs when a message that is conveyed by one modality alters the content of a message conveyed by another modality. A person's facial expression, for example, can greatly alter the meaning of the words he or she pronounces.

User Benefits

It is widely recognized that multimodal interfaces, when carefully designed and implemented, have the potential to greatly improve human-computer interaction, because they can be more intuitive, natural, efficient, and robust.

Flexibility is obtained when users can use the modality of their choice, which presupposes that the different modalities are equivalent (i.e., they can convey the same information). Increased robustness can result from the integration of redundant, complementary, or disambiguating inputs. A good example is that of visual speech recognition, where audio signals and visual signals are combined to increase the accuracy of speech recognition. Naturalness results from the fact that the types of modalities implemented are close to the ones used in human-human communication (i.e., speech, gestures, facial expressions, etc.).

5 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/overview-multimodal-interaction-techniques-applications/24271

Related Content

Distributed Representation of Compositional Structure
Simon D. Levy (2009). *Encyclopedia of Artificial Intelligence (pp. 514-519).*
www.irma-international.org/chapter/distributed-representation-compositional-structure/10295

Predicting Weather Conditions Using Machine Learning for Improving Crop Production
Shabnam Kumariand P. Muthulakshmi (2024). *AI Applications for Business, Medical, and Agricultural Sustainability (pp. 267-302).*
www.irma-international.org/chapter/predicting-weather-conditions-using-machine-learning-for-improving-crop-production/341761

An Optimized Intuitionistic Fuzzy Associative Memories (OIFAM) to Identify the Complications of Type 2 Diabetes Mellitus (T2DM)
Felix A.and Dhivya A. D. (2020). *International Journal of Fuzzy System Applications (pp. 22-41).*
www.irma-international.org/article/an-optimized-intuitionistic-fuzzy-associative-memories-oifam-to-identify-the-complications-of-type-2-diabetes-mellitus-t2dm/253083

Percentile Matching Estimation of Zigzag Uncertainty Distribution
S. Sampathand K. Anjana (2018). *International Journal of Fuzzy System Applications (pp. 56-73).*
www.irma-international.org/article/percentile-matching-estimation-of-zigzag-uncertainty-distribution/195677

Optimization of Queuing Theory Based on Vague Environment
Verónica Andrea González-López, Ramin Gholizadehand Aliakbar M. Shirazi (2016). *International Journal of Fuzzy System Applications (pp. 1-26).*
www.irma-international.org/article/optimization-of-queuing-theory-based-on-vague-environment/144201