Human-Harmonized Information Technology, Volume 2

Toyoaki Nishida
Editor

# Human-Harmonized Information Technology, Volume 2

## Horizontal Expansion

Springer

*Editor*
Toyoaki Nishida
Graduate School of Informatics
Kyoto University
Kyoto
Japan

ISBN 978-4-431-56533-8 ISBN 978-4-431-56535-2 (eBook)
DOI 10.1007/978-4-431-56535-2

Library of Congress Control Number: 2017932921



Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Japan KK
The registered company address is: Chiyoda First Bldg. East, 3-8-1 Nishi-Kanda, Chiyoda-ku, Tokyo 101-0065, Japan

*Dedicated to the late Prof. Yoh'ichi Tohkura*

# Foreword

The Japan Science and Technology Agency (JST) is a major organization in Japan to promote science and technology policies and to provide research funds to researchers mainly in academia in order to support innovative science and technology. CREST is one of the funding programs promoted by JST, and in the field of Information and Communications Technologies, eight research areas are ongoing, one of which is "Creation of Human-Harmonized Information Technology for Convivial Society." Under the Research Supervisor, Prof. Toyoaki Nishida, who took over from the late Prof. Yoh'ichi Tohkura, ten projects have been initiated. I have served as a research advisor for two different areas, one of which is this area. There are ten highly advanced projects selected in the area. I noted as an advisor that the important point of projects in this area is to focus on the technologies for "convivial society". There have been several element technologies studied up to now such as virtual reality, augmented reality, robotics, sensing technologies, and so on. These technologies have been rapidly progressing, but in order to realize a convivial society, these element technologies should be combined and harmonized based on the requirements for convivial societies. The chapters in this publication can contribute to the creation of new theories, new architecture, new systems and the invention of advanced applications to provide convivial societies. Those societies are required not only in the 2020s and 2030s, but in the 2040s, when "singularity of artificial intelligence" is envisaged to be reality.

Tomonori Aoyama, Professor Emeritus, the University of Tokyo
Visiting Professor, Keio University

A harmonization between human beings and machines is becoming a more important function in daily life. The human-harmonized information technology that can understand a human's internal intentions will be able to realize healthy and cultured living and good preparation for an aged society. Some interesting themes of the human–machine harmonization have been explored by Prof. Tohkura and

researchers in a CREST project. Chapters appearing in this book are bringing fascinating results not only to the harmonized society but also to Cyber Physical Systems and the Internet of Things.

Kazuo Asakawa, Fujitsu Laboratories, Ltd.

When the late Yoh'ichi Tohkura, Ph.D., started a field of research with the theme of harmony between human beings and the information environment, "information circulation" was emphasized as a key word that must never be forgotten. It not only referred to the processing of information but also pointed out the importance of feedback of the processed results to human beings. It expresses the idea that we should consider comprehensively the nature of the information environment being created, including "how" is "what feedback" given and whether the results will be useful at all. This book is the first to discuss the information circulation from three perspectives—with regard to people, information, and machines. The research findings introduced here serve as a persuasive guide to the design of our information society.

Eisaku Maeda, Vice President, Head of NTT Communication Science Laboratories

Technological singularity is widely considered as an Artificial Intelligence disaster triggered by highly advanced information technology. This idea of an exclusive relationship between human beings and machines is fascinating and seems to be a relationship inspired by the old story of Frankenstein's monster. The concept of human–machine harmonization, advocated by Prof. Tohkura, considers both human beings and machines as the necessary parts of the "information circulatory system" in human-harmonized information technology. This book predicts the future appearance of technological singularity that would not exclude humans; it will create a harmonious relationship between humans and machines in the information circulatory system. In the future, humanity will have its embodiment harmonized between humans and machines.

Taro Maeda, Professor, Graduate School of Information Science & Technology Osaka University/Center for Information and Neural Networks, NICT

This book guides the reader through cutting-edge research trends anticipating the way in which humans will live their lives in the rich information environments of the near future, receiving intellectual assistance while they work, study, eat, and have fun. The book's chapters cover a wide range of topics, but the reader may find that they come together under a coherent idea of looking ahead into the future. The entirety of this book is related to a research field of JST CREST, one of the most

prestigious research grant schemes in Japan. This fact alone, however, does not fully account for the impression the book strikes on its readers. I am fond of the late Prof. Tohkura's stance in which a JST CREST research area should virtually function as a national research institute, with the research supervisor acting as its director. This was undoubtedly his role as the first research supervisor of the "Creation of Human-Harmonized Information Technology for Convivial Society" research area. From the beginning, he clearly showed his views regarding the definition of a convivial society, as well as how human-harmonized information technologies should be in such societies. In this regard, he emphasized that an information environment should enhance humans' intellectual capabilities. He showed his strong and talented leadership and excellent discernment while selecting all the research directors, who are responsible for each of the chapters in this book now with the prominent research outcomes. Moreover, he, as well as his successor, Prof. Nishida, continuously encouraged the research directors to not only pursue their original research objectives, but also to actively seek the opportunity to discuss and collaborate in joint research projects with other colleagues, in particular those working in this research area. I can say that those stances resulted in a great success. I am sure the reader will enjoy going through every chapter.

Yôiti Suzuki, Professor, Research Institute of Electrical Communication
Tohoku University

Japan is becoming a super-aged society ahead of the rest of the world, and with it, the number of persons who have cognitive disabilities is sharply increasing. Moreover, persons who have developed visual or auditory function disorders or communication disorders in their youth must keep living in a society while burdened by a major handicap. Information technology (IT) compensates for weaknesses in human information processing, and the hope has been that IT will serve as a tool to assist those with language and/or communication disorders. While it is said that current artificial intelligence (AI) has functions that exceed human potential, we must wonder if it has become a technology that is beneficial to everyone. There is also the chance that globalization of useful IT will even lead to the homogenization of human thoughts and sensitiveness.

This project does not simply aim for IT that exceeds human ability of intelligence, but explores the underlying potential and diverse functions of humans while examining what IT should be, such as AI, for harnessing the potential and diverse functions of humans. Through my own research I have experienced how the underlying potential and diverse functions of humans can often surface for the first time after persons develop a physical disability. Persons who have developed language and/or communication disorders will try to converse through gestures or tactile means, while elderly persons who have developed cognitive impairments will increasingly try to convey something through facial expressions and gestures. However, that is insufficient to properly communicate with others and leads to

social isolation, and society has yet to come up with a good approach for how to assist these people.

Fortunately, the researchers of the themes in this project have taken on the challenge of elucidating information on the processing function that lies deep within the brain; not just “intelligence”, but also “emotion” and “consciousness” that lie in the background of intelligence. Furthermore, several of the themes feature highly creative approaches, such as discovering the possibilities of humans from new perspectives and helping to revive the esthetic sense and spiritual cry that lie deep within the brain during the course of evolution. I believe that these approaches will contribute greatly to revealing and fostering human creativity, which are major goals of the project.

This book summarizes the results to date of this project. For example, it describes IT that grasps meanings conveyed by tactile means or movements in addition to words, whereby a computer answers using expressions with universal understanding. Persons with cognitive impairments and those with language and/or communication disorders have been waiting for this technology, which will become highly useful as a tool to promote their social participation. Additionally, research that reveals and fosters human creativity could give rise to various arts, such as new and yet familiar-sounding music that crosses racial and ethnic boundaries. This will give the joy of creativity to humans and give new purpose to living. This book describes how AI and IT should be developed, so that humans can truly understand each other and share the joy of creativity, and describes many hints and approaches for realizing this.

Professor Yoh’ichi Tohkura, the program officer of this project until 2014, was a friend of mine for more than 40 years. We shared the same dream of creating IT that is truly useful to humans, the kind described in this book. I hope that those of you who also wish to realize the same dream will enjoy this book.

Tohru Ifukube, Professor Emeritus, Institute of Gerontology
The University of Tokyo

# Preface

This is the second in a pair of collections describing the major outcomes of the JST-CREST research area on the creation of human-harmonized information technology (IT) for a convivial society, launched in 2009. Even in the short year following the publication of the first volume, we have witnessed much technological progress in information and communications technology, such as Uber's first real-world test of self-driving taxis in Pittsburgh in September 2016. Recent surges in artificial intelligence, however, have significantly increased society-wide concern over the risks of its use to humanity, as illustrated by shocking accidents of Google self-driving cars and Tesla autopilot. The more we hear about the role of artificial intelligence in human society, the more convinced we are of the rightness of the direction of our research, which the late Prof. Yoh'ichi Tohkura set out upon almost 8 years ago. Increasingly, people have come to think that technology should be carefully oriented not to weaken humanity but enhance it, if not challenge it.

We have become more confident than in our early days that we are on the right track in attempting to establish technology to enhance human and social potential. Human potential is the power of an individual that enables her or him to actively sustain the endeavor to reach goals in maintaining social relationships with other people. This involves vision, activity, sustainability, empathy, ethics, humor, and an esthetic sense. Social potential is the power that a society of people as a whole possesses. It encompasses generosity, support, conviviality, diversity, connectedness, and innovativeness. We believe that human and social potential complement each other to enable conviviality.

Professor Tohkura focused on human perception to research human-harmonized IT on the road toward a convivial society. Human-harmonized IT centers on understanding and enhancing cognitive dynamics resulting in an interaction between pathos based on embodied perception and logos based on modern civilization. It is evident that people will need to find a non-traditional style of self-actualization and society will aspire to a new principle of endorsing harmony. Even if the current development of artificial intelligence eventually releases us from our labor, whether physical or informational, individuals and society as a whole will

need to find new styles and ways of living for wellness in a new technological world.

The preceding volume emphasized the vertical aspects of science and technology, in which basic theories on human perception and embodied behaviors form the core of human-harmonized IT, which in turn serves as a foundation for what may be called human-harmonized services in the convivial society we see before us. In this volume, we stress higher layers of human-harmonized IT oriented to a broader range of applications, including content creation, the human-harmonic information environment, health care, and learning support. This volume consists of eight chapters.

Chapter 1 addresses technology to help people not only consume musical content but also use it in a creative fashion. Toward this end, Goto and his colleagues have developed a suite of technologies for building a similarity-aware information environment. Songle is a web-based active music appreciation service that can automatically determine four types of musical descriptions: musical structure (chorus sections and repeated sections), hierarchical beat structure (musical beats and bar lines), melody line (fundamental frequency, f0, of the vocal melody), and chords (root note and chord type). Songle's web service permits anonymous users to correct errors in the musical archive, to cope with the incompleteness of the automated tool. Songle Widget is a web-based multimedia development framework that allows the control of computer-graphic animation and physical devices, such as robots, in synchronization with music publicly available on the web. Songrium is a music-browsing assistance service that allows the visualization and exploration of a large amount of user-generated music content. Goto and colleagues have also developed content-creation support technologies, such as TextAlive, which enables the creation of music-synchronized lyrics animation.

Chapter 2 addresses 3D sound-scene reproduction. Ise and his colleagues have succeeded in the world's first implementation of an immersive auditory display, named the Sound Cask, which implements the principle of boundary surface control (BoSC), a theory of 3D sound-field reproduction. BoSC features the ability to reproduce a sound field, not using points but in three dimensions. As a result, the system can provide high-performance spatial information reproduction, including sound localization and sound distance, even as the listener freely moves her or his head. The performance evaluation of the system is reported, which encompasses physical performance, localization, and the psychological and physiological evaluation of the feeling of reality in a 3D sound field. The Sound Cask system helps music professionals such as musicians, acoustic engineers, music educators, and music critics enhance their skills and further explore their creativity by providing them with the means to experience 3D sound in a telecommunications environment. Applications include a sound-field simulator, sound table tennis, and sound-field sharing.

Chapter 3 describes a framework for user-generated content creation. Tokuda and his colleagues created the MMDAgent toolkit to build voice-interaction systems by incorporating speech recognition, HMM-based flexible speech synthesis,

embodied 3D agent rendering with simulated physics, and dialogue management based on a finite state transducer. MMDAgent was released as an open-source software toolkit. Tokuda and colleagues have constructed an all-in-one set of materials on the use of MMDAgent and the production of dialogue content, including guidebooks/tutorials, slides, reference manuals, and sample scripts. The results have been demonstrated in public installations, including the ones in front of the main gate of the Nagoya Institute of Technology and at City Hall in Handa City, Aichi, Japan.

Chapter 4 presents a project that enables a mobile social robot to adapt to an open public space in a city. Toward this end, Kanda and his colleagues developed a series of techniques to harmonize their robots in daily human contexts. They addressed common-sense problems in the domain of open public spaces, such as a shopping mall corridor where pedestrians walk. They focused on technologies for sensing pedestrians. Their pedestrian model includes collision avoidance and task-oriented human–robot interaction (HRI) encompassing such activities as shopping and observation. They also introduced high-level harmonized HRI features to avoid collision, prevent congestion, and escape "robot abuse"—the nasty treatment of robots, by children in particular. They conducted several field studies and found that they were able to harmonize mobile robots in daily human contexts, and they encouraged people to acquire information from them.

Chapter 5 focuses on the varieties of gait a person has to uncover the relation between gait variation and inertial states, i.e., attention (gaze direction), human relation (group segmentation), and cognitive level (assessment of dementia). For attention estimation, Yagi and his colleagues conducted numerous experiments studying the relationship between gaze and whole-body behaviors. They found similar eye–head coordination in different conditions, which suggests that head orientation is directly related to visual perception; the distribution of the eye position varies systematically with head orientation; the angles of the gaze, head, and chest have linear relationships, under non-walking and walking conditions; not only head but also arm and leg movements are related to the gaze locations; and so on. They also propose a method of determining whether two people belong to the same group, combining motion trajectory, chest orientation, and gesture. Researching dual-task analysis for cognitive-level estimation, they conducted data collection at an elder-care facility and at the National Museum of Emerging Science and Innovation, or Miraikan, in Tokyo. The data obtained from the latter are immense, with more than 95,000 participants. The analysis of these data is in progress.

Chapter 6 addresses the design and demonstration of the future of the information environment where people get together. Naemura and his colleagues focus on three issues: privacy control of display content for promoting discussion in groups, projection-based control of physical objects for suppressing the incompatibility between the physical and digital worlds, and spatial imaging for augmented reality among people without wearable displays. For the privacy-control issue, they propose a privacy-control method called SHelective for sharing displays

and a group-work facilitation system called Inter-Personal Browsing for collaborative web search. For the projection-based augmentation issue, they propose the concept of a bit-data projection system called the Pixel-Level Visible Light Communication Projector and a chemical augmentation system called Hand-Rewriting for paper-based computing. These latter two are functional extensions of existing image projectors to create a more advanced augmentation of the physical world. For spatial imaging, they propose EnchanTable, which can display a vertically standing mid-air image on a table surface using reflection; MiragePrinter, for interactive fabrication on a 3D printer with a mid-air display; and fVisiOn, a glasses-free tabletop 3D display viewable from 360° to augment ordinary tabletop communications.

Chapter 7 describes a reading-life log technology to help people leverage characters to live an intellectual life. Kise and his colleagues, based on character detection, recognition, and generation, transfigure traditional character and document media into new active media, using technologies such as high-speed character recognition and document image retrieval. Their technology is comprehensive, including real-time character recognition for alphanumeric and Japanese characters, omnidirectional character recognition that allows recognition of all characters in a 360° scene image, and real-time document image retrieval based on basic character detectors and recognizers, a large-scale character dataset, and an automatic font generator. On this basis, Kise's group developed reading-life log technology, which not only builds a record for one's reading life: the time spent on, amount of, and attitude toward reading activities, but also analyzes the content of the reading to support the user's intellectual activities. Using reading-life log technology, Kise's group prototyped applications such as Wordometer, which counts the number of words one reads to diagnose one's reading life, a scene-text detector and generator, an automated text annotator, a system for recording texts together with the facial expressions of the reader, and an augmented narrative that uses bio-feedback in a text–body interaction.

Chapter 8 addresses pedagogical machines that can teach and be taught. Hiraki and his colleagues take a threefold approach: the development of cognitive science, machine intelligence, and field studies in an educational environment. In the first approach, they found that infants are very sensitive to temporal contiguity in interaction with their mothers. In particular, it became clear that nowness and responsiveness are very important for the design of a pedagogical machine. With the second approach, they developed a pedagogical agent with gaze interaction (PAGI) that is designed to teach Korean words to Japanese students, capable of simulating mutual gaze, gaze following, and joint attention. Experiments with PAGI showed that even adults are implicitly affected by nowness and responsiveness during word learning with artificial agents. As a result of the third approach, several novel findings have been obtained, e.g., interactions among children where the interactions seem to affect each child's learning and altruistic behaviors.

Chapter 9 is the epilog. I summarize the results obtained in research activities over the past 8 years. In a nutshell, what we have built can be called a perceptually rich common ground between humans and computers. As potential next challenges, I suggest companion agents and robotic apprentices with a more comprehensive common ground, ranging from perception to cognition, that can build and maintain longitudinal companionship with us to help explore larger information spaces.

Kyoto, Japan Toyoaki Nishida
January 2017

# Contents