

## Preface

Human Computer interaction (HCI) deals with the relationships among people and computers. As the digital world is getting multi-modal, the information space is getting more and more complex. In order to navigate this information space and to capture and apply the implicit and explicit knowledge to appropriate use an effective interaction is required. Such an effective interaction is only possible if computers can understand and respond to important modalities of human perception and cognition, i.e., speech, image and language, including other modalities, e.g., haptic, olfactory and brain signals. HCI researchers have to respond to these challenges by developing innovative concepts, models and techniques. There have been efforts in the areas of language, speech, vision and signal processing. However, the general techniques may not be applicable to or may degrade in HCI environment. This book attempts to bring all relevant technologies (for Language, Speech, Image and other signal processing) at one place in an interaction framework.

The signal processing community and HCI researchers can refer the book to improve their understanding of the state-of-the-art in HCI and broaden their research spheres. It will help postgraduate and doctoral students in identifying new and challenging research problems in the HCI areas.

We invited chapter proposals for *Speech, Image and Language Processing for Human Computer Interaction: Multimodal Advancements* and received around forty proposals. After three rounds of reviews, revisions, suggestions and editing we are herewith fifteen chapters in this book. The book is neither a handbook consisting of numerous research papers nor a textbook describing each aspect of HCI coherently. It is not possible to cover all the aspects of HCI in a single book, especially when HCI is still not a matured area. The 15 chapters in the book provide contribution on models, techniques and applications of HCI and are organized in three sections which are as follows: 1) Modeling Interaction 2) Interaction based on Speech, Image and Language, and 3) Multimodal Developments. We had to include the first section, which describes cognitive user models and underlines the need of a general framework for effective interaction. Chapters contained in the second section present methods of audio, visual and lingual processing for multimodal interaction. The third section introduces issues related to multimodal interfaces and conversational systems. It describes current efforts to develop intelligent, adaptive, proactive, portable and affective multimodal interfaces and discusses mobile vision and health-care applications.

The book opens with a brief summary of different user modeling techniques in HCI in the first chapter. This chapter is useful for system analysts and developers who find it difficult to choose an appropriate user model for their applications. The models discussed in the chapter include GOMS family of models, cognitive models and application specific models. Chapter 2 emphasizes on the separation of specific problem solving skills and problem related knowledge from the general skills and knowledge. In this line of argument the chapter proposes a general three layer architecture of HCI, consisting of the human layer, the interaction layer and the computer layer. The authors categorize existing architectures for HCI based on the use of no cognition, individual cognition and social cognition, highlight their main features

and outline a framework for collaborative HCI. The next chapter (chapter 3) analyzes the cognitive and interactive behavior of users in the collaborative design activities. The chapter presents features of Metaverse and compares it with two other cooperative design environments. The first section ends with chapter 4 presenting a fuzzy logic-based methodology for designing an adaptive user interface in imperfect, vague and multimodal environment.

The second section of the book collects methods for speech, Image and language processing for multimodal interaction. To improve human computer interaction, it is necessary to create immersive audio environments where the user feels like a part of the system. For this, chapter 5 presents a structural model for the pinna for the synthesis of head related transfer functions. Chapter 6 offers a brief review of classical and recent approaches of Markov modeling for speech recognition while chapter 7 gives an overview of spoken dialog systems. In chapter 8, the authors give a broad introduction to audio-visual speech recognition systems and address some of the issues related to visual feature extraction and integration of audio-visual information for speech recognition. Shape analysis and recognition is a well known problem in the areas of HCI and computer vision. Chapter 9 presents a novel idea of using Farey sequence to represent the edge slopes, and the vertex angles to get an efficient description in integer domain. In chapter 10, the authors report their work on gesture recognition by finding fingertip point locations aiming at the development of an 'accessory-free' or 'minimum accessory' interface for communication and computation. Considering the importance of HCI aspects of web search we have a chapter (chapter 11) focusing on effective meta-searching. An interesting extension to text-based messages is emoticons which have been almost overlooked in the area of HCI. The last chapter of this section describes a prototype system, CAO (emotiCon Analysis & decOding), for affect analysis of Eastern style emoticon.

The three chapters contained in the third section of the book discuss the applications and research issues involved in the multimodal development of interfaces and interactive systems. The section begins with an overview of architectures and toolkits for the development of multimodal interface in chapter 13. The revolutionary changes in communication led to the development of vision recognition applications on mobile phones. Chapter 14 identifies a number of such applications. Human Computer Technology has found applications in healthcare domain also. A number of gadgets are being used to monitor health of patients automatically. One such application is covered in the final chapter of the book. This chapter focuses on the use of automatic speech recognition technology for evaluating speech disorder. It reviews existing systems and discusses different types of speech disorders, the main innovations in the field, and the available resources that can be used to develop such systems.

We would like to thank all the authors for their contributions. We would also like to thank the members of the advisory board for extending their support and suggestions. Special thanks go to reviewers for providing thoughtful and valuable comments on the initial and revised chapters. Their comments and suggestions helped in improving the quality of the book.

Finally, we are deeply indebted to our family members for their patience and understanding while we were busy with this book.

We wish you happy reading.

*Uma Shanker Tiwary*  
*Indian Institute of Information Technology Allahabad, India*

*Tanveer J. Siddiqui*  
*University of Allahabad, India*