

Preface

Our interest in human–computer interaction started with our participation in the TT2 project (“Trans–Type-2”, 2002–2005—<http://www.tt2.atosorigin.es>), funded by the European Union (EU) and coordinated by Atos Origin, which dealt with the development of statistical-based technologies for computer assisted translation.

Several years earlier, we had coordinated one of the first EU-funded projects on spoken machine translation (EuTrans, 1996–2000—<http://prhlt.iti.es/w/eutrans>) and, by the time TT2 started, we had already been working for years in machine translation (MT) in general. So we knew very well which was one of the major bottlenecks for the adoption of the MT technology available at that time by professional translation agencies: Many professional translators preferred to type by themselves all the text from scratch, rather than trying to take advantage of the (few) correct words of a MT-produced text, while fixing the (many) translation errors and sloppy sentences. Clearly, by post-editing the error-prone text produced by a MT system, these professionals felt they were not in command of the translation process; instead, they saw themselves just as dumb assistants of a foolish system which was producing flaky results that they had to figure out how to amend (the state of affairs about post-editing has improved over the years but the feeling of lack of control persists).

In TT2 we learnt quite a few facts about the central role of human feedback in the development of assistive technologies and how this feedback can lead to great human/machine performance improvements if it is adequately taken into account in the mathematical formulation under which systems are developed. We also understood very well that, in these technologies, the traditional, accuracy-based performance criteria is not sufficiently adequate and performance has to be mainly assessed in terms of estimated human–machine interaction effort. In one word, assistive technology has to be developed in such a way that the human user feels in command of the system, rather than the other way around, and human-interaction effort reduction must be the fundamental driving force behind system design. In TT2 we also started to realize that multimodal processing is somehow implicitly present in all interactive systems and that this can be advantageously exploited to improve overall system performance and usability.

After the success of TT2, our research group (PRHLT—<http://prhlt.iti.upv.es>), started to look at how these ideas could be applied in many other Pattern Recognition (PR) fields, where assistive technologies are in increasing demand. As a result, we soon found ourselves coordinating a large and ambitious Spanish research program, called Multimodal Interaction in Pattern Recognition and Computer Vision (MIPRCV, 2007–2012—<http://miprcv.iti.upv.es>). This program, which involves more than 100 highly qualified Ph.D. researchers from ten research institutions, aims at developing core assistive technologies for interactive application fields as diverse as language and music processing, medical image recognition, biometrics and surveillance, advanced driving assistance systems and robotics, to name but a few.

To a large extent, this book is the result of works carried out by the PRHLT research group within the MIPRCV consortium. Therefore it owes credit to many MIPRCV researchers that have directly or indirectly contributed with ideas, discussions and technical collaborations in general, as well as to all the members of PRHLT who, in one manner or another, have made it possible.

These works are presented in this book in a unified way, under the PR framework of Statistical Decision Theory. First, fundamental concepts and general PR approaches for Multimodal Interaction modelling and search (or inference) are presented. Then, systems developed on the base of these concepts and approaches are described for several application fields. These include interactive transcription of handwritten and spoken documents, computer assisted language translation, interactive text generation and parsing, and relevance-based image retrieval. Finally, several prototypes developed for these applications are overviewed in the last chapter. Most of these prototypes consist in live demonstrators which can be publicly accessed through the Internet. So, readers of this book can easily try them by themselves in order to get a first-hand idea of the interesting possibilities of placing Pattern Recognition technologies within the Multimodal Interaction framework.

Chapter 1 provides an introduction to Interactive Pattern Recognition, examining the challenges and research opportunities entailed by placing PR within the human-interaction framework. Moreover, it provides an introduction to general approaches available to solve the underlying *interactive search* problems on the basis of existing methods to solve the corresponding non-interactive counterparts and, an overview of modern machine learning approaches which can be useful in the interactive framework.

Chapter 2 establishes the common basics and framework on which are grounded the computer assisted transcription approaches described in the three subsequent *Chaps.*: 3, 4 and 5. On the one hand, *Chaps.* 3 and 5 are devoted to handwritten documents transcription providing different approaches, which cover different aspects as multimodality, user interaction ways and ergonomics, active learning, etc. On the other hand, *Chap. 4* focuses directly on transcription of speech signals employing a similar approach described in *Chap. 3*.

Likewise, *Chap. 6* addresses the general topic of Interactive Machine Translation, providing an adequate human–machine-interactive framework to produce high-quality translation between any pair of languages. It will be shown how this also allows one to take advantage of some available multimodal interfaces to increase the

productivity. Multimodal interfaces and adaptive learning in *Interactive Machine Translation* will be covered in *Chaps. 7* and *8*, respectively.

With significant differences in relation to previous chapters, *Chaps. 9–11* introduce other three Interactive Pattern Recognition topics: *Interactive Parsing*, *Interactive Text Generation* and *Interactive Image Retrieval*. The second one, for example, is characterized by not using input signal, whereas the first and third by not following the *left-to-right* protocol in the analysis of their corresponding inputs.

Finally, *Chap. 12* presents several full working prototypes and demonstrators of multimodal interactive pattern recognition applications. As previously commented, all of these systems serve as validating examples for the approaches that have been proposed and described throughout this book. Among other interesting things, they are designed to enable a true human–computer interaction on selected tasks.

Valencia, Spain

E. Vidal
A.H. Toselli
F. Casacuberta

Multimodal Interactive Pattern Recognition and
Applications

Toselli, A.H.; Vidal, E.; Casacuberta, F.

2011, XVI, 274 p., Hardcover

ISBN: 978-0-85729-478-4