EURASIP Journal on Applied Signal Processing 2004:11, 1617–1618

Editorial

Ryohei Nakatsu
School of Science and Technology, Kwansei Gakuin University, 2-1 Gakuen, Sanda 669-1337, Japan
Email: nakatsu@ksc.kwansei.ac.jp

Richard Reilly
Department of Electronic and Electrical Engineering, University College Dublin, Belfield, Dublin 4, Ireland
Email: richard.reilly@ucd.ie

The use of the keyboard for typing and the mouse for clicking and dragging has been the traditional methods of text and data entry. However, during human-human interaction, we convey a wealth of information through many modes of communication, among them, speech, gesturing, writing, drawing, facial expressions, gaze, and many others. The research challenge therefore is to design and develop human-computer interface systems that incorporate these other communication modes, providing intuitive and seamless interaction.

Research attention to multimodal interfaces has been focused on acquisition, integration, and recognition of different modes of communication; audio and video, speech and gesture, and so forth. One of the greatest research challenges in this area is the combination of these modes, with the aim being the development of robust, reliable applications and usable interfaces to technology.

It is generally accepted that the user can "adapt" their response to suit the interface device, but a more appropriate solution would include the ability of the system to adapt to the user. Therefore, another challenge in this area for the multimodal approach to function efficiently is that each mode must be customisable to the individual. This requires each of the individual elements of the system to be highly configurable to meet the individual's needs

One goal of such research should be to move these intelligent, adaptable interfaces into everyday life in such a way that they become as commonplace as the more traditional, yet restrictive, interfaces. Ensuring that such interfaces are seamless and adaptable poses another significant challenge that can only be addressed in a multidisciplinary manner.

To highlight this area, this special issue titled "Multimedia Human-Computer Interface" focuses on recent developments and applications in the area of multimedia-based user interfaces. It is the accepted theme of this special issue that not one but rather a combination of modes of information is deemed the best practice for human-computer interaction.

Following the reviewing process, a number of papers were selected for this special issue. These papers cover three broad categories: recognition and interpretation of movement, image, and speech; recognition and interpretation of physiological signals; and implementation architectures for multimodal signals.

The initial paper by Quek focuses on gesture recognition. The use of gesture as a form of human-computer interaction has long been a challenging area of research and combining it with speech analysis is an interesting approach. This paper focuses on the combination of video and audio information at the level of discourse planning to improve understanding and performance. The author employs a catchment feature model which uses a feature decomposition approach and which facilitates cross-modal fusion. The author gives three examples of how this type of model can be employed in other multimodal applications.

In the second paper, Morishima and Nakamura describe a multilingual translation system that utilises a multimodal approach. The system focuses on Japanese-to-English and English-to-Japanese translation. The authors report on the use of a 3D face model combined with both the speaker's lip motion and speech recognition to provide the basis for the translation system. The system also translates the speaker's speech motion and synchronises it to the translated speech. The authors believe their multimodal approach is suitable for automatic dubbing from one language into another.

In their paper, Huang and Chung describe a real-time model-based human motion tracking and analysis system. Their approach uses no physical markers and provides 3D motion analysis based on two orthogonal camera views. The authors extract body definition and body animation parameters, and using a classification system based on hidden Markov models, use these two modes of information to provide recognition of fifteen different postures.

The next two papers can be grouped together in that they aim to aid the human visual processing system or extract interpretation from the human autonomic nervous system. The fourth paper in this special issue by Yoshitaka et al. describes a personalised database retrieval system driven by the user's gaze. By using a head mount camera together with a gaze tracker, and by estimating the type of visual information being viewed by the user and calculating the duration of the user's gaze, the context associated with the visual image is retrieved from the database and presented to the user. The authors' "digital reminder" system aims to provide information on which items in the visual field are being attended to by the user.

Lisetti and Nasoz in their paper present a multimodal recognition system to highlight the correlation between affect and cognition. Their system is based on noninvasive wearable signal acquisition and multimodal physiological signal processing. The authors report on their studies into the recognition of six different emotions generated by the autonomic nervous system. Subjects are exposed to a series of video sequences while being monitored for changes in galvanic skin resistance, heart rate, and temperature. Application for this research in multimodal human computer interfaces includes telemedicine, driving safety, and learning.

The final paper by Djenidi et al. addresses the challenging task of implementation of multimodal systems. The authors propose an architecture for the implementation of a multimodal system based on linked multimodal agents. An expert system monitors each multimodal component and dynamically reconfigures the system depending on the application needs.

As can be appreciated from the range of papers in this special issue, this area is a multidisciplinary one and continued advances in this area will be dominated by teams working from different background and perspectives. We encourage further research in this area and the continued submission of results to the EURASIP Journal on Applied Signal Processing.

Ryohei Nakatsu
Richard Reilly

Ryohei Nakatsu received the B.S., M.S., and Ph.D. degrees in electronic engineering from Kyoto University in 1969, 1971, and 1982, respectively. After joining NTT in 1971, he mainly worked on speech recognition technology. In 1994, he joined Advanced Telecommunications Research Institute (ATR) as the President of ATR Media Integration & Communications Research Laboratories. From the spring of 2002, he has been a Professor at School of Science and Technology, Kwansei Gakuin University. At the same time, he established a venture company, Nirvana Technology Inc., and became a President of the company. His research interests include emotion extraction from speech and facial images, emotion recognition, nonverbal communications, and integration of multimodalities in communications. He received the Best Paper Award from the IEEE International Conference on Multimedia Computing and Systems (1996); Telecom System Award from Telecommunication System Foundation (1999 and 2000); the Best Paper Award from Virtual Reality Society of Japan (1999, 2000, and 2001); and the Best Paper Award from Artificial Intelligence Society of Japan (2000). He is a Fellow of the IEEE and the Institute of Electronics, Information and Communication Engineers Japan (IEICE-J).

Richard Reilly received his B.E., M.Eng.Sc., and Ph.D. degrees in 1987, 1989, and 1992, all in electronic engineering, from the National University of Ireland. In 1988, he joined Space Technology Ireland and the Departement de Recherche Spatiale (CNRS Group) in Paris, developing DSP-based onboard experimentation for the NASA satellites. In 1990 he joined the National Rehabilitation Hospital, and in 1992 became a Postdoctoral Research Fellow at University College Dublin, focusing on signal processing for speech and gesture recognition. Since 1996, Dr. Reilly has been on the academic staff in the Department of Electronic and Electrical Engineering at University College Dublin. He is currently Senior Lecturer and researches into neurological signal processing and multimodal signal processing. In 2004, he was awarded a US Fulbright Award for multisensory neuroscience collaborative research with the Nathan Kline Institute for Psychiatric Research, New York. Dr. Reilly was the 1999/2001 Silvanus P. Thompson International Lecturer for the IEE. He is currently an Associate Editor for IEEE Transactions on Multimedia.