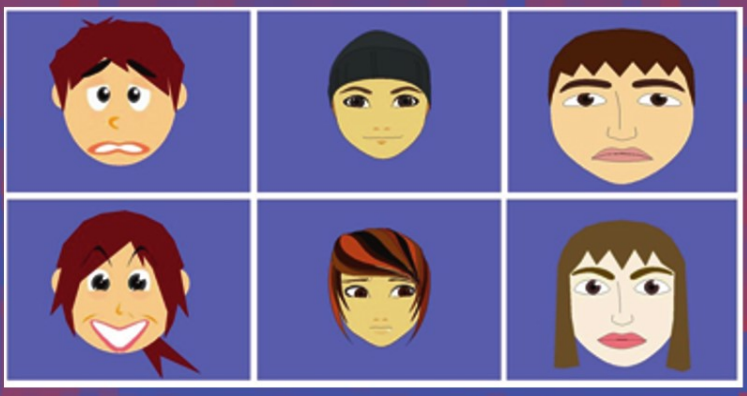
State-of-the-Art Survey

LNCS 6259

Toru Ishida (Ed.)

Culture and Computing

Computing and Communication for Crosscultural Interaction



Springer

Lecture Notes in Computer Science 6259

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Toru Ishida (Ed.)

Culture and Computing

Computing and Communication for Crosscultural Interaction

Springer

Volume Editor

Toru Ishida
Kyoto University, Department of Social Informatics
Yoshida-Honmachi, Kyoto 606-8501, Japan
E-mail: ishida@i.kyoto-u.ac.jp

Library of Congress Control Number: 2010938609

CR Subject Classification (1998): H.5, J.1, K.4, K.4.2, H.2.3

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN 0302-9743
ISBN-10 3-642-17183-4 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-17183-3 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

Preface

In the near future, international communities will face tough global issues, such as problems with population, energy, the environment, and food. Though it is necessary to build a consensus to resolve those issues, there are difficulties with communication among cultures. Contributions from information and communication technologies are required to understand and overcome these difficulties.

In this context, the First International Conference on Culture and Computing was held February 22–23, 2010, in Kyoto, the heart of culture in Japan, to provide an opportunity to share research issues and discuss the future of culture and computing. This conference was designed as a collection of symposiums, panels, workshops, exhibitions, and guided tours organized by various parties including universities, institutes, NPO/NGOs, and industries located in the Kyoto area and special interest groups of Japanese academic societies. We decided to make this conference more than just an academic event by basing it on real and vital activities related to culture and computing. The events forming this conference included:

- Symposium: “Culture and Computing” by Graduate School of Informatics, Kyoto University
- Symposium: “Multilingual Support for Medical Care” by NPO, Center for Multicultural Information and Assistance Kyoto
- Symposium: “Buddhist Culture and Computing” by Kyoto Buddhist Culture Forum
- Panel: “A New Trend in the Humanities Research: Digital Humanities” by Ritsumeikan University
- Panel: “Cultural Computing with Meister, and Shinto Priests in Kyoto” by Academic Center for Computing and Media Studies, Kyoto University
- Workshop: “The Current State and Problems of Cultural Sciences and Computer Study in Japan” by SIG Computers and the Humanities, Information Processing Society of Japan
- Workshop: “Language Grid and Intercultural Collaboration” by SIG-AI and SIG-IC, Institute of Electronics, Information and Communication Engineers (IEICE)
- Workshop: “The Formation of Academic Knowledge Base of Buddhism” by SAT Daizokyo Text Database Committee
- Workshop: “Agents in Cultural Context” by Agents Workshop Committee
- Workshop: “The Scope and Perspectives e-Culture” by APAN e-Culture WG

The first conference featured two prominent keynote speakers, Makoto Nagao, President of the National Diet Library, and Taichi Yuasa, professor of Kyoto University, followed by four joint talks as outlined below.

- "Intercultural Collaboration" by W. Lewis Johnson, Alelo Inc. and Toru Ishida, Kyoto University
- "Digital Humanities and New Trends in the Study of Japanese Culture" by Ellis Tinios, University of Leeds, UK, and Ryo Akama, Ritsumeikan University
- "Web Media and e-Culture" by Julien Masanes, European Archive and Katsumi Tanaka, Kyoto University
- "Cultural Computing" by Ryohei Nakatsu, University of Singapore and Naoko Tosa, Kyoto University

This volume includes 17 invited and selected papers. In parallel with various events, we encouraged academic participants of the conference to submit papers on their presentations at the conference. Those papers were reviewed and selected by program committee members. The accepted papers are categorized into the following areas:

– Culturally Situated Agents
– Intercultural Collaboration and Support Systems
– Culture and Computing for Art and Heritage
– Culture and Computing in Regional Communities

The conference was successfully concluded with 499 attendees. The next conference has not been decided yet, but it will likely be held in Kyoto again in the near future.

May 2010 Toru Ishida

Organization

Organizing Committee

Kozaburo Hachimura	Department of Media Technology, Ritsumeikan University
Toru Ishida	Department of Social Informatics, Kyoto University (Chair)
Michihiko Minoh	Academic Center for Computing and Media Studies, Kyoto University
Katsumi Tanaka	Department of Social Informatics, Kyoto University
Naoko Tosa	Academic Center for Computing and Media Studies, Kyoto University

Program Committee

Ruth Aylett	Heriot-Watt University, UK
A. Charles Muller	The University of Tokyo, Japan
Makoto Goto	Hanazono University, Japan
Toru Ishida	Kyoto University, Japan (Chair)
Takaharu Kameoka	Mie University, Japan
Yasuhiro Katagiri	Future University Hakodate, Japan
Yasuhiko Kitamura	Kansai University, Japan
Tomoko Koda	Osaka Institute of Technology, Japan
W. Lewis Johnson	Alelo Inc., USA
Kiyonori Nagasaki	Institute for Research in Humanities, Japan
Toyoaki Nishida	Kyoto University, Japan
Ana Paiva	Instituto Superior Te'cnico, Portugal
Alicia Sagae	Alelo Inc., USA
Yoshinori Sato	Tohoku Gakuin University, Japan
Naomi Yamashita	NTT, Japan
Takashi Yoshino	Wakayama University, Japan

Organized by

The Organizing Committee of the International Conference on Culture and Computing

Co-sponsored by

Graduate School of Informatics, Kyoto University
Academic Center for Computing and Media Studies, Kyoto University
Kyoto University Global COE Program Informatics Education and Research Center for Knowledge-Circulating Society
Ritsumeikan University Global COE Program Digital Humanities Center for Japanese Arts and Cultures
Advanced Scientific Technology & Management Research Institute of Kyoto
Kyoto Research Park Corporation

Supported by

The Institute of Electronics, Information and Communication Engineers
Information Processing Society of Japan
The Japanese Society for Artificial Intelligence
Kyoto Prefecture
Kyoto City

Table of Contents

Culturally Situated Agents

Intercultural Collaboration and Support Systems

Culture and Computing for Art and Heritage

Culture and Computing in Regional Communities

Using Immersive Simulations to Develop Intercultural Competence

W. Lewis Johnson

Alelo Inc., 19210 Culver Bl., Suite J, Los Angeles, CA 90066 USA
ljohnson@alelo.com

Abstract. This chapter describes a simulation-based approach to cultural competency training realized in the Alelo family of learning products. It is intended to help people who may not be cultural specialists to quickly develop the cultural skills they need to be effective in intercultural settings. This includes the knowledge and skills necessary to handle common situations involving intercultural interactions, as well as the adaptability needed to cope with unexpected intercultural interactions. Crucially, it utilizes simulations of intercultural encounters that learners are likely to experience in the course of carrying out their jobs or missions, which help learners develop the skills and confidence necessary to be effective in those situations. The approach also supports the assessment of cultural competence by testing trainees in simulated intercultural encounters. The approach makes use of a number of innovative technologies, but most importantly immersive game and artificial intelligence technologies.

Keywords: Cultural competence, cultural skills, agent-based computing, simulation-based learning environments, serious games.

1 Introduction

Intercultural skills are increasingly recognized as important for many jobs and everyday situations. The workforces of multinational corporations are frequently organized into multinational teams in which people from various cultural backgrounds need to work together effectively. Companies that deliver products internationally need to understand the cultural backgrounds and perspectives of their potential customers in each target market. Professionals engaged in health care, law enforcement, education, humanitarian relief, and peacekeeping operations all require skills in interacting with people with other cultural backgrounds [2], [5], [10], [11], [14].

This need for intercultural knowledge and skills poses a significant educational challenge. Experts in the language education community, such as the American Council on the Teaching of Foreign Languages (ACTFL), commonly assert that true intercultural competence arises only after years of immersion in the target culture, as part of a language education program [9]. Unfortunately, relatively few specialists in other professions get the opportunity to devote that much time to cultural training. Thus a major challenge for cultural skills training is finding ways to help people develop the intercultural skills they need, in the time they have to learn.

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 1–15, 2010.

This chapter describes a simulation-based approach to cultural competency training realized in the Alelo family of training products. It is intended to help learners who may not be cultural specialists to quickly develop cultural skills they need to be effective in intercultural settings. This includes the knowledge and skills necessary to handle common situations involving intercultural interactions, as well as the adaptability needed to cope with unexpected intercultural interactions. Crucially, it utilizes simulations of intercultural situations that trainees are likely to encounter in the course of carrying out their jobs or missions, which helps trainees develop the skills and confidence necessary to apply intercultural skills in those situations. The approach also supports the assessment of intercultural competence by testing trainees in simulated intercultural encounters.

The approach makes use of a number of innovative technologies, but most importantly immersive game and artificial intelligence technologies. Immersive game technologies are used to simulate the intercultural encounters that serve as the context for skills training. Artificial intelligence technology is employed to create interactive computer characters that can engage in dialog with the learners, in a culturally appropriate way.

2 Example Courses

Many users of Alelo courses are members of military services around the world, who require training in cross-cultural communication skills so that they can perform missions overseas, such as reconstruction and humanitarian assistance, and work together with coalition military partners. Fig. 1 shows a scene from one such course, the US version of the Operational Dari language and culture trainer. Operational Dari includes interactive lessons that teach task-relevant communication skills and interactive scenarios in which learners can practice those skills. In the example in the figure, the learner, playing the character on the left, is engaged in a meeting with the elders in an Afghan village. He needs to develop rapport with the elders in order to gain their trust and cooperation. To play the scenario, the learner must speak into the microphone in Dari, speaking phrases appropriate for his or her character in the scenario, and the non-player characters respond accordingly. The dialog history is shown in the top center of the figure. This is a relatively advanced scenario; more elementary modules and scenarios help trainees who initially have no knowledge of Dari get to the point where they can engage in such conversations with confidence.

We also develop learning materials for other platforms to provide comprehensive, anywhere-anytime learning and refresher training. We deliver interactive lesson content on iPods and other handheld devices. This can be used either as a low-cost self-contained learning tool or as a refresher practice supplement. We have also developed an artificial intelligence plug-in for popular immersive multi-player training environments, making it possible to populate virtual worlds with artificially intelligent characters playing a range of different roles. This makes it possible to integrate language and culture training seamlessly with other types of job training.

Fig. 1. Operational Dari language and culture training system

Fig. 2. Exercise from the goEnglish Web site

This scenario-based approach is being applied more broadly to teach language and culture. For example, Fig. 2 shows a dialog exercise from goEnglish, a Web site developed for Voice of America to teach American culture and language worldwide.

The site is designed for learners who have acquired a basic reading knowledge of English in school, but have a limited understanding of American culture, and have had a limited opportunity to develop and practice their conversational skills. It contains a set of lessons that focus on the language and culture of situations that arise in everyday life in the United States, as well as cultural topics of particular interest. To develop the set of topics, we interviewed recent immigrants to the United States to find out what situations they found particularly surprising or challenging. The site contains a set of lessons where they learn conversational language and cultural points relevant to these situations and topics, and exercises in which they practice in simulated conversational situations. A speech recognition plug-in is loaded into the user's Web browser, enabling learners to speak English into a microphone and have the non-player characters in the situation respond to what they say. In the situation shown in Fig. 2, the learner is getting to know an acquaintance named Malika and is talking about their relationships. The learner has found out that Malika has a boyfriend (see the dialog history in the top center of the figure) and can now ask further questions, such as when and where they met. If the learner is uncertain as to what to say, he or she can consult a menu of hints about conversational topics (top left, in Chinese) and possible ways of expressing each particular topic (center left). The learner's score for the exercise depends upon the total number of dialog exchanges, and the extent to which the learner relied on hints to get through the scenarios.

There are multiple versions of the goEnglish Web site, aimed at different learner communities. In addition to the Chinese site shown in Fig. 2, there is a site in Persian, and sites for speakers of Russian and other languages are currently under development.

3 Cultures Modeled to Date

Developing Alelo courses requires knowledge of the culture being taught. This is used both to develop learning materials that teach about the culture, and to develop animated agents that behave in accordance with the norms of the culture, as in Figs. 1 and 2. Depending upon the target audience and the learning objectives of the course, some courses provide extensive instruction in the local language, and include characters that understand and speak only the local language, while others include a more limited amount of language instruction and engage in conversation in the learner's native language.

It is also important to have a good understanding of the learner's culture to help ensure that the learning materials and objectives are presented in a style, and with language and phraseology, that match the learner's expectations. As noted above, goEnglish is available in multiple languages. Likewise, Operational Dari is available in versions intended for use by the military services of the United States, Australia, the United Kingdom, Germany, and other NATO countries. The instructional language of the one for Germany is in German; all the others are in English. Course details vary slightly in each case to reflect the learner's culture. Differences include uniforms and physical appearance of game characters, artwork, as well as terminology, spelling, and phraseology. Voiceovers and narrations may be recorded

using voice actors from the local country. Military services and other large organizations have their own distinctive subcultures with specialized terminology and writing style, which we try to respect where possible. The extent to which a particular course needs to be adapted to the local culture typically depends upon the size of the country. For example, the German and British versions required extensive adaptation for the local culture, while most smaller NATO countries are willing to use more generic courses.

To date we have developed courses with cultural instruction and extensive language instruction for the following countries: Iraq, Afghanistan, Indonesia, French-speaking sub-Saharan Africa, and the United States. The United States courses include a course on the majority language (goEnglish), as well as a pilot course in a Native American language (Cherokee). More culture-focused courses have been developed for the countries of the Horn of Africa, and additional courses are being developed for Northern Africa and other countries.

4 Course Structure and Development Methodology

Cultural knowledge potentially encompasses a broad range of topics, such as social structure, material culture, economics, political structure, art, as well as the culture of day-to-day life. It is difficult to cover all these topics in enough depth that learners can competently and effectively apply the cultural knowledge they gain. We employ a course development methodology that enables us to identify which types of situations are most important, and what cultural skills are needed to master those situations. We then design the courses to help learners develop and master those skills. This offers practical benefits for learners, because it gives them cultural competencies that they can immediately put to use when interacting with people from the other culture. It helps learners develop confidence in their ability to engage in intercultural interaction. Once learners start becoming effective in interacting with people from the other culture, they will learn more about the culture from those interactions, and are likely to continue to broaden and deepen their cultural knowledge on their own.

This course development methodology, called the Situated Culture Methodology (SCM) [3], [15], works as follows and is illustrated in Fig. 3. The first step, outlined in the top left corner of the figure, is to determine scope of the course. This takes into account the situational contexts in which learners are likely to employ their cultural skills. First, it is necessary to identify the intended scope of the course, in terms of the breadth of the geographic region being covered and the duration of the course. To further focus the course, we work with cultural subject matter experts to identify the particular jobs and missions that learners will likely want to train for. This process identifies typical scenarios that require intercultural skills, situations that learners might encounter during those scenarios, and tasks that learners might need to perform in those situations. Given these situations and tasks, it is then possible to identify specific *situated culture learning objectives*: knowledge and skills that learners need to acquire in order to be effective in those situations. If the learners must employ the local language in these situations, then a set of situated language learning objectives are implied as well.

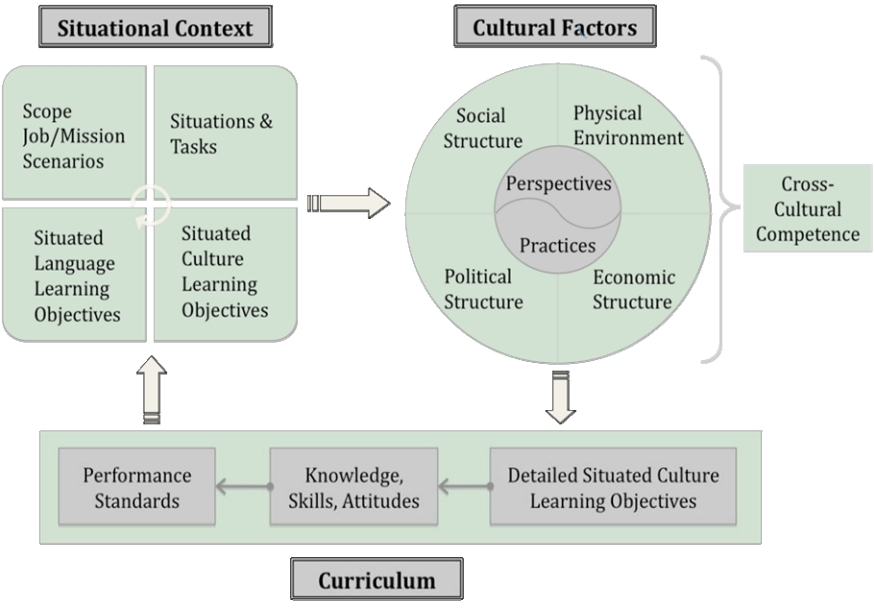


Fig. 3. Situated Culture Methodology

Once these situated learning objectives are defined, we go through a process of identifying the specific cultural factors that need to be covered in the course to meet these objectives. These factors are grouped in the following general categories: the physical environment or milieu in which the culture is situated, and the political, economic, and social structures common within the culture. The perspectives that are typical of people within the culture are important, such as people's attitudes toward time, personal relations, work, and the role of the individual within the community. We also identify relevant cultural practices, including norms of conversational interaction and nonverbal communication.

Note that the process of identifying the relevant cultural factors often involves research and the development of data about culturally appropriate interaction in the target situations. Published sources of cultural information frequently lack sufficient detail about how people in the culture would actually behave in a given situation. Using a combination of ethnographic interviews and role-playing with subject matter experts, we elicit details of how people typically interact in these situations. We then codify this information in the form of dialogs that illustrate how people in the culture interact with each other, and annotate them with explanations of the dialog moves and indications of how people in the culture commonly interpret them. These dialogs serve as the basis for designing simulated dialog exercises and the agents that play appropriate roles within those exercises.

In addition to specific knowledge about particular cultures, learners also require general cross-cultural competence, i.e., general skills for handling intercultural situations [8]. Cross-cultural competence helps learners recognize and understand cultural differences, and provides a general set of skills for coping with such differences when they arise in intercultural contexts.

Fig. 4. A goEnglish lesson page

Once the particular cultural factors are identified, they are used to develop the specific curriculum content to be included in the course. Detailed learning objectives are defined for each relevant cultural factor. These typically comprise *knowledge* about the culture, intercultural *skills*, and *attitudes* that are conducive to success in intercultural settings. The knowledge and skills tend to be interrelated; learners first acquire knowledge about the culture and then acquire the skills necessary to apply this knowledge to intercultural situations. We then identify the target performance standards for each learning objective, e.g., demonstrate the ability to apply the cultural skills in particular simulated scenarios.

The result is a computer-based curriculum containing lesson pages that introduce language and cultural knowledge, exercises that reinforce knowledge and develop intercultural skills, and simulated scenarios for practicing and assessing those skills. For example, Fig. 4 shows one of the lesson pages in goEnglish that introduces the language and cultural knowledge that learners apply in the simulated small talk scenario in Fig. 2. This page explains that, in modern American culture, being "single" indicates not only that a person is unmarried, but also that they do not have boyfriend or girlfriend. The page also gives the learner the opportunity to practice asking someone whether they are single, and giving an appropriate reply. The learner speaks the phrase into the microphone, and the quality of their speech is scored. Learners advance through a series of progressively more difficult practice exercises

until they are ready to practice in a simulated scenario such as those seen in Figs. 1 and 2.

Once the course has been developed and fielded, we evaluate trainee performance in the course and seek feedback from learners who have tried to apply what they learned in real-life intercultural interactions. This results in an iterative process, in which experience in the field results in revisions to the situational contexts, cultural factors, and detailed learning objectives of the course.

5 Cultural Modeling Technologies

Immersive simulation technologies, together with other supporting technologies, play critical roles in developing and implementing these courses. Knowledge management tools are being developed to capture and organize the cultural knowledge. Authoring tools have been developed to help authors create interactive lessons and activities for teaching cultural knowledge, skills, and attitudes. Immersive simulation technologies, incorporating intelligent agents, are used to simulate culturally appropriate behavior and assess the learners' ability to behave in culturally appropriate ways.

These technologies come together in the conversational agent architecture used in Alelo courses to generate the behavior of the non-player characters in the simulated scenarios. Fig. 5 gives an overview of the agent architecture common to the learning environments described here, and its interaction with the learners and the immersive game environment. Further details are available in [4]. The most fully developed implementation of this architecture is in the Virtual Role-Player (VRP) system, a system that enables trainers to populate immersive environments with artificially

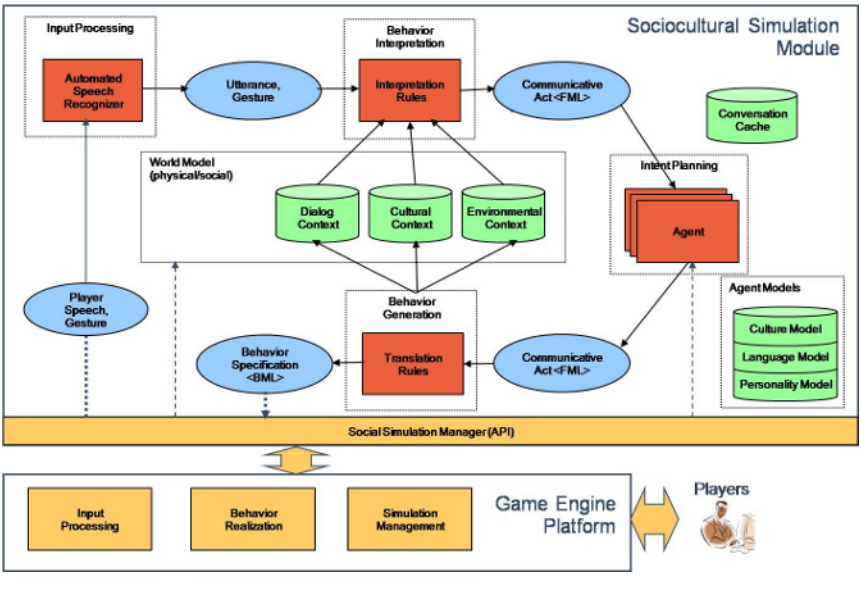


Fig. 5. Conversational agent architecture

intelligent non-player characters that are capable of producing culturally appropriate behavior in 3D multi-player training scenarios. This discussion will focus primarily upon the VRP implementation, which is being developed for military mission rehearsal training. Other Alelo learning environments, such as Operational Dari and goEnglish, currently employ simplified versions of this architecture, but are migrating to the new VRP architecture.

5.1 Game Engine Platform

Game engines commonly incorporate features for creating virtual worlds, populating those worlds with objects and animated characters, and interface controls that enable users to interact with the objects and the animated characters. We build on these to create a set of software platforms for building and running virtual environments for learning culture. These platforms have a common set of capabilities, whether they support 3D interfaces as in Fig. 1, or 2 ½ D Web-based interfaces as in Figs. 2 and 4. This is labeled the Game Engine Platform in Fig. 5.

The game engine platform provides a conversational interface that allows the learners to choose which non-player characters to talk to, and a set of graphical user interface controls for managing the dialog, as shown in Figs. 2 and 4. These include buttons for starting and stopping speech input (top right corner in the figures), and controls for choosing nonverbal gestures for the learner's character to perform. The interface provides other information such as a summary of the objectives to complete in the science, hints about what to say and what to do, and a transcript of the dialog.

The game engine platform is also responsible for generating the behavior of each character. This includes speech, animated gestures, and body movement. We record a set of possible spoken lines for the characters to say and gesture animations for the characters to perform, and load them into the game engine platform. The characters are then able to perform any combination of speech and gesture, as directed by the social simulation module, which selects the appropriate speech and gestures.

The game engine platform includes a simulation management capability that keeps track of the location of objects, characters, and landmarks in the virtual world, many of which have cultural significance. It controls and tracks the movement of people and objects in the virtual world, and can trigger event notifications in response, e.g., cause a non-player to react when a player approaches within a set distance.

5.2 Sociocultural Simulation Module

The Sociocultural Simulation Module is responsible for controlling the behavior of the non-player characters in the simulation, and ensuring that they behave in culturally appropriate ways. The agent dialog models follow an interpretation-decision-behavior cycle, in a variant of the SAIBA agent architecture [16].

The first phase of processing starts with interpreting the learners' spoken utterances and gestures. We develop speech recognition models for each language, specially designed for learners of foreign languages. Acoustic models are trained on both native speech and the speech of learners of the language, to ensure that the speech recognizer is tolerant of the accents of language learners. The recognizers utilize grammar-based language models constructed from a database of sample

utterances in the language. As course developers author scenario dialogs and other learning materials, the example utterances in these materials are entered into the database and then used to create the language models. The speech recognition system runs using the Julius speech recognition decoder, an open-source speech recognition engine developed under the leadership of Kyoto University [6].

Next, the agent evaluates and interprets the communicative intent of the learner's utterance and gesture. Communicative intents are represented using a library of communicative acts, derived from speech act theory originating in the work of Austin [1], and further developed by Traum and Hinkelman [13]. Each communicative act has a *core* function, i.e., the illocutionary function of the utterance (to greet, inform, request, etc.), and *grounding* function, i.e., the role of the utterance in coordinating the conversation (e.g., to initiate, continue, acknowledge, etc.). The grounding functions help to determine the current dialog context, which in turn can influence how subsequent utterances are interpreted. At each point in the dialog, the agent is expecting to hear and respond to one of a set of possible communicative acts, which changes over the course of the conversation. If the learner says something that is not appropriate at that stage of the conversation, e.g., greeting a character at the end of the conversation instead of the beginning, the agent will act as if the learner said something odd that does not make sense.

Note that the interpretation of the utterance and gesture depends upon the particular culture being modeled. The mapping from utterances to communicative acts is specified for each language. Some gestures have meaning only in certain cultures, e.g., placing the palm of the right hand over the heart in greeting only has meaning in Islamic countries. Some gestures are appropriate only in some social contexts; for example, American culture and Arab cultures differ as to when it is acceptable to shake hands with the opposite sex, or kiss the cheek of someone of the same sex.

Depending upon the type of dialog exercise, the agent may not just evaluate the appropriateness of the learner's communication, but also identify and classify the learner's mistakes. Operational Pashto and goEnglish both include so-called mini-dialog exercises, in which learners practice individual conversational turns with a non-player character and receive feedback regarding any mistakes they may have made. Detected errors include grammatical errors, semantic errors (e.g., confusing words with similar meanings), and pragmatic errors (e.g., inappropriate use of expressions of politeness, honorifics, etc.).

Once the learner's input is interpreted, the intent planning stage occurs, in which each agent in the conversation decides how to respond. Intent planning is challenging because it must address multiple conflicting needs: accuracy, versatility, authorability, and run-time performance. The agents should choose communicative acts that are culturally appropriate, e.g., that match the dialog examples created in the cultural data development process described in section 4. However the agent models cannot simply follow the example dialogs as scripts, but need to be versatile enough to respond in a culturally appropriate way regardless of what the learners might say. The agent modeling language needs to be powerful enough to achieve such versatility, yet be authorable by instructional designers who lack the computer science background required for sophisticated agent programming languages. It also is important for the intent-planning module to have good runtime performance, so that the intent planning

process does not interfere with speech processing or 3D scene rendering, both of which require significant amounts of computing resources.

Our most common approach for specifying agent intent planning is through finite state machines. Our dialog authoring tools allow authors to specify the phases of each dialog, the communicative acts the learner might perform at each phase, and the acts that the agent should choose in response. This approach is highly authorable and has been used to create hundreds of dialog models for a variety of different cultures and agents. Unfortunately, it is somewhat lacking in versatility; agent models must be authored specifically for each agent in each scenario. To address this limitation, VRP provides a rule-based intent-planning engine that enables authors to specify cultural behavior rules. Each rule specifies a learner communicative act that the agent should respond to, a set of conditions that must be met, and the effects of applying the rule. Conditions can refer to properties of the learner's communicative act (e.g., its degree of politeness), the state of objects in the virtual world, and properties of the social relationships between the characters in the scenario. For example, the response of a non-player character such as the village elder in Fig. 1 depends upon the extent to which the learner has developed the village elder's trust, which in turn depends upon what the learner has said and done up to that point.

Once each agent in the conversation chooses an action to perform, the sociocultural simulation module generates behavior to realize the action. This typically includes selecting an utterance that realizes the utterance in the target language, as well as one or more animated gestures. These in turn are passed to the game engine platform for execution.

Using the agent behavior authoring tools, we create libraries of reusable agent models. Authors can then create cultural training scenarios by taking a virtual world representing the cultural environment, populating it with non-player characters, and assigning agent models to each character appropriate for that character's role in the scenario. Example roles include village elders, shopkeepers, and passers-by on the street. Each agent will respond to the learners in a manner appropriate for the culture, the agent's role in the scenario, and the current state of the unfolding situation. Authors can extend and modify the rule sets of the agents if needed to reflect the specific characteristics of the scenario.

6 Evidence of Effectiveness

The effectiveness of this approach has been documented in various published studies (e.g., [12]). The most dramatic evidence of effectiveness comes from a study conducted by the US Marine Corps Center for Lessons Learned (MCCLL) [7], which studied the experience of the 3rd Battalion, 7th Marines (3/7 Marines) in Anbar Province, Iraq in 2007. Prior to deployment to Iraq, the battalion assigned two members of each squad of approximately thirteen marines to spend forty hours in self-study training with Alelo's Tactical Iraqi course. It should be noted that (1) forty hours is not a long time to spend learning a foreign language, (2) Arabic is a difficult language for most English speakers, and (3) self-study computer-based language learning tools often fail to produce significant learning gains. However the 3/7's experience drew attention both prior to deployment and after deployment. In final

exercises prior to deployment, Iraqi speaking role players commented that the 3/7 demonstrated Arabic language ability far beyond the skills of typical marine units preparing for deployment. During their entire tour of duty in Iraq, the 3/7 Marines did not experience a single combat casualty. The MCCLL interviewed officers in charge of the unit and conducted surveys of the individual trainees, and found numerous instances where language and cultural skills contributed directly to mission effectiveness. Most importantly, it demonstrated an appreciation and willingness to learn about Iraqi culture, which caused the Iraqis to be more positive and cooperative, and set in motion a virtuous cycle of cooperation leading to operational effectiveness leading to mutual trust, leading to further cooperation, etc.

7 Current Work

We are now working to extend our sociocultural models and apply them to new types of learning environments. An important area of development is in the area of environments that support multiple learners as well as multiple non-player characters. When learners reach a sufficient level of communicative proficiency, multi-player learning environments can become advantageous, as a way to motivate learners and encourage further practice. We are using the Virtual Role-Player (VRP) architecture to create multi-player mission rehearsal environments in which teams of trainees work together on mission rehearsal scenarios. Trainees must interact with non-player characters representing local people in order to complete their mission successfully. We plan to extend the VRP framework so that a player can assume control of one of the local people characters and direct its actions without the learners being aware of it. This makes it possible to offer the best possible combination of human-human interaction and human-agent interaction in intercultural skills training. Individuals with communicative competence in the target language and culture can play roles within the scenario and artificially intelligent virtual role players can fill in and play additional roles as needed.

Another learning game under development, the ISLET Game, takes the concept of multi-player game-based cultural learning even further. The ISLET Game is part of a larger language learning environment called ISLET (Integrated System for Language Education and Training), that is exploring methods for making language learning highly motivating, so that learners will devote their free time to learning a language, just as they would an entertainment game. The ISLET Game incorporates a sophisticated reward system, similar to what is found in multi-player entertainment games such as *World of Warcraft*, that motivates learners to keep practicing and developing their communication skills. It incorporates engaging multi-player quests which encourage collaboration between learners. It also supports team conversations (see Fig. 6), in which learners work together in conversing with a non-player character and learn from each other in the process.

The CultureCom project is developing formal models of the cultural influences underlying dialog and utilizing them to increase the flexibility and realism of the behavior of non-player characters in training simulations. The work is being conducted in collaboration with Dr. Michael Agar of Ethknoworks and Prof. Jerry Hobbs of the University of Southern California. Cultural and linguistic

Fig. 6. Scene from the ISLET Game

anthropologists are developing validated sociocultural data sets for Afghanistan and other cultures of interest, consisting of annotated dialogs of cross-cultural interactions. Experts in artificial intelligence then use these data to develop logical models of sociocultural behavior in different cultures, based upon a formal ontology of microsocial concepts underlying interpersonal communication. This in turn is being used to create an enhanced version of the VRP architecture in which agent intent planning utilizes explicit validated models of sociocultural reasoning for different cultures, which can be swapped in and out to enable agents to model a variety of different cultural characteristics.

The following example illustrates how CultureCom cultural models will be developed and used. American culture and Afghan culture differ in the way they express promises and commitments. Afghans sometimes agree to a request as a way of being socially agreeable, without making a firm commitment. In CultureCom, we explicitly model the sociocultural inferences that can be made from communicative acts, such as whether a statement of agreement constitutes a firm promise and commitment. This in turn can be used to ensure that the non-player character's actions are consistent with the culture throughout, and can also provide helpful feedback to the learner. For example, it can help learners to recognize when intercultural misunderstandings can arise due to different views of what has been promised and agreed to.

The C-CORE project is developing a set of authoring tools and workflow management tools to facilitate the creation and maintenance of cultural models. It enables culture researchers to incorporate sociocultural data from a variety of sources,

including subject matter expert first-person accounts, field reports, hand-authored dialogs and dialog transcripts, and other sources. It supports the process of managing, processing, and utilizing streams of updates to sociocultural data to keep the cultural models up-to-date as new information is obtained or properties of the cultural environment changes. It provides tools for comparing the behavior produced by different character models, and for comparing different versions of the same character model resulting from changes in the sociocultural data. The output of C-CORE will be specifications for cultural simulations, including the cultural environment, scenario logic, non-player characters and their behavior.

C-GAME, developed in coordination with C-CORE, is extending the overall approach for cultural competence training, and developing it into an interoperable framework that can be utilized with a variety of game engines and artificial intelligence models. The cultural simulations in C-GAME will serve as rich learning environments in which learners can develop their knowledge and awareness of the culture, and then develop intercultural skills, all within the same environment.

The scenario construction toolset will enable trainers to develop their own sociocultural training scenarios through a combination of libraries of reusable sociocultural models and authoring tools. It will include mechanisms for incorporating current sociocultural and human terrain data into the scenarios (e.g., from C-CORE), to keep them up-to-date. It will provide a way to annotate a variety of entities in the virtual world in terms of their cultural significance, so that learners can learn their meaning in the context of the culture. This may include clothing and other examples of the material culture, locations (public, private, or sacred places), food and other items associated with cultural practices, and even nonverbal gestures and other human behaviors. These elements will be drawn from an ontology of cultural entities developed in C-CORE, which in turn will relate back to the dimensions of culture shown in Fig. 3.

C-GAME will also provide a way for trainers to define trainee cultural performance standards. Trainers will be able to assess those performance standards themselves by participating in or observing the training scenario, and they can rely on automated cultural skills assessments built into the non-player characters.

The C-GAME run-time execution engine will incorporate instrumentation to compare trainee performance to the training objectives and provide automated performance review and remediation capabilities. This includes assessment of trainee performance at each stage in the interpretation-decision-behavior cycle of the virtual role players. In the interpretation phase, whenever learners say or do something that is linguistically or culturally inappropriate, the role players will attempt to analyze and classify the nature of the action. This analysis is similar to the analysis that takes place in mini-dialogs, as described in section 5, but can apply to any immersive scenario. Learners will get feedback on their cultural competence both from the moment-to-moment reactions of the non-player characters and from the performance review and the end of the scenario.

Acknowledgments. The author wishes to express his thanks to the members of the Alelo team who contributed to this work, and to LeeEllen Friedland for her help in editing. This work was sponsored by USMC PMTRASYS, Voice of America, Office of Naval Research, and DARPA. Opinions expressed here are those of the author and not of the sponsors or the US Government.

References

1. Austin, J.L.: How to Do Things with Words. Harvard University Press, Cambridge (1975)
2. Earley, P.C.: Intercultural training for managers: A comparison of documentary and interpersonal methods. The Academy of Management Journal 30(4), 685–698 (1987)
3. Johnson, W.L.: A simulation-based approach to training operational cultural competence. In: Proc. of ModSIM World 2009 (2009)
4. Johnson, W.L., Valente, A.: Tactical Language and Culture Training Systems: using AI to teach foreign languages and cultures. AI Magazine 30(2), 72–84 (2009)
5. Kosoko-Lasaki, S., Cook, C.T., O'Brien, R.L.: Cultural Proficiency in Addressing Health Disparities. Jones & Bartlett, Boston (2008)
6. Lee, A., Kawahara, T.: Recent Development of Open-Source Speech Recognition Engine Julius. In: Proc. of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (2009)
7. MCCLL: Tactical Iraqi Language and Culture Training System. MCCLL Newsletter 4(8), 4 (2008)
8. McDonald, D.P., McGuire, G., Johnson, J., Selmeski, B., Abbe, A.: Developing and managing cross-cultural competence within the Department of Defense: Recommendations for learning and assessment. Technical report, RACCA WG (2008)
9. National Foreign Language Center (NFLC): Cultural Proficiency Guidelines (3.2), http://www.nflc.org/culture-prof-guide_3-2.pdf
10. Schneider, P., Sadowski, D.: The effects of intercultural collaboration strategies on successful PhD education. In: Proc. of IWIC 2009. ACM Press, New York (2009)
11. Sorcher, M., Spence, R.: The interface project: Behavior modeling as social technology in South Africa. Personnel Psychology 35(3), 557–581 (1982)
12. Surface, E.A., Dierdorff, E.C. and Watson, A.M.: Special Operations Language Training Software Measurement of Effectiveness Study: Tactical Iraqi Study Final Report. SWA Consulting technical report 2007010602 (2007)
13. Traum, D., Hinkelman, E.: Conversation acts in task-oriented spoken dialogue. Computational Intelligence 8, 575–599 (1992)
14. US Dept. of Health and Human Services (USDHHS): Cultural competency curriculum for disaster preparedness and crisis response, https://cccdpcr.thinkculturalhealth.org/
15. Valente, A., Johnson, W.L., Wertheim, S., Barrett, K., Flowers, M., LaBore, K., Johansing, P.: A Dynamic Methodology for Developing Situated Culture Training Content. Technical Report, Alelo (2009)
16. Vilhjálmsson, H., Marsella, S.: Social Performance Framework. In: Proceedings of the AAAI Workshop on Modular Construction of Human-Like Intelligence (2005)

Cross-Cultural Study on Facial Regions as Cues to Recognize Emotions of Virtual Agents

Tomoko Koda[1], Zsofia Ruttkay[2], Yuka Nakagawa[1], and Kyota Tabuchi[1]

[1] Faculty of Information Science and Technology, Osaka Institute of Technology
1-79-1 Kitayama, Hirakata, 573-0196 Osaka, Japan
koda@is.oit.ac.jp
[2] Moholy-Nagy University of Art and Design
Zugligeti ut 9-25, Budapest, Hungary
ruttkay@mome.hu

Abstract. This paper reports the preliminary results of a cross-cultural study on facial regions as cues to recognize the facial expressions of virtual agents. The experiment was conducted between Japan and Hungary using 18 facial expressions of cartoonish faces designed by Japanese. The results suggest the following: 1) cultural differences exist when using facial regions as cues to recognize cartoonish facial expressions between Hungary and Japan. Japanese weighed facial cues more heavily in the eye regions than Hungarians, who weighed facial cues more heavily in the mouth region than Japanese. 2) The mouth region is more effective for conveying the emotions of facial expressions than the eye region, regardless of country. Our findings can be used not only to derive design guidelines for virtual agent facial expressions when aiming at users of a single culture, but as adaptation strategies in applications with multicultural users.

Keywords: facial expression, virtual agents, character, cross-culture.

1 Introduction

Virtual agents are frequently used in virtual worlds, online applications and (serious) games. In the current state of development of the virtual agent technology, virtual agents can express emotions in their bodily behavior, first of all, by displaying facial expressions. The culture of the virtual agent is relevant in relation to the culture of the real human interlocutor. Moreover, there are training and simulation applications emerging where the cultural identity and particularly, the bodily behavior is the major learning component of the training application [1]. Such agents have been designed under the assumption that their expressions are interpreted universally among all cultures. The basis for this assumption is the early finding about the universality of the 6 basic expressions – joy, surprise, fear, sadness and disgust - by Ekman [2]. Later works by Ekman and colleagues indicated cultural differences in perceived intensity of emotions [3, 4]. However, recent research indicates cultural differences in recognizing human facial expressions. Elfenbein et. al. [5] have coined the term cultural dialects of facial expressions: the cultural dialect, unlike a personal

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 16–27, 2010.

idiosyncratic variant, is a well identifiable specific usage of some facial signal. Such cultural dialect can be seen as an in-group advantage in human facial expression recognition, whereby human facial expression recognition is generally more accurate for perceivers from the same cultural group as expressers. These works all have been using photographs, where the cultural identity of the face was clear. Ruttkay [6] discusses further empirical findings and provides a detailed scheme of the possible factors of cultural differences in interpreting facial emotions of virtual agents.

When designing virtual agents, it is a challenging and not very much exploited possibility to use non-realistic faces. There are two motivations for going for cartoon-like faces. On the one hand, it has been shown that the more realistic the design is, the more critical the human perceivers are. As the realism increases, the "uncanny valley" effect occurs [7, 8]. However, in most application contexts it is the "suspension of disbelief" which is to be achieved, not the full realism. Further on, when using non-realistic faces there are additional means of expressivity (exaggeration, usage of non-realistic features or additional signals). Hence it is interesting to study possible cultural variations in perception of cartoon-like faces. Do findings from psychology on interpreting realistic facial expressions carry over to cartoon-like faces? How do the drawing style and familiarity with non-realistic facial expressions (e.g. in the tradition of comics) influence the interpretations of cartoon-like facial expressions? Koda's cross-cultural study [9] on the recognition of cartoon-like agent facial expressions drawn by Asian and Western designers suggests that the recognition accuracy of facial expressions is higher for virtual agents designed by the same cultural group as the subjects. E.g. Japanese cartoon-like facial expressions are recognized most accurately by Japanese, and Western cartoon-like facial expressions are recognized most accurately by western countries. The results suggest an in-group advantage is also applicable to cartoon-like facial expressions of virtual agents.

Recent psychological study also investigates the cultural differences of facial expressions by focusing on the facial regions. Research on human eye movements to interpret photo realistic human facial expressions showed East Asian participants mostly focused on the eyes, and Western participants scanned the whole face [10]. Yuki et al. used pictograms and photorealistic human facial images and suggest that Americans tend to interpret emotions based on the mouth, while Japanese tend to focus on the eyes [11]. Yuki et al. state this cultural difference arises from cultural norms: that people in cultures where emotional subduction is the norm (such as Japan) would focus on the eyes, and those in cultures where overt emotional expression is the norm (such as U.S.) would focus on the mouth shape.

This study applies the findings of [10, 11] to animated cartoon-like virtual agent faces to improve the culturally effective facial expression design of virtual agents. Such findings can be used not only to derive design guidelines when aiming at users of a single culture, but as adaptation strategies in applications with multicultural users. E.g. in an ATM machine or on-line shop, if the user's cultural identity is established; the virtual agent's facial expressions may be fine-tuned for optimal recognition for the given culture. Animating virtual agents' facial expression is rather easily done, but much more difficult in case of physical robots. Recent social robots have cartoonish faces with limited facial expressions, e.g., Kismet [12], Nexi MDS

Robot [13], iCat [14] when their research focus is not on increasing realism of a humanoid robot such as geminoids [8]. We believe providing research results on the perception of cartoonish virtual agents' facial expressions is also meaningful in order to minimize the effort to develop social robot's facial expressions.

We investigated cultural differences in using the eye and mouth regions as cues to recognize the facial expressions of cartoonish agent faces between Hungary and Japan. We conducted a web-based survey to confirm the following hypothesis. In cartoonish faces, Japanese weigh facial cues in the eye regions more heavily than Hungarians, who weigh facial cues in the mouth region more heavily than Japanese. Section 2 describes the facial expression design and experiment design, section 3 describes the results, section 4 discusses the results, and section 5 concludes the research.

2 Experiment

2.1 Design of Facial Expressions

Two Japanese designers designed the three agent faces shown in Fig. 1. Each face design has neutral, happy, and sad expressions. The examples of the three original facial expressions are shown in Fig. 2. The face and facial expressions were created using CharToon [15], a design and animation tool for 2D cartoon faces. The facial expressions were designed by taking the emotional expressions displayed in Ekman's FACS (Facial Action Coding System) training material. The facial features were discussed in case of happy and sad expression [16].

Pre-evaluation of the original expressions was conducted by ten Japanese and eight Hungarians to validate that each expression conveyed the intended emotions of the designer. The static images of happy and sad expressions of the three face designs were shown randomly to the evaluators after showing the neutral expression in each session. They were asked to select the perceived emotion of the each expression (happy or sad expression) from the following four adjectives: happy, sad, surprised, fear. They wrote an adjective if they don't find appropriate adjectives from the four provided. As a result of the pre-evaluation, all the original face designs had higher than 90% recognition accuracy among Japanese, higher than 85% among Hungarians, although we used cartoon faces and facial expressions designed by Japanese. Thus, we can assume the happy and sad expressions correctly convey the intended emotions in both countries. We also asked the perceived age of each agent face (discussed in section 4).

We then created six static expressions per agent design by combining the eyes and mouths. The six combinations are: happy eyes and neutral mouth (HN), happy eyes and sad mouth (HS), neutral eyes and happy mouth (NH), neutral eyes and sad mouth (NS), sad eyes and happy mouth (SH), and sad eyes and neutral mouth (SN). The total number of combined facial expressions is 18 (six expressions x three agents). Fig. 3 shows the six combined expressions created from the original expressions in Fig. 2.

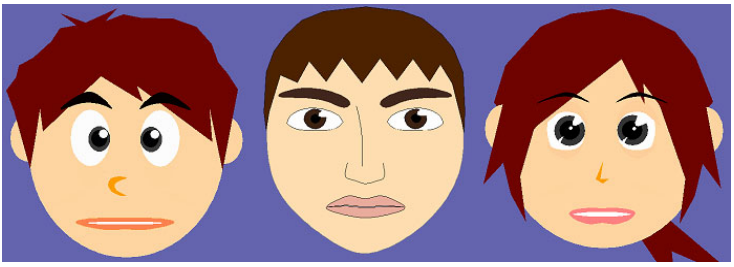


Fig. 1. Three agent designs created by two Japanese designers. Each design shows neutral expression. From left: Boy 1, Boy 2, Girl designs.

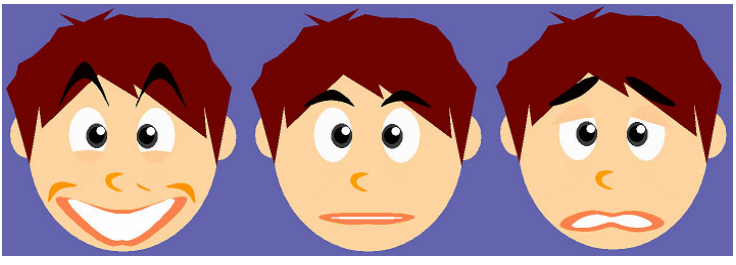


Fig. 2. Three original facial expressions of the Boy 1 design. From left: happy, neutral, and sad.

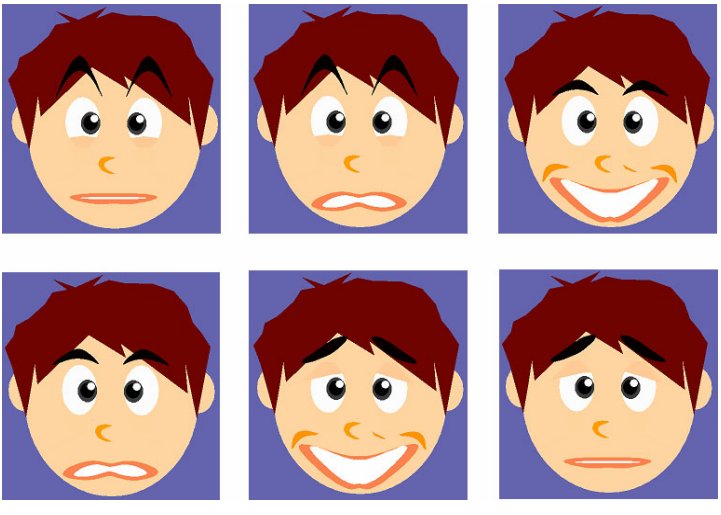


Fig. 3. Examples of six combined facial expressions created from images in Fig. 2. From top left: happy eyes and neutral mouth (HN), happy eyes and sad mouth (HS), neutral eyes and happy mouth (NH); from bottom left: neutral eyes and sad mouth (NS), sad eyes and happy mouth (SH), sad eyes and neutral mouth (SN).

Next, we created 18 animations that start from the neutral expressions of each agent design and end with one of the combined expressions by using CharToon's animation editor. Each animation lasted for four seconds; one second for a neutral expression, two seconds for the transition to a combined expression, and one second for the combined expression, and the animation ends with the combined expression. The animations were converted to the Adobe Flash movie format.

2.2 Experiment Procedure

The experiment was conducted on the web from December 2009 to January 2010. Subjects were invited to participate in the experiment by email. Invitations were made to gather participants with minimum exposure to and experiences with other cultures.

The experiment procedure was as follows:

1) Participants accessed the experiment web site. Instructions were made in Japanese for the Japanese participants and in English for the Hungarians.
2) Participants first read the instructions on the procedure. No explanation on the objective of the experiment was made.
3) They accept the terms and condition by clicking the accept button.
4) They answered a questionnaire about their demographics: name or nickname, age, gender, nationality, first language, and experiences of living abroad.
5) The 18 animations are randomly shown one by one. Subjects watched each animation by clicking the play button. They can replay the animation multiple times.
6) They evaluated the perceived emotion of the final expression of each animation using a 6-point Likert scale (6: very happy - 1: very sad). The instruction was made to follow their first impression upon rating.
7) The experiment continued until they finished all 18 animations. The estimated time to complete the experiment is approximately 5 minutes.
8) Participant answers (demographic information and ratings) were stored on a secured disk on the web server. Each answer was labeled with the name/nickname (which participants entered) and the timestamp when the participants clicked the accept button in order to differentiate participants with the same name or nickname.

We gathered 53 valid answers from Japan and 31 from Hungary. Valid participants resided in Hungary or Japan, spoke either Japanese or Hungarian as their first language, and had no experience or had lived abroad less than one year. The Hungarian participants are fluent in English. Incomplete answers that lack participants' demographic information or those did not finish evaluating the all 18 animations were not used for later analysis. The participant demographics of valid answers in Japan were 29 male and 24 female, whose average age is 17.9 years old (standard deviation (SD) 4.5 years). The demographics in Hungary were 15 male and 16 female, whose average age is 24.3 years old (SD=11.0 years).

Fig. 4 shows a screenshot of one of the animation and evaluation pages of the experiment.

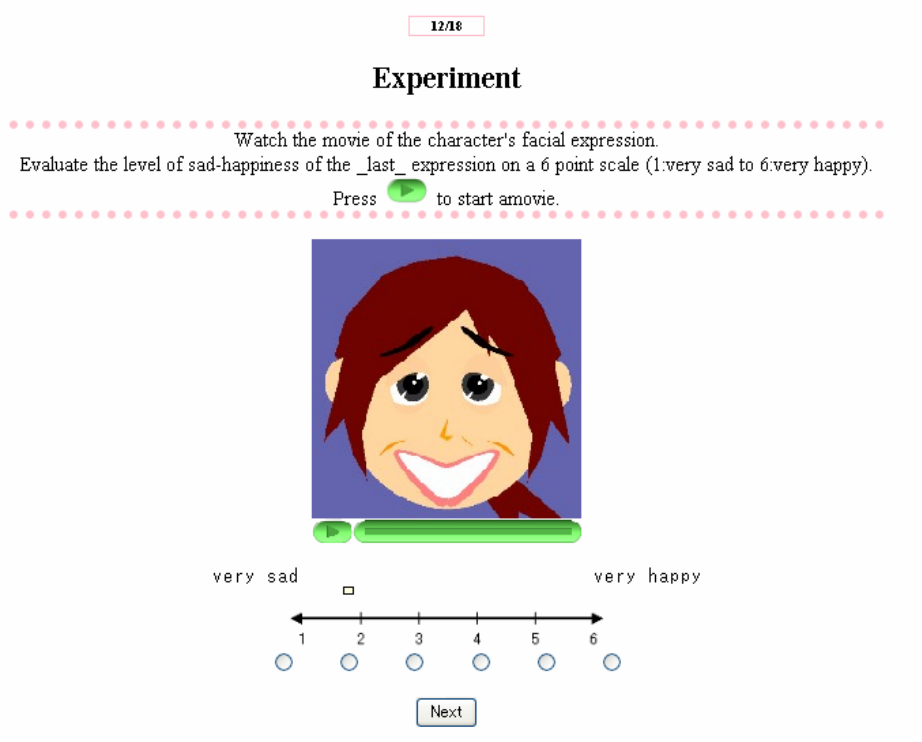


Fig. 4. Screenshot of web experiment. Flash movie shows animation from a neutral expression to one of 18 combined expressions shown randomly. Subjects rated perceived emotion of each combined expression on 6-point Likert scale.

3 Results

3.1 Analysis of Perceived Emotions by Facial Regions

The participant answers about the perceived emotions are shown in Fig. 5. By focusing on the shape of either the eye or mouth regions, the combined expressions were placed into four categories: happy eyes, sad eyes, happy mouth, and sad mouth. For example, the happy eyes category includes happy eyes and neutral mouth and happy eyes and sad mouth expressions.

First, we focused on the answers by facial categories regardless of the country. As shown in Fig. 5, the happy mouth category was rated the happiest and the sad mouth category was rated the saddest in all categories by both countries. The eye region (happy eyes/sad eyes) did not convey the intended emotions as effectively as the mouth region (happy mouth/sad mouth).

Second, we focused on the answers between Hungary and Japan. Japanese rated the happy eyes category as significantly happier than Hungarians regardless of the

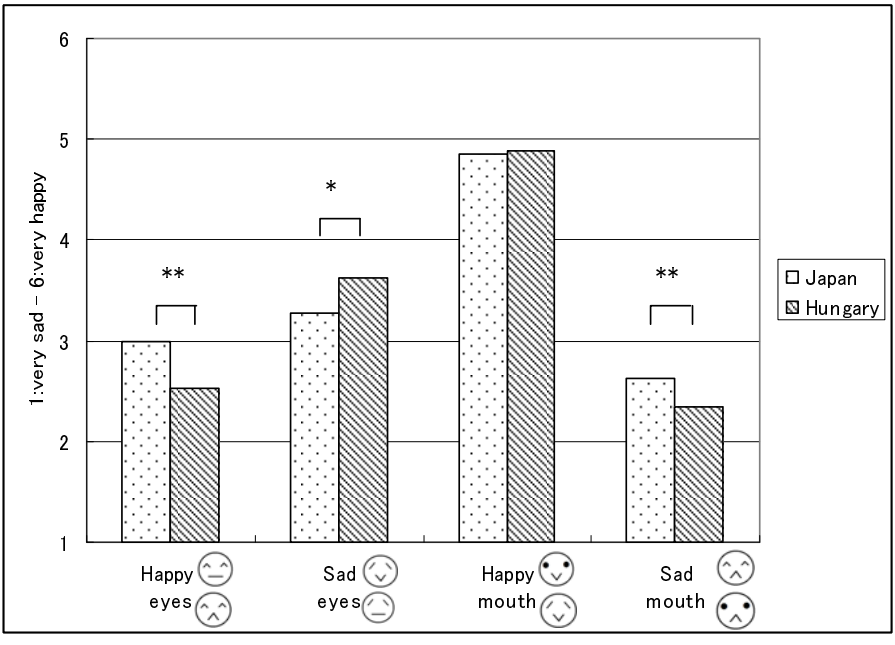


Fig. 5. Perceived emotions of Japan and Hungary categorized by facial regions. Scale indicates 1: very sad to 6: very happy. Number of subjects: Japan (n=53), Hungay (n=31). ** indicates $p<0.01$, * indicates $p<0.05$).

mouth's shape (Japan=3.00 (n=53), Hungary=2.52 (n=31), F=24.77, $p<0.01$) and the sad eyes category as sadder (Japan=3.27 (n=53), Hungary=3.63 (n=31), F=5.98, $p<0.05$). Hungarians rated the sad mouth category significantly sadder than Japanese regardless of the shape of the eyes (Japan=2.63 (n=53), Hungary=2.34 (n=31), F=14.59, $p<0.01$). There were no significant differences in the ratings of the happy mouth category (Japan=4.87 (n=53), Hungary=4.76 (n=31), F=0.13, not significant).

3.2 Analysis of Perceived Emotions by Combined Expressions

In this section, we examine the results by the six combined expressions for more detailed analysis. Fig. 6 shows the perceived emotions of the combined expressions. Analysis by Two-way ANOVA indicates significant cultural differences in the perceived expressions between Japan and Hungary in the HN (Japan=3.57 (n=53), Hungary=2.91 (n=31), F=27.31, $p<0.01$), SN (Japan=2.09 (n=53), Hungary=2.49 (n=31), F=10.74, $p<0.01$) and NS (Japan=2.82 (n=53), Hungary=2.56 (n=31), F=8.43, $p<0.05$) expressions. There were no other significances in the perceived emotions between Japan and Hungary.

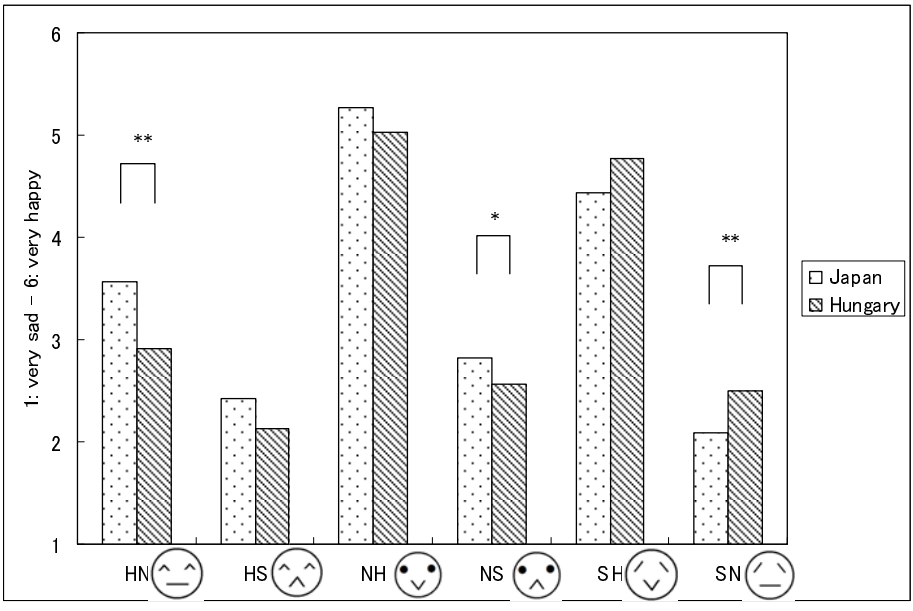


Fig. 6. Perceived emotions of six combined expressions in Japan and Hungary. Scale indicates 1: very sad to 6: very happy. Number of subjects: Japan (n=53) , Hungay (n=31). ** indicates $p<0.01$, * indicates $p<0.05$). Acronyms: HN: happy eyes and neutral mouth, HS: happy eyes and sad mouth, NH: neutral eyes and happy mouth, NS: neutral eyes and sad mouth, SH: sad eyes and happy mouth, SN: sad eyes and neutral mouth (SN).

3.3 Analysis of Perceived Emotions by the Shape of the Facial Regions

This section analyzes the changes of the perceived emotions within country by the shape of the facial regions. Table 1 shows the perceived emotions shown by the shape of the mouth region when the shape of the eye region is fixed. Table 2 shows the perceived emotions shown by the shape of the eye region when the shape of the mouth region is fixed.

The perceived emotions between HN and HS (mouth shape change when the eye region is fixed on happy eyes), and SN and SH (mouth shape change when the eye region is fixed on sad eyes) are significantly different in both countries (Table 1). This result again indicates that the shape of the mouth effectively displays the emotions of cartoon faces both in Japan and Hungary.

The emotions perceived by the shape of the eyes are not as consistent as those by the shape of the mouth (Table 2). The emotions perceived by Japanese differ significantly between NH and SH (eye shape change when the mouth region is fixed on a happy mouth), but the differences of emotions perceived by Hungarians are not significant. Contrary to designer intentions, Hungarians rated the perceived emotions of NS and HS (eye shape change when the mouth region is fixed on sad mouth) significantly differently, but the differences in emotions perceived by Japanese are not significant.

Table 1. Perceived emotions by shape of mouth region

Country	Design	Happy Eyes			Sad Eyes		
		HN 	HS 	p<0.01	SN 	SH 	p<0.01
Japan	boy 1	3.87	2.53	**	1.79	4.43	**
	boy 2	3.17	2.4	**	3.02	4.94	**
	girl	3.66	2.36	**	1.47	3.94	**
Hungary	boy 1	3.06	2.25	**	2.35	4.57	**
	boy 2	2.55	2.16	**	2.87	4.77	**
	girl	3.13	1.87	**	2.76	4.74	**

Table 2. Perceived emotions by shape of eye region

Country	Design	Happy mouth			Sad mouth		
		NH 	SH 	p<0.01	NS 	HS 	p<0.01
Japan	boy 1	5.47	4.43	**	3	2.53	
	boy 2	4.75	4.94		2.75	2.4	
	girl	5.58	3.94	**	2.72	2.36	
Hungary	boy 1	5.32	4.77		2.48	2.35	
	boy 2	4.32	4.77		2.58	2.16	**(opposite)
	girl	5.42	4.74		2.61	1.87	**(opposite)

4 Discussion

First, we examined the results by categorizing the facial regions. The results from 3.1 suggest cultural differences in the perceived emotions in the happy eyes, sad eyes, and sad mouth categories in the expected direction. The highest happiness ratings in the happy mouth category in both countries suggest that the mouth region more effectively conveys the emotions of the facial expressions than the eye region. The mouth's effectiveness is understandable since the mouth is the most expressive part of the face, since it is evolved as a primary means of verbal communication [17, 18] while the eyes are more difficult to control than the mouth when people express emotions.

Second, we examined the results by the combined expressions. The results from 3.2 again indicated cultural differences in the expected direction in the hypothesis.

Japanese rated the perceived emotions of HN (happy eyes and neutral mouth) significantly happier than Hungarians and SN (sad eyes and neutral mouth) as significantly sadder than Hungarians. Hungarians rated the perceived emotions of NS (neutral eyes and sad mouth) significantly sadder than Japanese. This means the expressions where either eyes or mouth is/are neutral confirmed the hypothesis, except NH (neutral eyes and happy mouth). However, the contradictory expressions (HS, SH), where the eye expressions and mouth expressions show opposite expressions, fail to show significant cultural differences.

The effectiveness of the mouth region is shown again in NH and SH (both with happy mouths). Both countries' perceived emotions of NH have the happiest scores. SH's score are higher than HN.

Third, we focused on the differences in ratings by the shape of the eye/mouth when the other facial region was fixed. Both countries responded to the differences in the mouth shape as significantly and dynamically as expected. However, Japanese responded to the differences of the eye shape more significantly and dynamically than Hungarians, implying that Japanese focused more strongly on the eyes even when the mouth region has the most convincing happy expression; Hungarians continued to focus on the mouth region.

However, the differences in the eye shape with a sad mouth did not result in the expected direction. Perhaps the reason reflects the nature of the effectiveness of the mouth region in facial expression recognition and eye designs. Some subjects commented that the happy eye designs resembled surprised expressions when isolated from the mouth region. Although the perceived emotions of the happy eyes using static images of the eye region in the pre-evaluation test had 90% in Japan and 85% accuracy in Hungary, the result implies we must design the eyes more carefully when they are used solely apart from other facial regions.

Former studies [10, 11] used emoticons and photorealistic human facial expressions. Emoticons can convey emotions simply and effectively because we use them in our daily lives. Since the photorealistic human facial expressions used in these studies were made by professional actors, they naturally and effectively conveyed the intended emotions. However, cartoonish facial designs have wide variations caused by the designers' drawing styles and techniques. Although we used CharToon as a design tool to minimize possible variations, the agent faces in the experiment had more room for improvement.

Another consideration should be put on cultural diversity in designing cartoon faces. Koda reports in [9] that the recognition accuracy of facial expressions is higher for virtual agents designed by the same cultural group as the subjects. The facial expressions used in the experiment were designed by Japanese, thus the drawings followed Japanese stylistic conventions for expressing emotions, and Japanese have more exposure and experience in such drawings. Although the pre-evaluation of facial expressions did not show significant cultural differences, we cannot exclude the possibility that cultural differences in recognizing expressions in cartoon faces might be a result of cultural differences in drawing cartoon faces. We had an interesting result from the pre-evaluation of the agent faces. Hungarians' perceived age of the agent faces used in the experiment are much younger than the Japanese ones. Hungarian perceived the age of the agent faces as in their teens or younger, and Japanese as in their twenties.

In order to investigate the cultural differences in recognizing cartoon facial expressions in both ways, we plan to conduct a subsequent experiment with Hungarian designed cartoon faces. We expect to have different drawing styles of virtual agent faces, since Hungary has not been exposed to Japanese comic/anime culture compared to other European/American countries. Evaluation from other countries than Hungary and Japan is also needed to eliminate in-group advantage within a country or a culture.

5 Conclusion

This paper applies the psychological findings on cultural differences on facial regions used as cues to recognize human facial expressions to the cases of animated virtual agent faces. Our preliminary results support the hypothesis. There are cultural differences when using facial regions as cues to recognize cartoonish facial expressions between Hungary and Japan. Japanese weighed facial cues in the eye regions more heavily than Hungarians, who weighed facial cues in the mouth region more heavily than Japanese. We also confirmed that, regardless of the country, the mouth region more effectively conveys the emotions of facial expressions than the eye region.

We believe the results can be used not only to derive facial expression design guidelines of virtual agents when aiming at users of a single culture, but as adaptation strategies in applications with multicultural users. The virtual agent's facial expressions may be fine-tuned for optimal recognition for the given culture. The results also can be applied to designing a physical robot facial expressions to minimize mechanical movements to create its facial expressions according to the culture the robot serves.

Acknowledgements. This research is supported by a Grant-in-Aid for Scientific Research (C) 20500196 (2008-2010) from the Japan Society for the Promotion of Science. CharToon was developed by Zsofia Ruttkay and Han Noot.

References

1. Rickel, J., Marsella, S., Gratch, J., Hill, R., Traum, D., Swartout, B.: Towards a New Generation of Virtual Humans for Interactive Experiences. IEEE Intelligent Systems, 3–38 (July/August 2002)
2. Ekman, P., Friesen, W.V.: Constants Across Cultures in The Face And Emotion. Journal of Personality and Social Psychology 17, 124–129 (1971)
3. Ekman, P., Friesen, W.V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W.A., Pitcairn, T., Ricci-Bitti, P.E., Scherer, K.R., Tomita, M., Tzavaras, A.: Universals and Cultural Differences in The Judgment of Facial Expressions of Emotion. Journal of Personality and Social Psychology 53, 712–717 (1987)
4. Matsumoto, D., Ekman, P.: American-Japanese Cultural Differences in Intensity Ratings. Motivation and Emotion 13, 143–157 (1989)

5. Elfenbein, H., Beaupré, M., Leveque, M., Hess, U.: Toward a Dialect Theory: Cultural Differences in Expressing and Recognizing Facial Expressions. Emotion 7, 131–146 (2007)
6. Ruttkay, Z.: Cultural Dialects of Real and Synthetic Emotional Facial Expressions. Journal of AI & Society 24(3), 307–315 (2009)
7. Mori, M.: Bukimi no Tani: The Uncanny Valley (Translated by MacDorman, K.F., Minato, T.). Energy 7(4), 33–35 (1970)
8. Ishiguro, H.: Android Science: Toward a New Cross-disciplinary Framework. In: CogSci 2005 Workshop: Toward Social Mechanisms of Android Science, pp. 1–6 (2005)
9. Koda, T., Ishida, T., Rehm, M., Andre, E.: Avatar Culture: Cross-cultural Evaluations of Avatar Facial Expressions. Journal of AI & Society 24(3), 237–250 (2009)
10. Jack, R.E., Blais, C., Scheepers, C., Schyns, P.G., Caldara, R.: Cultural Confusions Show Facial Expressions are Not Universal. Current Biology 19(18), 1543–1548 (2009)
11. Yuki, M., Maddux, W.W., Masuda, T.: Are the Windows to the Soul the Same in the East and West? Journal of Experimental Social Psychology 43, 30–311 (2007)
12. Breazeal, C.: Emotion and Sociable Humanoid Robots. International Journal of Human-Computer Studies 59(1-2), 11–155 (2003)
13. Breazeal, C.: Sociable Robots. Journal of the Robotics Society of Japan 24(5), 591–593 (2007)
14. Van Breemen, A.J.N.: ICat: Experimenting with Animabotics. In: AISB Creative Robotics Symposium (2005)
15. Ruttkay, Z., Noot, H.: Animated CharToon Faces. In: International Symposium on Non-photorealistic Animation and Rendering, pp. 91–100 (2000)
16. Ekman, P., Friesen, W.V.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto (1978)
17. Ekman, P., Friesen, W.V.: Unmasking the Face: A Guide to Recognizing Emotions from Facial Cues. Prentice Hall, Englewood Cliffs (1975)
18. Fridlund, A.J.: Human Facial Expression: An Evolutionary Review. Academic Press, San Diego (1994)

Availability of Multilingual Chat Communication in 3D Online Virtual Space

Takashi Yoshino and Katsuya Ikenobu

Faculty of Engineering, Wakayama University,
930 Sakaedani, Wakayama, Japan
yoshino@sys.wakayama-u.ac.jp
http://www.wakayama-u.ac.jp/~yoshino/

Abstract. There is a possibility that a 3D online virtual space can become a more familiar communication medium because of the widespread use of the Internet. However, language differences pose significant barriers to intercultural communications. We have developed a multilingual chat communication support system in Second Life as an initial research. Second Life is one of the most popular 3D online virtual spaces in the world. We have carried out a multilingual chat communication experiment in Second Life. From this experiment, we have found the following. (1) It is important that a machine translation machine make it visible in a 3D virtual space. A user can easily learn the usage of a machine translation. (2) There is not understanding other party's language in the problem because appearances of each avatar are free. It is important for users to show translation language pairs not to give confusion.

Keywords: multilingual chat communication, 3D online virtual space, machine translation.

1 Introduction

The number of accounts on Second Life[1], which is a three-dimensional virtual space, has increased to more than 14 million. Approximately 51% of these users are English-speakers[2], and English is used even in chats with non-English speakers. However, there are many users who cannot speak English well, and therefore, language can become a communication barrier.

Several support methods for communication in a virtual space have been proposed[3,4]. These researches aim to enable communication by using voice conversations in the same language. Research on multilingual chat communications support in a virtual space has been carried out a little. The Internet and technologies supported by high-performance computers have enabled communication in the virtual world for the first time.

This research presents the development of a multilingual chat communications support system in a three-dimensional virtual space. Further, we show the result of experiments that use multilingual chat communications support in a virtual space.

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 28–40, 2010.

2 Related Works

Sugawara et al. researched communication in a three-dimensional virtual space [5]. They developed a three-dimensional virtual space, called InterSpace, for their research. In InterSpace, there is an agent who mediates communications in a virtual space. Asynchronous communications in a virtual space have been achieved by this agent. Then, an open experiment was carried out on the Internet using InterSpace. The effect of the development of communities and the use of the communications media was studied in this experiment.

Matsuda et al. developed a virtual society, called PAW, and carried out an open experiment [6]. From the result of the open experiment, they presented the user's profile, the effect of an event, the feature of communities, and the communication tools for using all these together.

Nakanishi et al. studied the impact of a three-dimensional virtual space communication environment [7]. From a comparison of this environment with a face-to-face environment, they have found that the environment of the three-dimensional virtual space contributes to an increase in the switch frequency of the speaker.

Zhang and Bach have developed a context-aware machine translation system, named Virtual Babel, for virtual worlds [8]. They enhance the Second Life viewer to intercept the incoming/outgoing chat messages and reroute the message to a machine translation. The returned translations are appended to the original text message to help users to understand the foreign language. They have the same issue as us. They provide a context-aware language model to estimate a suitable domain for chat.

Our research differs from the abovementioned researches; we discuss the possibility of communication among speakers of several different languages by using machine translation in a virtual space.

3 Multilingual Chat Support System

We have developed a multilingual chat support system by using an object[1] that could be built in an event-driven-type script in a three-dimensional virtual space. Portable and fixed translation systems are expected as multilingual chat communications support. In the following subsections, we discuss the support form in the three-dimensional virtual space and then describe the portable and fixed translation systems.

3.1 Support Form of Multilingual Chat Support System

Table 1 shows the advantages and the disadvantages of a fixed and a portable communication support system in a virtual space.

In the case of the fixed system, the translation object was set up in a certain pre-decided place. When a user (avatar) went to this place, he/she could use the

[1] An "object" is a form combined by "prim," which is the basic object in Second Life.

Table 1. Advantages and disadvantages of a fixed and a portable communication support system in a virtual space

	Advantages	Disadvantages
Fixed	- It is easy to manage and update software because the software is in an object. - The language of the other person is shown clearly by the bench. - It is easy for users to operate the avatar because a user can use translation only while sitting on the bench. - The translation bench can be shown clearly using a translation function.	- The translation chat can be carried out only in that particular area. - One translation object can be translated only in one pair of languages.
Portable	- A user can use the translation function everywhere. - One HUD can support multiple language pairs.	- A user needs to update the translation system by himself/herself because the translation function is in the object. - The language of the other person is not shown clearly. A user needs to ask his/her language in advance. - A user has to operate the HUD for translation. - Because the user is not aware that the other person is using a machine translation function, there could be a lot of confusion.

translation chat system. It was necessary to prepare a separate translation object for each translation language for such a system[2]. The advantage of this system was that the other party's language was specified by the place where it sat.

The operation "Sit an avatar" is an easy operation that is often used in Second Life. The fixed machine translation function was displayed in the three-dimensional virtual space. Therefore, other users could recognize that the other party is using the machine translation system. In the case of the portable system, the translation object was displayed on the user's screen as a head-up display (HUD)[3]. This feature of the portable system did not place limitations on where the user could use the translation functions, and communications that used translation chat could be carried out everywhere. It was necessary to select the translation language for a portable system on HUD; further, it was necessary to know

[2] This limitation is a problem in Second Life.

[3] A head-up display (HUD) is a transparent display that presents information without requiring the user to look away from his/her usual viewpoint.

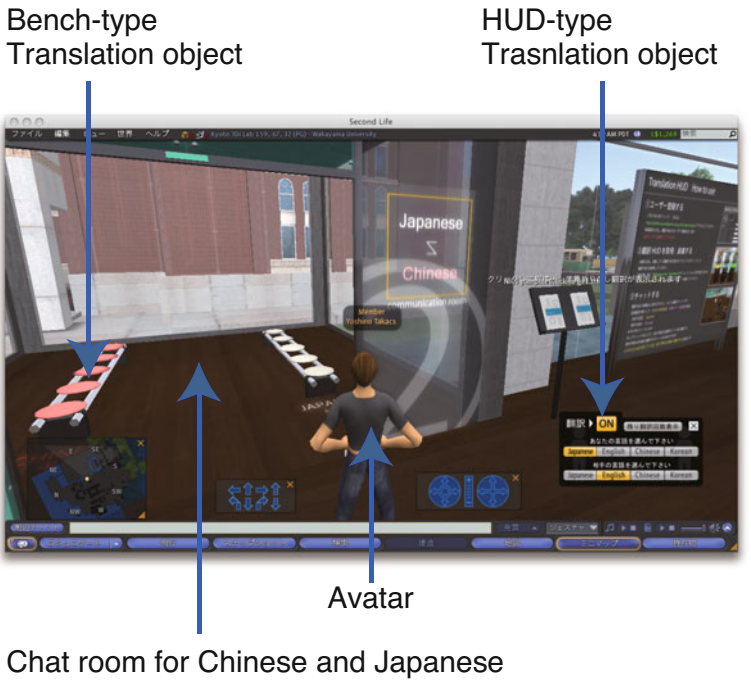


Fig. 1. Screenshot of the multilingual chat support system

the other party's language beforehand[4]. When a portable system was used, sentences in two or more languages were suddenly displayed while chatting because the machine translation function was not displayed in the three-dimensional virtual space. This might confuse the other party.

3.2 Structure of Multilingual Chat Support System

The multilingual chat support system that we have developed is shown in Figure 1.

1. Fixed translation system (Bench-type translation object)
   "Chat room for Japanese and Chinese" in Figure 1 is a fixed translation object. We set up Japanese–English, Japanese–Chinese, and Japanese–Korean chat rooms. The procedure for using the translation chat room was as follows:
   (a) A user clicked the bench-type translation object in the translation chat room, and then the avatar of the user sat on the bench.

[4] When this system was developed, neither the user's national origin nor information on the user's language was offered in Second Life. However, now, a user's language can be known.

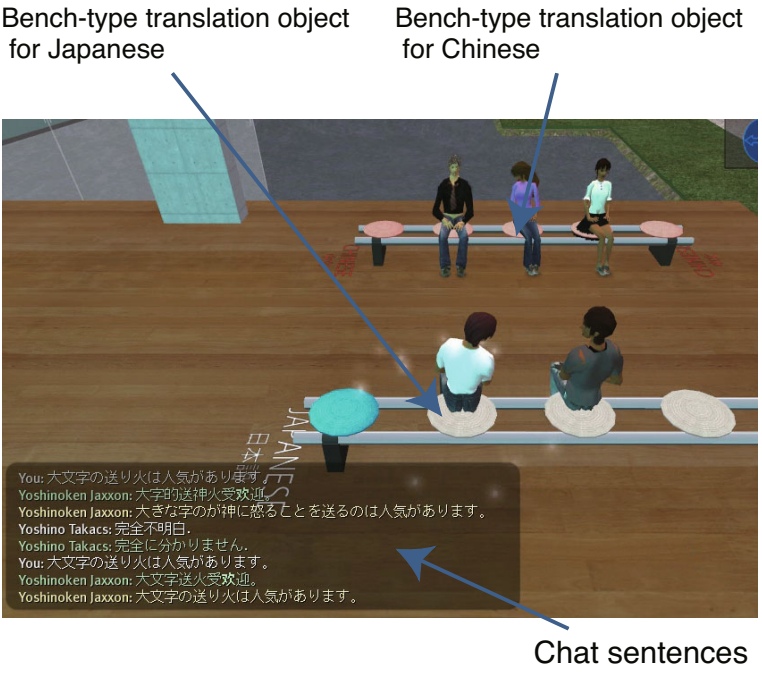


Fig. 2. Screenshot of a chat communication in the translation chat room

(b) The chat sentence was translated automatically during the chat.

The appearance of the chat communication in the chat room is shown in Figure 2. Each user sat on the bench on the native language side. The input chat sentences were machine translated and displayed as a chat sentence.

2. Portable translation system (HUD-type translation object)
   The HUD-type translation object in Figure 1 is a portable translation object. Figure 1 shows an expanding HUD-type translation object. This could translate among Japanese, English, Chinese, and Korean by switching the language tab. The procedure for using the HUD-type translation object was as follows:
   (a) The user installed the HUD-type translation object in his/her avatar.
   (b) The user clicked the language selection tab according to the partner's language and his/her own language.
   (c) The chat sentence was translated automatically during the chat.

3.3 Back Translation

Back translation is the process of translating a sentence that has already been translated into an intermediate language back to the original language. Back translation enables users to check the accuracy of a message using their own native language.

Fig. 3. Screenshot of the heads-up display (HUD)-type translation chat object

Fig. 4. Example of multilingual chat communication

An example conversation is shown in Figure 4. The translation sentence was displayed only on the input user's side and not on the other party's side. When the input sentence had a mistake, the sentence could not be machine-translated correctly. When an unknown word or a complex syntax was used even when the input sentence was correct, the translation would most likely be incorrect. In previous studies, the user confirmed the back-translation result before he/she presented it to the other party's side [9]. When the user judges that there was a problem in the back-translation result, he/she can correct the input sentence [10,11].

A chat in a virtual space communication demanded real-time translation. Therefore, the user could correct the sentence if he/she found that the translation result was not correct.

3.4 System Configuration

Figure 5 shows the system configuration of the multilingual chat support system. The input sentence that the translation object (Bench-type translation object in the translation chat room and the HUD-type object) received from the Second Life server passed through the relay server (Figure 5(B)).

The relay server authenticated the user, limited the translation frequency, and recorded the user log. Then, the input sentence was sent to the translation server (Figure 5 (D)) through the language grid [12] server (Figure 5 (C)).

The relay server sent the translated sentence to the Second Life server (Figure 5 (B)). Finally, the translation object displayed the translated sentence in Second Life. The back translation was carried out using an almost similar procedure.

4 Experiment 1: Portable Multilingual Chat Communication System

We carried out an experiment using the portable multilingual chat communication system. In the experiment, five Japanese university students chatted with some unspecified general users in Second Life.

The purpose of the experiment was to investigate what problems occurred during the use of the portable multilingual chat communication system in Second Life. The procedure of the experiment was as follows:

1. A subject installed the portable translation system (HUD-type translation object) in his/her avatar.

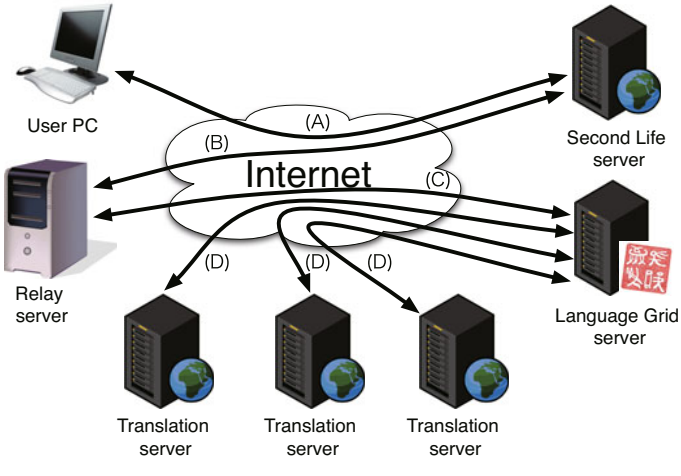


Fig. 5. System configuration of the multilingual chat support system

Table 2. Questionnaire result of the experiment using a portable translation system

Questionnaire items	Average	Standard deviation
(1) The translation result of my chat sentence was correct.	2.2	0.16
(2) I think that the back translation is necessary.	4.6	0.24
(3) I understood the meaning of the foreigner's (the other party of the chat) sentence from the translation.	2.8	0.56
(4) I am satisfied with the accuracy of the translation.	2.0	0.40
(5) I understood the usage of the translation function at once.	4.0	0.40
(6) A problem occurred in the translation system while chatting.	2.2	1.36
(7) I think that the machine translation in the chat is useful.	3.6	1.04

We used a five-point Likert scale for the evaluation, with 1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, and 5 = Strongly agree.

2. The subject went to an island (called SIM in Second Life) where the other avatar existed.
3. The subject read a conversation among the general users on Second Life and chatted to them.

The duration of each experiment was approximately 20 or 30 min.

In the experiment, the translation system was not able to translate sentences in many cases. Many users in Second Life used clipped forms for chat. Some users incorrectly used capital and small letters while spelling out a word. Moreover, even if the chat with a general user on Second Life was translated, it was not possible to follow the conversation because the translation could not keep up with the speed of the conversation. Subjects were ignored by almost everyone on Second Life.

Table 2 shows the questionnaire result of the experiment using the portable translation system.

We obtained good results with respect to the necessity and ease-of-use of the system from the questionnaire (Table 2 (2), (5), and (7)). However, the results with respect to the actual use of the system in Second Life were poor (Table 2 (1), (3), and (4)). We found that the sentences of the other party and those of the subject were often translated incorrectly. The main reasons that the subjects gave for not being able to carry out a good chat conversation (Table 2 (3)) were as follows:

- I hardly understood the content of the chat. However, I think that the translation depends on the content of the sentence.
- I think a peculiar word was often used in the chat ('You' is shown 'u'). I did not know what was said at all because the contents of the chat were too unusual.
- Because the translation results were not too correct, I could not understand the conversation well.

We found that the chat in Second Life was similar to the chats on the Internet. The users did not care much about grammar and spelling. When the subjects used the portable translation system, the other party did not answer in consideration of such a special situation.

5 Experiment 2: Fixed Multilingual Chat Communication System

5.1 Experimental Procedure

We carried out the experiment using the fixed multilingual chat communication system. The purpose of the experiment was to investigate what problems occurred during the use of the fixed multilingual chat communication system in Second Life.

In the conversation experiment, we verified the following items:

- Can you carry out a serious conversation using this system?
- Does the meaning of the other party's sentence change because of the machine translation?
- How can we verify that the machine translation function is effective?

From the result of the experiment using the portable system, it was concluded that most of the conversation was not established. We thought that this was because of the following two reasons:

- The other party did not understand the purpose of the conversation.
- The conversation was not meant for machine translation.

Therefore, we gave the subject a purpose "Question your partner by using machine translation." The subjects were seven Second Life users whose native language was English. We recruited the experimental subjects from the official Web page of Second Life. We told the participating subjects that they would be paid 10L$‘30L$ in Second Life currency as a reward.

The procedure of the experiment was as follows:

1. We guided the subject to the translation chat room.
2. We told the subject that this was an experiment to verify the effectiveness of machine translation.
3. We asked the subject to ask a Japanese interlocutor 10 different questions.
4. The Japanese interlocutor requested rephrasing when the question could not be understood from the translation result.
5. When the translation of the rephrased sentence also did not make much sense, the Japanese interlocutor asked the subject to change the question.

The Japanese interlocutor was one of the authors. In the experiment, the subjects did not use back translation. They could only see their input sentences and the translated sentences. We screen-captured the entire experiment. We recorded all the sentences that were exchanged during the chat.

Table 3. Judgment result of the Japanese interlocutor's understanding of the subjects' question

Subjects	Question of subject			
	Apprehensible	Inapprehensible	Average number of characters	Average number of words
A	7	3	74.8	14.6
B	10	0	39.5	7.6
C	10	0	31.9	6.2
D	8	2	53.1	10.0
E	9	1	57.9	10.7
F	9	1	46.7	9.9
G	10	0	66.2	13.1
Average	9.0	1.0	52.9	10.3

5.2 Result and Discussion

Table 3 shows the Japanese interlocutor's understanding of the subjects' questions. The Japanese interlocutor judged whether a translated question was understandable. The subjects were asked to rephrase questions that the Japanese interlocutor did not understand. From Table 3, it could be concluded that the translation result for many of the questions was understandable.

Some of the problems that hindered accurate translation, and hence the overall chat communication, were as follows:

1. The subjects hardly used clipped forms in the chat. We had explained to the subjects that the experiment was about verifying the effectiveness of machine translation in advance. However, we did not explain the features of machine translation and the problems caused by the use of clipped forms.
2. The subjects often made spelling mistakes. Therefore, the machine translation failed. A couple of examples are shown below.
   Input sentence: Who are your favorate singers/music groups?
   Translation result: favorate/yO[vNH
   (The spelling of "favorate "is incorrect.)
   Input sentence: It is a very well desgined welcome area.
   Translation result: desginedGAB
   (The spelling of "desgined "is incorrect.)
3. When the Japanese interlocutor said that he did not understand the translated sentence, the subjects rephrased the sentence. The subject followed the following rephrasing method:
   – The subject numbered each sentence.
   – The subject supplemented the meaning by using parentheses.
   Some of the examples are as follows:
   First input sentence: I am making up questions, it is ok. Next question: This skirt comes in Copy/No Mod/No Trans, can you make the skirt in Copy/Mod/No Trans? I am small, so I have to edit the skirt.
   Second input sentence: (1) I am making up questions, it is ok. (2) This skirt

comes in Copy/No Mod/No Trans, can you make the skirt in Copy/Mod/No Trans? I am small, so I have to edit the skirt.
(After the subject placed (1) and (2) at the beginning of the sentence, he asked the Japanese interlocutor to specify which sentence was poorly translated.)
Third input sentence: (1) This skirt (item of clothing) comes in (has the Second Life Permissions) of Copy/No Modify/No Transfer. Can you (the creator) change the Second Life Permissions to Copy/Modify/No Transfer. (3)I (Kokoro) am a small person, so I have to edit the skirt (item of clothing)[5].
(The subject explained the meaning of the word by using parentheses.)

4. The subjects divided a long sentence.
   First input sentence: When I am speaking (typing) to a Japanese person, what symbol should I use to make each idea/statement easier to seperate?
   Second input sentence: The English tend to put many subjects in their statements. Is there a symbol that would make it easier for you to understand the change of thought?
5. Colloquialism
   When the subjects used colloquial words, the sentence could not be machine-translated well.
   Example: Yay! I look forward to that. Do you enjoy flying or would you rather teleport all the time?
   ("Yay" is colloquialism term that means the affirmative.)
6. The subject used some emoticons.
   Example: And bad spelling does not help :)
7. The foreigner subject asked only simple questions for the conversation experiment. In particular, all questions asked by subjects B and C were simple (Table 3).
   Example 1: What is your name?
   Example 2: What kind of car do you drive?
8. There were subjects who did not begin their sentences with a capital letter. In the case of subjects C and D (Table 3), most of the sentences in their conversations began with a small letter. However, this did not affect the machine translation, and the sentences were translated correctly.
   Example 1: how is the weather there?
   Example 2: what do you think is the biggest source of conflict between most couples?
9. When a subject noticed a mistake in the input, he/she would input only the corrected form of the incorrect word in the next line.
   First input sentence: can you tell mehow many cans of beer would be left in a carton of 24 cans, if I drank eight cans of beer?
   Second input sentence: me how*
   (In the first input sentence, he input "me how " as "mehow")

[5] There is no '(2)' as is written

6 Conclusion

We have developed a multilingual chat communication support system in Second Life as an initial research. We carried out a multilingual chat communication experiment in Second Life.

From the experiments, we found the following:

(1) It is important that a machine translation machine make it visible in a 3D virtual space. A user can easily learn the usage of a machine translation.
(2) There is not understanding other party's language in the problem because appearances of each avatar are free. It is important for users to show translation language pairs not to give confusion.

The present machine translation could not handle colloquialisms well. However, it was possible to talk enough if there is motivation of communications.

Currently, the multilingual chat communication support system in Second Life has been opened to the public as a trial at the following URL:
http://slurl.com/secondlife/3D%20Community%20Lab/209/140/32

Acknowledgment. We would like to thank the reviewers for the careful consideration of the paper and for constructive comments and suggestions made to improve the paper. This work was partly supported by a Grant-in-Aid for Scientific Research (B) (No. 19300036, 2007-2009), and by the Okawa foundation for information and telecommunications(2008).

References

1. Linden Lab: Second Life, http://jp.secondlife.com/
2. Second Life - Economic Statistics: Key metrics, http://static.secondlife.com/economy/stats_200807.xls
3. Minoura, D., Ishibashi, S.: A 3-D Virtual Communication Environment for Thousands of Users. Transactions of Information Processing Society of Japan 42(11), 2595–2606 (2001) (in Japanese)
4. Isbister, K., Nakanishi, H., Ishida, T., Nass, C.: Helper agent: designing an assistant for human-human interaction in a virtual meeting space. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 57–64 (2000)
5. Sugawara, S., Suzuki, G., Nagashima, Y., Matsuura, M., Tanigawa, H., Moriuchi, M.: InterSpace: Networked Virtual World for Visual Communication. IEICE Transactions on Information and Systems E77-D(12), 1344–1349 (1994)
6. Matsuda, K., Miyake, T., Kawai, H.: Culture formation and its issues in personal agent-oriented virtual society: PAW^2. In: Proceedings of the 4th International Conference on Collaborative Virtual Environments, pp. 17–24 (2002)
7. Nakanishi, H., Yoshida, C., Nishimura, T., Ishida, T.: FreeWalk: A 3D Virtual Space for Casual Meetings. IEEE Multimedia 6(2), 20–28 (1999)
8. Ying, Z., Bach, N.: Virtual Babel: Towards Context-Aware Machine Translation in Virtual Worlds. In: Proceedings of the 12th Machine Translation Summit MT Summit XII (August 2009)

9. Yoshino, T., Fujii, K., Shigenobu, T.: Availability of Web Information for Intercultural Communication. In: Ho, T.-B., Zhou, Z.-H. (eds.) PRICAI 2008. LNCS (LNAI), vol. 5351, pp. 923–932. Springer, Heidelberg (2008)
10. Miyabe, M., Yoshino, T., Shigenobu, T.: Effects of Undertaking Translation Repair using Back Translation. In: Proceedings of the 2009 ACM International Workshop on Intercultural Collaboration (IWIC 2009), pp. 33–40 (2009)
11. Miyabe, M., Yoshino, T.: Accuracy Evaluation of Sentences Translated to Intermediate Language in Back Translation. In: Proceedings of 3rd International Universal Communication Symposium (IUCS 2009), pp. 30–35 (2009)
12. Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration. In: Proceedings of IEEE/IPSJ Symposium on Applications and the Internet, pp. 96–100 (2006)

Capture and Express Behavior Environment (CEBE) for Realizing Enculturating Human-Agent Interaction

Yoshimasa Ohmoto, Akihiro Takahashi, Hiroki Ohashi, and Toyoaki Nishida

Graduate school of Informatics, Kyoto University
Yoshidahonmachi, Sakyo-ku, Kyoto-shi, Kyoto-hu, Japan
{ohmoto@,a.takahashi@ii.ist,ohashi@ii.ist,nishida@}i.kyoto-u.ac.jp

Abstract. We are studying how Embodied Conversational Agents (ECAs) express communication behavior with cultural background. The objective of this study is the proposition of the modified Capture and Express Behavior Environment (CEBE) in which a person can interact with ECAs controlled by the captured behavior of another person with cultural background. In this paper, we discuss modifications and concepts of CEBE to apply CEBE for investigations to realize an ECA with cultural background. The prototype system could capture basic human behavior, such as head direction, posture of the upper body, and 3D angles of arms, when each part of the body, such as head, hands, arms and trunk. In addition, the system could control a robot or a virtual agent based on the detected data. We have to develop some implementations to interact with people with cultural background.

Keywords: human-agent interaction, measuring system, WOZ system, Robot operation interface.

1 Introduction

In recent years, there has been a growing exchange in goods, information and people in the world amidst the progression of internationalization. In cross-cultural exchange, people can relativize their own values and obtain new values from different cultures. Although there are some cross-cultural differences in habits and rules, the experience of recognizing these differences gradually is important in acquiring communication skill. The reason why people can get though them is that they have great abilities to learn, adapt and communicate with other people.

On the other hand, Embodied Conversational Agents (ECAs) which are expected to provide an effective interface for many users, including computer novices in diverse fields ([1], [9], [12]), however, have such ability in the distant future. It is important to interact with people in the fields. When ECAs interact with people, the cross-cultural features influence human-agent interactions. ECAs can have databases of cross-cultural rules and habits which are known well in advance. There are however body motions, behavior, and mental stances

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 41–54, 2010.

which influence cross-cultural interactions, and it is hard to describe them as rules or habits. These differences appear not only in languages and their use, but also in facial expressions, face directions, gestures, the range of movements, postures, interpersonal distance, and so on [8].

We are studying how ECAs express communication behavior which is not known as rules in a culture, such as emotional representation, the degree of agreement or disagreement, and commitment to the communication, based on cross-cultural background. In this paper, "cultural background" means meaningful customs, habits, gestures, nonverbal information and mental stances in a particular culture. We expect that the cultural background influences communication behavior. For the studies, we pay attention to gaze (head orientation), postures, hand gestures and paralinguistic features (pitch and power of voice). We will investigate them in a negotiation situation in which people have to express their opinions and speculate the inner states of their communication partner.

Cultural behavior is observed in diverse situations and representations. Different layers of phenomena exist that are influenced by cultural aspects such as the verbal and nonverbal behavior, appearance, proxemics, learning strategies and many more. We thus have to consider different appearance and physical limitations when ECAs express cultural behavior. Moreover, it is hard to pinpoint down the constituents of culture and its effect on interaction [13]. Therefore, we have to design ECAs based on data which is obtained in actual human-agent interaction. It is hard to apply the foundations of cross-cultural communication by analyzing data obtained in human-human communication to ECAs which have different appearance and physical limitations than humans.

We, however, did not have an environment to obtain the data in actual human-agent interaction. In other words, we need to analyze enculturated agent-human interaction in an environment in which a human can interact with another human who has the appearance and physical limitations of ECAs.

The purpose of this study is to propose an environment in which a person can interact with ECAs controlled by captured behavior of another person with cultural background, so that we can analyze the data obtained in actual enculturated agent-human interaction. The proposed environment is called Capture and Express Behavior Environment (CEBE). Human behavior in CEBE is expressed by an ECA which has a different appearance and physical limitations from a human. If the ECA could express enculturated behavior by CEBE, a standalone ECA which has the same appearance and physical limitations could express enculturated behavior without pinpointing down the constituents. Therefore, we expect that CEBE is a useful tool to analyze cross-cultural behavior expressed by ECAs. We discussed the prototype of CEBE in general situations in another paper [10]. In this paper, we introduce the basic architecture and the prototype, and we discuss modifications and concepts of CEBE to apply CEBE for investigations to realize an ECA with cultural background.

CEBE will be used as a Wizard of OZ (WOZ) system in two types of research. The first is data capturing to realize agents which have a communication skill

with existing cultural background for cross-cultural communication. The other is the investigations of human-agent interaction on the supposition that humans and agents which controlled a human belong to different culture.

The rest of the paper is organized as follows. Section 2 reviews related works and discusses achievements and non achievements of previous work. Section 3 explains the design and architecture of CEBE. Section 4 describes implementations to detect human behavior with cultural background in CEBE. Section 5 discusses concepts and limitations of CEBE to capture and analyze human behavior with cultural background. Finally, Section 6 contains conclusions and future works.

2 Previous Works

A number of research groups have studied the use of ECAs for intercultural communication.

Huang et al. [6] developed a culture-adaptive virtual tour guide agent that is implemented in a modular way with the GECA Framework [5] to minimize the development cost. It can switch its behavior and speech language to three culture modes: general Western, Japanese, or Croatian. This study focuses only on the surface traits of culture, that is, languages, symbolic gestures, and probably culture-dependent characteristics of gestures. Since they used scripts to describe human-agent interactions, the range of possible interactions will be relatively limited and the quality of the whole system heavily depends on the knowledge and skill of the agent designers.

Iacobelli and Cassell [7] showed an attempt to use virtual peers to encourage African American children to switch their language coding to increase school-based literacy. They implemented two virtual peers by applying two models of behavior to an existing virtual child with a racially-ambiguous appearance in their previous study [2]. They implemented their enculturated behavior on the basis of observations of the verbal and non-verbal behavior of children. Their agents, therefore, had the same limitations as an agent in Huang et al [6].

Some researchers developed their systems to generate culturally adequate behavior based on a human's observable behavior instead of implementations of cultural backgrounds by heuristics. For example, Rehm et al. [14] proposed a Bayesian network model of cultural adaptation, which was then employed in two different sample applications that illustrate the great potential of culture adaptive systems. They, however, realized the model with empirical data from the corpus study under standardized conditions at hand. In addition, the empirical data was cross-cultural rules and habits which could be known in advance. This is a common issue when a behavior model is developed from empirical data at hand.

There is another method to learning cross-cultural behavior, which is through by behavior data obtained from encultured human-agent interaction in which an agent directly expresses behavior of a human operator. This method is our approach. There are some systems in which an agent directly reflects human actions.

Ghobadi et al. [3] proposed a system to control a hand based robot using 2D/3D images. These images are used as the input for hand detection, classification and a tracking system which is used as an interface for sending the commands to an industrial robot. Hirsch et al. [4] also developed BiDi (bidirectional) screen to support 2D multi-touch and walk-up 3D gesture interaction. An image sensor placed a small distance behind an LCD forms a mask-based light-field camera, allowing passive depth estimation. The estimated scene depth can be used to support real-time 3D gesture interaction. These systems are specialized for hand operations, and they could not capture cross-cultural behavior in human-agent interaction.

Tachi et al. [15] had developed a master-slave manipulation system with the function of mutual telexistence (TELEsarPHONE). The system composed of 3 subsystems; a slave robot TELESAR II, a master cockpit, and a viewer system. The operator sitting in the cockpit manipulated the slave robot through two master arms, two master hands multi-stereo display system, speakers and a microphone. The operator had to equip these master arms and master hands. By using this system, the robot reflects the arm and hand actions of the operator. The devices, however, would prevent natural interaction with a cultural background.

We have discussed achievements and non achievements of previous work for the purpose of our study. Some researchers presented ECAs for intercultural communication. Most ECAs were, however, implemented their cross-cultural behaviors and behavior models on the basis of observations of human-human interaction. Possible interactions of these ECAs are relatively limited and the quality depends on the agent designers. To solve this issue, there is a method to learning cross-cultural behavior by behavior data obtained from HAI, in which an agent reflects human behavior. Some systems where an agent reflects human actions could not achieve natural interactions with a cultural background.

The objective of this study is the proposition of CEBE design. This study is different from previous works in following ways; the agent reflects human actions without preventing natural interaction by using an immersive environment and non-contact sensors. In addition, a person in CEBE can intuitively understand the differences of appearance and physical capabilities between ECAs and human. These features are necessary to capture human behavior in natural interaction with a cultural background. In addition, we discuss modifications and concepts of CEBE to apply CEBE for investigations to realize an ECA with cultural background, and we partly implemented.

3 Framework

In this section, we propose a design and architecture of CEBE.

Two kinds of ECAs are expected to perform in CEBE; virtual agents and robots. Virtual agents are suitable for interactions in which expressive gestures are needed because most of them can express facial expressions, gaze directions and subtle movements. We had better use virtual agents when we focus on

those features in cross-cultural communication. The virtual agents, however, are unsuitable for interactions in which ECAs have to interact with substantial objects in real world because most of them do not have their body in real world. On the other hand, interactions which are suitable for robots are vice versa. For example, we had better use robots to investigate a multiple-party interaction in which people in different cultures take part because virtual agents cannot join the multiple-party conversation. Both these kinds of ECAs are used depending on the situation. Therefore, we develop CEBE to be able to use both types of ECAs and we will investigate them in both situations.

3.1 Necessary Conditions for CEBE

CEBE is an environment in which a human can interact with another human who has the appearance and physical limitations of ECAs. CEBE has two aspects; one is a system to obtain data of cultural behavior in a human interaction through a body and perception of an ECA. Another is a system to operate a body of an ECA by using perception of the ECA and obtained data.

In either case, a person in CEBE has to be able to perceive the same level of information which an ECA can perceive. We need to consider the difference of information processing between ECAs and humans. For example, a human can recognize his/her surroundings by using only video in most cases, because visual information is easily interpretable by humans. ECAs, however, need lots of sensor data, such as a range sensor, 360-degree camera, microphone, and so on, to recognize their surroundings at the same level. It is important to provide both a person and an ECA with the same level of information. It is, however, difficult to process all of the information which is used in human-human communication. Therefore, we need to control the information provided to a person depending on the situation.

We also have to be able to obtain and express cultural interaction behavior in a human-agent interaction. For example, we behave in various ways, such as shaking our heads, making hand gestures and changing prosody, when we negotiate with others. We also pay attention to nonverbal behavior to speculate others' intentions. We expect that significant information which expresses cultural interaction behavior is gaze directions (head orientation), movements of the upper body and hands, and prosody of utterances (pitch and power of voices). The accuracy of measuring and reproducing behavior is dependent on the situation. We expect that ECAs cannot completely reproduce human behavior in most cases. We also expect that people can convert their cultural behavior which ECAs can reproduce by mutual adaptations, in the case that they understand the appearance and physical limitations of ECAs.

To sum up, necessary conditions for CEBE are as follows:

- Both a person and an ECA are provided with the same level of information which is needed to interact with cultural behavior.
- Cultural interaction behavior in a human-agent interaction can be obtained and expressed.

– A person in CEBE can intuitively understand appearance and physical limitations of ECAs, and he/she can convert their cultural behavior which ECAs can reproduce.

The accuracy of measuring and reproducing behavior is depending on the situation.

3.2 Prototype

On the basis of necessary conditions for CEBE, we developed a prototype CEBE which detects head direction, posture of the upper body, and 3D angles of arms, and controls a head and arms of a robot or a virtual agent based on the detected data. We currently implemented the methods to capture human behavior to apply CEBE for using robots. We thus did not implement the methods to capture gaze directions, facial expressions and hand gestures.

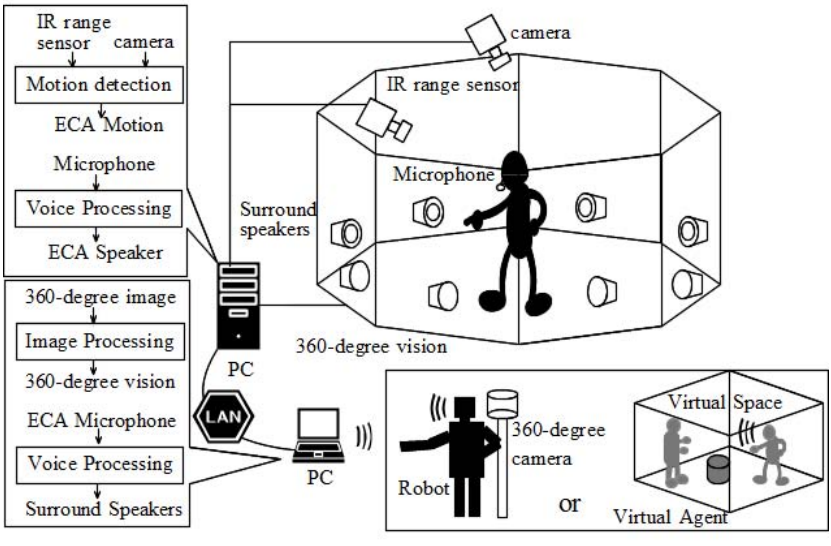


Fig. 1. An architecture of prototype CEBE

Devices which we used in the prototype. Figure 1 shows the whole image of the prototype system.

The prototype system had an immersive display, which reproduces 360-degree image around an ECA. The display intuitively provides an operator with surroundings of the ECA. We expected, therefore, that the operator's cognitive load to know the surroundings was reduced by using this display.

We used non-contact or small sensors to detect the operator's behavior in the immersive display because the sensors would not prevent natural interaction. Optical motion capture system is a typical non-contact sensor which can measure

human body motion. The system needs many cameras for robust measuring. We, however, could not set enough cameras for robust measurement in the prototype system because the system used the immersive display in which there was small space. We therefore developed a motion capture system by using infrared range sensors (IR-range sensors) for the system.

ECAs which are used in this architecture are virtual agents which are implemented using the GECA Framework [5] and multi-degree of freedom robots (NAO, Aldebaran Co., Ltd). NAO is a robot which can be controlled by using wireless LAN and control programs. We can control angles of the head, shoulders, elbows, wrists, fingers, hip joints, knees and ankles of NAO.

Fig. 2. Devices which we use in the prototype

Figure 2 shows the devices in the prototype system. The 360-degree camera and NAO is used when we use a robot as an operated ECA. The immersive display composed of eight portrait orientation LCD monitors with a 65 inch screen size in an octagon shape. The diameter of the octagon is about 2.5 meters. The infrared range sensor is used to detect human upper body motions.

System architecture. Figure 1 also shows an architecture of the prototype system which we are implementing. The prototype system can obtain and express head orientation, movements of upper body and hands, and prosody of utterances (pitch and power of voices) for the preliminary experiment, in which a negotiation task will be conducted.

An immersive display reproduces a 360-degree image around an ECA to provide a person in CEBE with information which the ECA perceives. When we use a virtual agent, the virtual space image which is around the virtual agent is displayed. When we use a robot, the image which is captured by the 360-degree camera is displayed. The person in the 360-degree vision system can see all around the ECA. Movements of the upper body and hands of the person are not prevented by the displays. Eight surround speakers play sounds which the ECA can listen. The person can detect the direction of the sound source to some degree.

Head orientation and movements of the upper body and hands of the person are captured by IR-range sensors (Swissranger, mesa imaging), acceleration sensors and image processing. When we use a virtual agent, a body of the virtual agent is moved by the captured data. When we use a robot, the robot is controlled by data which is converted to meet the limitations of robot's degree of freedom. The person can see the movements of the ECA in a window which is displayed in the 360-degree vision system. The voices of the person are captured by a headset microphone, and the ECA plays the voice. The voice passes through a voice processing component. The voice processing component can filter the voice depending on the purpose. When the voice is filtered, the person can hear the filtered voice.

Head orientation and movements of the upper body and hands of the person are captured by IR-range sensors (Swissranger, mesa imaging), acceleration sensors and image processing. When we use a virtual agent, a body of the virtual agent is moved by the captured data. When we use a robot, the robot is controlled by data which is converted to meet the limitations of robot's degree of freedom. The person can see the movements of the ECA in a window which is displayed in the 360-degree vision system. The voices of the person are captured by a headset microphone, and the ECA plays the voice. The voice passes through a voice processing component. The voice processing component can filter the voice depending on the purpose. When the voice is filtered, the person can hear the filtered voice.

4 The Method of Human Motion Capture

We discussed the prototype of CEBE in general situations in another paper [10]. In this section, we describe the modifications to capture human motion with cultural background in the prototype system. In this paper, we only explain the method to measure head rotation and angles of arms because upper body movements are important in the immersive environment. These implementations are currently in progress.

The prototype system could capture basic human behavior, such as a face direction and pointing, when each part of the body, such as head, hands, arms and trunk, were occluded below 30%.

4.1 The Implementation to Measure Head Rotation

In the CEBE for general situations, we implemented the method to measure head rotation by using an reflection intensity image (we call this as "amplitude") of IR-range sensors. This method was not suitable for measuring detailed degree of the head rotation angle. We implemented the method to measure head rotation by using a three-axis acceleration sensor and image processing because we expected that we had to measure detailed head motion for investigations about cultural differences. We could easily measure the vertical head rotation by using the three-axis acceleration sensor. In addition, the sensor would not prevent human

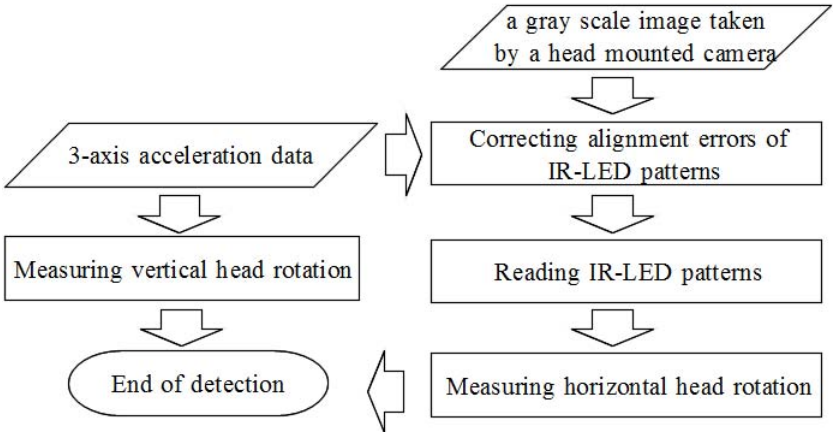


Fig. 3. The overview of the head rotation measuring process

actions because it was adequately small. Meanwhile, we measured the horizontal head rotation by using image processing, for which we used a head mounted camera and infrared-LEDs.

The overview of the measuring process is shown in Figure 3. The inputs of the measuring process are three-axis acceleration data and a gray scale image taken by a head mounted camera. The gray scale image captures the IR-LED patterns which are set between the LCD monitors composed the immersive display. The IR-LED patterns indicate the horizontal angle. In the first step, the angle of vertical head rotation is extracted from the three-axis acceleration data. In the next step, alignment errors of IR-LED patterns in the gray scale image are correcting by using the acceleration data. In the third step, the angle of horizontal head rotation is estimated by reading the IR-LED patterns.

Although the method to measure head rotation is robust, the head mounted camera made some people feel uneasy in test installations. We can measure head rotation by using a geomagnetic sensor, infrared range sensors and only image processing. We are developing another method to measure head rotation by using a small sensor which can measure geomagnetism and 3-axis acceleration because of reducing this uneasiness of users.

4.2 The Implementation to Estimate Angles of Arms

In the CEBE for general situations, we implemented the method to estimate angles of arms by using distance data of IR-range sensors. We could easily detect the region of a human body in an image by using the IR-range sensor because the sensor could obtain many 3D positions of measurement objects' surface. In addition, the IR-range sensor was robust to occlusion than an optical motion capture system in the prototype system. We, therefore, expected that we could use this method for investigations about cultural differences. In this section, we briefly describe the method.

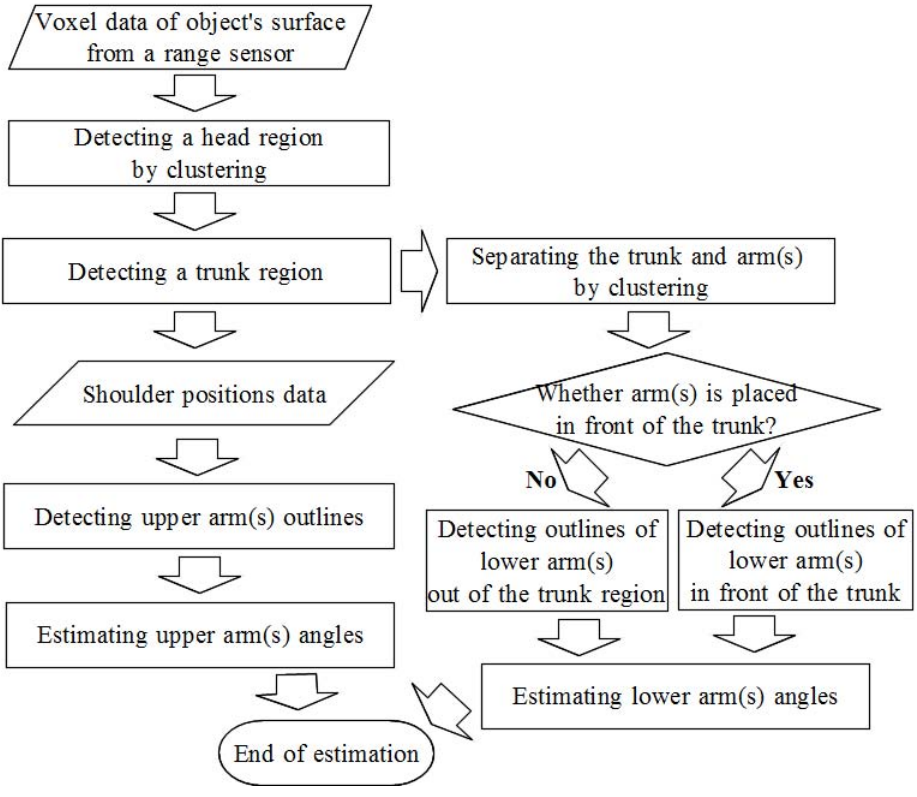


Fig. 4. The overview of the estimation of arm angles

The overview of the estimation process is shown in Figure 4. The input of the estimation process is voxel data of object's surface. In the first step, the horizontal widths of scattered parts of the measured object are used as the input features for a clustering technique to segment into head region and body region. In the next step, a trunk region is detected on the basis of structure of the body. From the third step, the estimation process diverges into two; estimation of upper arm angles and lower arm angles.

In the upper arms estimation, shoulder positions are firstly estimated from the trunk region on the basis of structure of the body. Secondly, the outline of the upper arms is detected by searching voxel data from the shoulder positions to the outside of the trunk. Finally, detected voxels of the upper arm outlines are used as the input features for a straight-line approximation of upper arm angles.

In the lower arms estimation, the distance of the voxel data is firstly used as the input features for a supervised clustering technique to segment into the trunk and the arms in the trunk region. When either arm is placed in front of the trunk, the outline of lower arms is detected by searching voxel data of the

lower arm clusters. When both of the arms are placed out of the trunk region, the vertical widths of scattered parts of the measured object are used as the input features for a clustering technique to segment into the arms region out of the trunk. The outline of the lower arms is then detected by searching voxel data of the arm clusters out of the trunk region. Finally, detected voxels of the lower arm outlines are used as the input features for a straight-line approximation of lower arm angles.

Fig. 5. An example of arm angles estimation

Figure 5 shows an example of an arm angles estimation; (a) shows the IR-range sensor data, (b) shows the result of arm region clustering when either arm is placed in front of the trunk and (c) shows the result of the arm angles estimation. In (b), the dark gray region indicates a right arm region. This means that we could correctly segment the trunk region and the arm region. In (c), two white lines indicate upper and lower arm angles. This means that we could correctly estimate the arm angles.

5 Discussions

5.1 The Concepts to Capture Gaze Directions, Facial Expressions and Hand Gestures

In this section, we discuss the concepts to capture gaze directions, facial expressions and hand gestures. We expect that these are needed for investigations about cultural differences.

Gaze direction. The first author developed a real-time system for measuring gaze directions without head restraint [11]. The system could measure gaze directions in a small space (30cm high, 30cm wide and 20cm depth) because stereo cameras could measure in the small space.

By using IR-range sensors, we can obtain many 3D positions and the rough texture of face in real time (Figure 6). We can thus improve the system with same accuracy (approximately $\pm$ 2 degrees) in larger space by using both IR-range sensors and cameras. The system can be used for CEBE.

Fig. 6. The examples of captured data of faces and hands by an IR-range sensor

Facial expressions. There are many software which convert facial features of a person into facial features of an avatar. We can easily invoke the external program as a plug-in from the virtual agent which we use because the virtual agent are worked on GECA Framework. When we use the software which can export facial features in real time, the virtual agent can express the facial expressions of the user in CEBE.

Hand gestures. We did not implement the method to capture hand gestures because we preferentially implemented the methods to capture human behavior to apply CEBE for using robots. It is difficult to capture hand gestures by using a motion capture system needed makers. We expect that we can capture hand gestures because we can obtain 3D positions of the surface by using IR-range sensors.

Ghobadi et al. [3] proposed a system to control hand based robot using IR-range sensors. Figure 6 also shows that IR-range sensors can capture the features of a hand. In addition, we can easily trace the hand positions because we estimate 3D-angles of arms. We thus expect that we can capture more detailed hand gestures by using both IR-range sensors and pan-tilt cameras.

5.2 Limitations

We will have to develop some implementations so that an agent reflects human behavior and can interact with people with cultural background.

It is most important for natural interaction to compensate for the delay time of agent actions. This delay was pronounced when a robot was controlled. To resolve this issue, we will have to implement a predictive control mechanism. On the other hand, we need to investigate the influence of the delay time of agent actions in encultured agent-human interaction.

It is also important to develop a better method to give feedback to a system user. In prototype system, we give feedback by using visual information. This method, however, is not effective when an ECA reflects the eye directions of the user. The user cannot perform appropriate actions without feedback. On the other hand, large feedback prevents the user from natural interaction. In addition, human behavior in cross-cultural communication is often expressed unconsciously. We have to consider whether we had better to give feedback or not.

5.3 Target Interactions

The prototype CEBE which we are implementing will be able to detect head direction, posture of the upper body and 3D angles of arms, and control a head and arms of a robot or a virtual agent based on the detected data. By using CEBE, we will be able to gather and analyze the information needed for realizing enculturated agent-human interaction. We can use the captured data for unsupervised simultaneous learning of gestures, actions and their associations, such as Embodied Interactive Control Architecture (EICA) [16].

6 Conclusion

6.1 Summary

In this paper, we proposed an environment in which a person can interact with ECAs controlled by captured behavior of another person with cultural background, and we can analyze the data obtained in actual enculturated agent-human interaction. In particular, we proposed a design and an implementation of the system, CEBE.

In future works, we will focus on a negotiation situation in which people have to express their opinions and speculate on the inner states of a communication partner, and we will conduct a preliminary experiment to confirm the effectiveness of CEBE in an enculturated agent-human interaction. In the experiment, we expect to reveal influences of cultural background in some situations, such as detecting the degree of agreement or disagreement, commitment to the communication, emotional representation, and how strong their assertion in the communication.

References

1. Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds.): Embodied Conversational Agents. The MIT Press, Cambridge (2000)
2. Cassell, J., Tartaro, A., Oza, V., Rankin, Y., Tse, C.: Virtual Peers for Literacy Learning. Educational Technology. Special Issue on Pedagogical Agents XLVII, pp. 39–43 (2007)

3. Ghobadi, S., Loepprich, O., Ah-Madov, F., Bernshausen, J., Hartmann, K., Lof-Feld, O.: Real time hand based robot control using 2D/3D images. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Remagnino, P., Porikli, F., Peters, J., Klosowski, J., Arns, L., Chun, Y.K., Rhyne, T.-M., Monroe, L. (eds.) ISVC 2008, Part II. LNCS, vol. 5359, pp. 307–316. Springer, Heidelberg (2008)
4. Hirsch, M., Lanman, D., Holtzman, H., Raskar, R.: BiDi screen: a thin, depth-sensing LCD for 3D interaction using light fields. In: ACM SIGGRAPH ASIA 2009 Art Gallery & Emerging Technologies: Adaptation, p. 62. ACM, New York (2009)
5. Huang, H.H., Cerekovic, A., Nakano, Y., Pandzic, I.S., Nishida, T.: The design of a generic framework for integrating ECA components. In: Padgham, L., Parkes, D., Muller, J.P. (eds.) The 7th international conference of autonomous agents and multiagent systems (AAMAS 2008), pp. 128–135 (2008)
6. Huang, H.H., Cerekovic, A., Pandzic, I.S., Nakano, Y., Nishida, T.: Enculturating human-computer interaction. AI and Society 24, 225–235 (2009)
7. Iacobelli, F., Cassell, J.: Ethnic identity and engagement in embodied conversational agents. In: International Conference on Intelligent Virtual Agents (IVA 2007), pp. 57–63 (2007)
8. Isbister, K.: Building bridges through the unspoken: embodied agents to facilitate intercultural communication. In: Agent culture: human-agent interaction in a multicultural world, pp. 233–244. Lawrence Erlbaum Associates, London (2004)
9. Nishida, T. (ed.): Conversational Informatics. An Engineering Approach. John Wiley and Sons, Chichester (2007)
10. Ohmoto, Y., Ohashi, H., Takahashi, A., Nishida, T.: Capture and Express Behavior Environment (CEBE) as WOZ system to realize effective human-agent interaction. In: Proceedings of International Workshop on Interacting with ECAs as Virtual Characters in AAMAS (2010) (in press)
11. Ohmoto, Y., Ueda, K., Ohno, T.: Real-time system for measuring gaze direction and facial features: Towards automatic discrimination of lies using diverse nonverbal information. AI and Society: The Journal of Human-Centred Systems 23(2), 187–200 (2009)
12. Prendinger, H., Ishizuka, M. (eds.): Life-Like Characters. Tools, Affective Functions, and Applications, Cognitive Technologies. Springer, Heidelberg (2004)
13. Rehm, M., Nakano, Y., André, E., Nishida, T.: Enculturating human-computer interaction. AI and Society 24, 209–211 (2009)
14. Rehm, M., Nakano, Y., André, E., Nishida, T., Bee, N., Endrass, B., Wissner, M., Lipi, A.A., Huang, H.H.: From observation to simulation: generating culture-specific behavior for interactive systems. AI and Society 24, 267–280 (2009)
15. Tachi, S., Kawakami, N., Nii, H., Watanabe, K., Minamizawa, K.: TELEsarPHONE: Mutual Telexistence Master-Slave Communication System based on Retroreflective Projection Technology. SICE Journal of Control, Measurement, and System integration 1(5), 1–10 (2008)
16. Mohammad, Y., Nishida, T., Okada, S.: Unsupervised Simultaneous Learning of Gestures, Actions and their Associations for Human-Robot Interaction. In: The 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2537–2544 (2009)

Hesitation in Intercultural Communication: Some Observations and Analyses on Interpreting Shoulder Shrugging

Kristiina Jokinen[1] and Jens Allwood[2]

[1] Department of Speech Sciences, University of Helsinki
PO Box 9, 00014 University of Helsinki, Finland
Kristiina.Jokinen@helsinki.fi
[2] SSKKII, University of Göteborg
Box 200, 405 30 Göteborg
jens@ling.gu.se

Abstract. This paper concerns the different ways in which hesitation, and hesitation related phenomena like uncertainty, doubt and other phenomena where lack of knowledge is involved are expressed in different cultures. The paper focuses especially on shoulder shrugging as a signal of hesitation or uncertainty, and starts from the observation that shoulder shrugging has different interpretations depending on the interlocutor's cultural background. It is not commonly used in Eastern cultures while in Western cultures it is a sign of uncertainty and ignorance. The paper reports a small study on the differences in interpretation of a particular video tape gesture, and draws some preliminary conclusions of how this affects intercultural communication between human interlocutors and between humans and conversational agents.

Keywords: hesitation signalling, gesturing, intercultural communication.

1 Introduction

Intercultural communication (ICC) is usually defined as communication between people who do not have the same ethnic or national cultural background, e.g. communication between a Chinese person and a German person (cf. [5], [1]). One of the goals of the study of ICC is to discover similarities and differences in the way people from different cultures communicate through cross-cultural comparative studies. With this cross-cultural information as a background, it is then possible to study whether and how such differences influence intercultural communication between people who have different cultural backgrounds.

There are many types of studies of intercultural communication. The most common type is questionnaire based cross-cultural studies of differences in attitudes and values. Probably the most well known is the IBM-study reported by [5]. Another approach is to observe people's communicative behaviour, based on audio or video recordings of actual interactions in different cultures (leading to cross-cultural comparison) or on recordings of intercultural interaction. This line of research ranges

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 55–70, 2010.

from linguistic-cultural studies to computational modelling of cultural behaviour patterns and has become more common with video corpora being collected and easily available for analysis. Usually, there are three different possible foci for such studies: (i) differences in produced communicative behaviour, (ii) differences in interpretation and understanding of the behaviour produced and (iii) studies of how context influences the communicative behaviour produced or its interpretation. For instance, [9] studied communicative behaviour of Finns and Estonians, and found differences in speaking rate, length of pauses and turns, and interruptions. As she discussed the differences in terms of high-context and low-context cultures ([4]), she concluded that the communication even among people from neighbouring countries with closely related languages like Finnish and Estonian can have differences that reflect cultural differences: Estonian speakers show higher speech tempo, short pauses, and frequent interruptions of interlocutors, which are claimed to be features of low-context cultures, whereas Finnish speakers exhibit opposite features typical to high-context cultures. [8], similarly, studied Finnish and English communication strategies from the point of view of shared context and common ground, and noticed that misinterpretations in various everyday dialogue situations depend on cultural presuppositions of the interlocutors. This can be related to the continuum of high-context vs. low-context cultures. They also pointed out that dialogue strategies are learnt through interaction, they also involve presupposition, and that awareness of cultural differences can help in avoiding miscommunication in ordinary activities such as everyday conversations. In virtual agent technology, cultural differences have been actively studied so as to produce appropriate behaviour in virtual conversational agents, and computational models for culture specific communicative behaviour of such agents have recently been built, see e.g. [6] and [3].

The three main foci of intercultural studies may be further subdivided in several ways. For instance, the produced communicative behaviour may be divided in (i) vocal (verbal and non-verbal), (ii) written (verbal and non-verbal e.g. pictures) and (iii) body movements (verbal and non-verbal), depending on the medium in which the communicative context is mediated. Interpretation may be subdivided in factual and emotional interpretation and context can be subdivided into, for example, physical setting and social setting. In this paper we are interested in body movement (shoulder shrug) and its factual interpretation in a social context, and we assume that the type of social setting is influenced by the social activity in which the interlocutors are involved in.

The goal of this paper is to study the different ways in which phenomena like hesitation, uncertainty, doubt and other phenomena where lack of knowledge is involved are expressed in different cultures. We will refer to this as "hesitation related phenomena". We are also interested in how a given communicative behaviour – the shoulder shrug (which in many Western cultures is used to express lack of knowledge) – is interpreted by people with different cultural backgrounds. This research is part of an on-going collaborative project between three universities in the Nordic countries (Sweden, Finland, Denmark), which also includes a larger intercultural focus on bodily communication in different cultural settings (see more information on the NOMCO-project at the project website: http://sskkii.gu.se/nomco).

In general, hesitation related phenomena can be expressed in a number of different ways:

1. Facial expression
2. Head movement
3. Shoulder movement
4. Prosody (e.g. lengthening or pausing)
5. Special words like *eh* or *hm*

In the annotations of conversational speech, such phenomena as fillers (e.g. filled pauses, discourse markers, editing terms, and parentheticals) and disfluencies are also considered as hesitation markers, as well as self-corrections (repairs) and retractions (reformulation or restart of one's utterance). Many of these phenomena have been grouped under the heading "Own Communication Management" (OCM), see [2], referring to their function when they are often simultaneously used to hold a turn and to gain time for the speaker to choose how to continue speaking.

Some of the ways used to express hesitation seem very wide spread, some perhaps even universal, while others are more specific to certain cultures. If we consider the shoulder shrug, we can, for example, hypothesize that the following three functions are good candidates for being universal; a shoulder shrug is produced, because of an itch, a muscular discomfort, or psychological tension (e.g. a so called tick). The interpretation of a shoulder shrug as lack of knowledge is, however, more uncertain and is what this article represents an attempt to study.

In order to study the connection between shoulder shrugging and hesitation in a more systematic way, especially the interpretation of shrugging as a sign of uncertainty, we conducted a small empirical study using our recorded video data. We showed a video of a shoulder shrug to people of different cultural background, and asked them to interpret the gesture and tell us about their own use of a similar gesturing. The main purpose was to get data on two aspects related to body movements as communicative signals: do people notice certain kinds of body movements, and how do they interpret them in the communicative context.

The article discusses the study and its findings, and is structured as follows. We will present the video data and the setup for the questionnaire and interview study conducted in Japan and Sweden in Section 2. In Section 3, we will first present the results and observations, and then continue with the discussion of their interpretation and significance in general. We will conclude and point to some future research lines in Section 4.

2 Data

2.1 Video Clip

The particular shoulder shrug that prompted the study occurs in a video clip that is part of the dialogue data collected under the auspices of ATR/NICT in Japan [7]. The data contains three free-flowing conversations among four participants and each conversation is about 1,5 hours long. One of the participants is Japanese, while the other three come from three different countries in Europe but are familiar with Japanese culture due to living or having lived in the country. The dialogues are conducted in English, and topics vary from casual chatting and story telling to travel information and cultural conventions.

The particular clip that is the focus of interest in this paper is part of a long discussion concerning how to address people in Japanese and in other languages, and how foreigners and people who do not master the language well usually make mistakes concerning the politeness code of each culture. The person using a shoulder shrug has been explaining the use of the French pronouns *vous* and *tu,* and how foreigners sometimes use the incorrect form. The speaker also uses his hand, which he first keeps close to his mouth but then stretches it towards the partners simultaneously extending his index finger and opening up his palm slightly. He then draws his hand back towards the mouth, and keeps his position during the clarification question by partner B, and finally shrugs at the end. The silenced finish of his sentence is then picked up by partner C who offers an explanation: "you get quite a lot of things excused if you're a foreigner". This prompts partner D to continue her earlier story about how to help foreigners to address Japanese people appropriately, while the partner B replies to this by taking it as a basis for his humorous utterance. This makes all the participants laugh, and thus functions as a release of the slightly embarrassed and tensed situation. The conversation then continues in a different mode, with partner C initiating a new although related topic.

The transcript of the pertinent part of the conversations is given below, and Figures 1 and 2 (next page) present snap-shots of the relevant shoulder shrug from sideways and from the front, by two cameras.

A: and but you know,… you can hear
hand, index finger open up towards the partners
that the person is English by accent or something….
larger movement with the hand palm open as if emphasizing "something"
B: and *vous* is rude?
A: no *tu*
B: *tu* is rude
A: and then … hmm … we just say ok … it's just a...
shrug
C: yeah, yeah … if if you're gaijin you can get excused … quite a lot of things
D: um, um, and or Japanese ... Japanese that person has to tell him … aaa … what he want to be called … how and or please call me aaa Takashi … or by his first name [or] … please call me Suzuki san…
B: [yeah] but I kk I don't like I'm Suzuki … you can call me John
C: ah *laugh*
D: we yeah we … *laugh*
A,B,C,D: *laugh*

Most people within Western cultures interpret the shrug as a sign of hesitation or uncertainty, as if the person does not really know how to express oneself or continue the sentence, but this seems not to be so for everyone in Eastern cultures. A Japanese annotator, for instance, noticed the shrug as a peculiar gesturing which is not common in Japanese culture: in fact, shrugging of one's shoulders is seldom used on any occasion. Instead, the common way of expressing hesitation in Japan instead is to tilt one's head sideways and to prolong the words in the utterance *so desu nee* (roughly translated as "well it is so"). This also seems to be the case with other Eastern cultures as confirmed by our interview and questionnaire study.

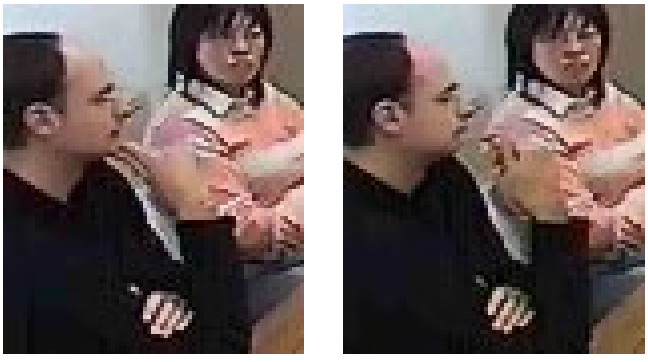


Fig. 1. Snapshot of the interlocutor on the front left shrugging his shoulders after uttering *it's just a....*

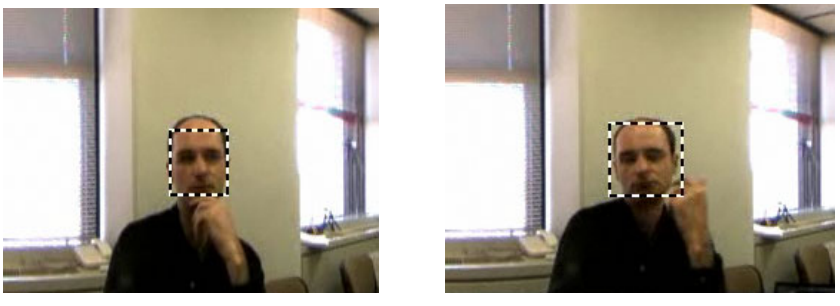


Fig. 2. Front view of the interlocutor shrugging his shoulders

2.2 Interview and Questionnaire

The interview study was conducted in international contexts at two locations, Japan and Sweden. In Japan, the subjects were university students and residents at a residence hall for foreign students and scholars in Kyoto, and the test took place in a quiet corner of the common lounge. The 14 subjects were mostly exchange students who had come to Japan to study their particular discipline, and for many, this was their first experience of living abroad. Half of them were just over 20 years of age (7 subjects out of 14), but two older scholars also took part in the experiment. A summary of the subjects' nationality and background is given in Table 1.

In Sweden the subjects were students in an English language communication program at the University of Gothenburg, and the test was part of their course assignment. The subjects were immigrants, exchange students, and native Swedish students, and their experience abroad varied. There were 27 subjects in their 20's, and they represented a wide variety of nationalities. The nationality distribution is given in Table 2.

Considering the subjects' nationality and experience abroad, we tried to select the participants so that there would be subjects from both Eastern and Western cultures, and the numbers are rather equal considering European vs. Asian participants in

Table 1. Background information of the subjects (N = 14) in Japan. The numbers after the country name indicate the years the person has lived abroad.

Nationality	Experience abroad	Age group	Gender
Catalan	No	20-24	F
Danish	Belgium (20), Japan (1)	20-24	M
Swedish	US (2), France (1), Spain, Uganda, Brazil	25-44	M
Canadian	US, France, Japan	45-65	M
Moroccan	No	25-44	F
Uzbek	No	20-24	F
Filipino	US (4)	20-24	F
Filipino	US (10)	45-65	F
Korean	No	20-24	M
Chinese	No	20-24	M
Chinese	No (English teacher)	25-44	F
Japanese	US (1)	25-44	F
Japanese	No (work with foreigners)	25-44	F
Japanese	No	20-24	M

general. However, as the aim was to explore and verify the previously presented hypothesis of the differences in interpretation between Eastern and Western cultures rather than to conduct quantitative study of the interpretation of the shoulder shrug, there was no rigorous effort to get a statistically balances sample of subjects.

Moreover, the setup of the experiment also supported subjective case studies rather than quantitative experiments. The interviews aimed at collecting data of the subjects' own view of the video clip: how they would interpret the particular gesturing and how they would express hesitation themselves. In the multicultural context of exchange students and researchers, the subjects are of course already used to very international communication, and their own cultural context is in the background. Although the interviews of the video clip were not conducted in void, they were conducted in the context which allowed the subjects' interpretation not only of the hesitation and uncertainty presented in the video clip but also of how hesitation and uncertainty are expressed in their own culture, their views being filtered through their own exposure and readiness to admit various differences in intercultural communication.

The setup of the interview study was fairly straightforward. The purpose of the study was first explained to the subjects together with some background of the cultural differences in the interpretation of gestures. In most cases this prompted enthusiastic presentation of the subjects' own experiences and observations of the cultural differences that they have noticed in their own interactions. The subjects were then shown the video clip and they were encouraged to express their observations and opinions of the clip. If they had not noticed the particular shoulder shrug, their attention was drawn to it by verbally describing what the person does. They could see the video as many times as they wanted, and usually 2-3 was enough.

Table 2. Nationality distribution among the participants (N=27) in Sweden

Nationality	Number
British	1
British/Mexican	1
Cameroonian	1
Chinese	6
German	1
Greek	1
Iranian	3
Iraqi	1
Lithuanian	1
Pakistani	1
Romanian	1
Russian	3
Swedish	5
Vietnamese	1

The subjects were also given a short questionnaire which included questions about their cultural background and interpretation of the video clip. The questionnaire is in the appendix of the paper. Some open-ended questions were aimed at eliciting information about the participants' own interpretation of the event, while the selection of the adjectives was aimed at directing the participants' attention towards the assumed interpretation related to hesitance and uncertainty. The adjectives to describe the gesture and body movement were picked from the Roget's Thesaurus as synonyms and antonyms for the words *hesitant* and *unsure*, and then mixed. The selection also included some less frequent words.

3 Results and Discussion

3.1 General Observations

The analysis of the data collected in Japan confirmed the observation that Chinese and Japanese participants do not use shoulder shrug as a sign of hesitation, and would rather shake their head or move hands sideways. The interpretation of the video clip was described as the speaker being doubtful or puzzled, uncertain, and unsure, and one male participant expressed the speaker's thoughts as "I don't think you are right but I don't want to tell you you're wrong. I don't agree with you". The Japanese subjects also recognized the shoulder shrug as hesitant and unsure, but noted that in Japanese culture, hesitation would normally be expressed verbally or by tilting the head sideways. The Korean male student recognized the gesture as uncertain, unsure, and impolite, but he also said he would use the gesture himself in cases when he was not sure of the situation or did not know the topic. It is interesting that the Swedish participant who was studying in Japan and had lived in various other countries for a

number of years, would also have expected eye-contact in the situation, so that for him the body movement did not appear only as hesitant, doubtful, reserved, and reluctant, but also impolite. For the Danish subject (who grew up in Belgium), the gesturing conveyed feeling of a reserved, shy, and timid person, who was polite and did not want to impose his views on the others. For the two Filipinos, who had long experience in living in the US, the gesture indicated, somewhat surprisingly, communication that is certain, sure, strong, and confident. They would use similar gesturing if confident and of control of the situation, but would also use more relaxed signaling with smiling and not having hands on chin. The female participants from Morocco, Catalonia, and Uzbekistan all considered the gesture fairly common, and interpreted it as uncertain, irresolute, unsure, indecisive, perplexed, and baffled. The Uzbek would also use it herself if not sure or could not decide what to do, while the Catalan and Moroccan pointed out that the gesture would also convey meaning of "doesn't matter", "as you want", "probably". The young Moroccan lady also noted that the unfinished sentence conveys the meaning that the speaker is not very sure of the idea he's trying to convey, but that the gesture also signals that the person does not care, and this is the way in which she would herself use the gesture.

Concerning the data collected in Sweden, an interesting tendency was that the four Middle Eastern subjects (from Iran and Iraq), seem to interpret the gesture as a sign of a sure, certain and confident speaker, like the Filipinos in the data collected in Japan. One of the subjects wrote that the person on the video is certain about the topic and shows certainty with a friendly gesture, while another wrote that the person tries to emphasize what he is saying. These subjects would not use the gesture themselves because it is regarded impolite and also shows lack of respect for older people, i.e. the gesture is related to behaviour that is socially inacceptable. It is interesting how this interpretation differs from that by the Japanese subjects: now the gesture is used but it expresses inacceptable behaviour unlike in Japan where the gesture is not used at all.

The selection of opposite adjectives to describe the gesture may of course be due to a confusion among the subjects concerning the meaning of the adjectives: e.g. one subject had first marked the "confident" words in the questionnaire but changed them to the "diffident" ones after realizing inconsistency in her markings and explanations. However, it may also be related to the cultural tendency of interpreting the shoulder shrug as conveying the speaker's self-confidence and certainty.

In fact, it is possible that that shoulder shrugging has different underlying interpretations, and "lack of knowledge" associated with it can be interpreted in different ways in different cultures. It can be interpreted as "it is not necessary to say any more" or, as the Moroccan lady noted, "I don't care", both of which can also be associated with confidence and certainty. In this way, a shoulder shrug is understood as a sign of non-continuation which can be associated with both lack of ability to continue and lack of willingness to continue, and thus with uncertainty and confidence/certainty (no need to continue), respectively. Perhaps in some cultures the uncertainty interpretation is primarily focused, while in others the confidence interpretation is primarily focused. Cultural conventions and the use of gestures in general would certainly require more investigations here.

3.2 Descriptive Adjectives

Concerning the word questionnaire, global frequencies of the answers of all the subjects are given in Table 3. As can be seen, the most common way to describe the gesture is to associate it with the adjectives *unsure* and *uncertain* (more than half of the subjects), and with *doubtful* and *lack of confidence* (about one third of the subjects). There is a tendency to describe the gesture as expressing the speaker's hesitation and lack of knowledge, but it is worth noticing that the adjective *hesitant* itself does not have especially high frequency (10 out of 42 subjects). On the other hand, the gesture is also related to the speaker being *reserved* or *unconvinced* (9 out of 42 subjects) or *perplexed* (8/42). Also the adjectives *ambivalent*, *indecisive*, *irresolute*, or *reluctant* were picked fairly often (7 out of 42), and the "confidence" adjectives *sure* and *confident* have the same frequency (7/42). The latter can be understood as supporting the above hypothesis that the interpretation of the gesture can focus on both the speaker's willingness to continue the conversation and the speaker's ability to say something: 7 (16%) of the subjects interpreted the gesture as expressing the speaker's confidence rather than uncertainty.

We should bear in mind, however, that the selection of appropriate words from the list of descriptive adjectives may be affected by the subjects' knowledge of English. Most of the subjects were non-native speakers, and since their English ability was not tested, the word-test need not accurately reflect their understanding of the situation.

Table 3. Global frequencies (N=42)

hesitant	**unhesitant**	troubled	**doubtful**	distracted
10	**4**	3	**15**	
confused	disturbed	puzzled	**certain**	**uncertain**
5	3	6	**6**	**23**
resolute	irresolute	**perplexed**	**reserved**	**unconvinced**
1	7	**8**	**9**	**9**
strong	wavering	weak	shy	incoherent
3	5	2	4	6
timid	bold	bashful	confused	unwavering
2	1	2	4	
reluctant	sorry	afraid	baffled	ambivalent
7		1	4	7
sure	**unsure**	decisive	**confident**	indecisive
7	**22**		**7**	7
polite	impolite	apologetic	regretful	**inconfident***
4	3		1	**12**

However, besides few markings on the other adjectives, the general tendency comes clear in the word test: the gesture is described as expressing the speaker's uncertainty and lack of knowledge. Further more, most subjects interpret the gesture as signalling the speaker is indecisive about what to say (hesitant about the topic), but 16% interpret the speaker's disinclination to speak as the speaker being sure and certain in the situation (no need or willingness to continue).

Typically the subjects selected 4-5 adjectives, except for the Greek and Uzbek, who had picked the total of 13 and 10 adjectives, respectively. Based on the selected adjectives, and a few background features of the subjects such as age, nationality, experience abroad, we then formed vectors that characterized each subject concerning their interpretation of the shrug, and then used these vectors to compare the interpretations automatically. Comparison was done with the Support Vector Machine technique in the Weka software [11], as a simple binary classification of the vectors into those that represented Western (European) vs. Eastern (non-European) culture groups. The classification reached 75% accuracy over 10-fold cross-validation, i.e. 3/4 of the subjects were correctly classified into either group on the basis of the features encoded in the vectors.

Considering then how similar the vector representations are, and if they form any natural clusters on the basis of the adjectives, we used the Expectation-Maximization clustering of Weka. The evaluation of the clusters with respect to the two culture groups produced 61% accuracy, i.e. almost 2/3 of the subjects had similar characterizing vectors (had selected similar adjectives) as those who belong to the same culture group.

Both accuracies can be considered fair results, especially as the dataset is small and the vectors are sparse. However, as already mentioned, our study was not intended as a statistical analysis of the interpretation of the shoulder shrug but as a preliminary exploration of intercultural communication. It is obvious that a larger group of subjects with a balanced sampling of various affecting characteristics and parameters would be necessary for a more detailed statistical study, but it is interesting that this kind of preliminary classification and clustering experiments already provide support for the cultural differences and the tendency to describe the shoulder shrug differently in the Eastern and Western cultures is visible in the answers drawn from a small set of 42 individuals only.

Finally, the word selection test also showed that a more detailed analysis of the semantics of hesitation related words is necessary. In the above experiments, using Weka's automatic feature analysis techniques, we found that the adjectives *certain* and *unconvinced* as well as *unhesitant* and *inconfident* were selected as the most predictive adjectives with respect to the two classes (culture groups). The selection was based solely on the statistical properties of the vectors by measuring the information gain with respect to the class, and considering the individual predictive ability of each adjective together with the degree of redundancy between them (i.e. correlation of the adjectives with the classes). The result is in accordance with what has been discussed above qualitatively: interpretation of the shoulder shrug can be described in terms of Western (European) view of the speaker who is *unconvinced* and *inconfident* or with the Eastern (non-European) view of a *certain* and *unhesitant* speaker (which further implies that the gesture is not much used as too arrogant behaviour is socially inacceptable). However, it would be useful to conduct a

linguistic analysis of the words describing hesitation related phenomena and to cluster their occurrence according to their semantic similarity and the context in which they occur. This will be future work.

3.3 Speech Analysis

Hesitation marking in speech is fairly varied. In English this ranges from different types of hesitation markers (*uhh, umm*) to pauses (silence) and slow speaking rate. Fundamental frequency F0 is also shown to rise before pauses occurring in major syntactic boundaries, but not if the pause occurs elsewhere. In Japanese, [10] found that the prosodic and temporal features of a response carry information about how the speaker has grounded the information expressed in the partner's previous utterance, especially if the speaker repeats a portion of that utterance. The features such as longer delays, higher pitch, slower tempo, and rising boundary tone signal lower integration degree, i.e. the information is not fully integrated in the speaker's body of knowledge, and in the context of this paper we can say that they express hesitation and uncertainty.

Figure 3 (a screenshot by the Praat software, http://www.fon.hum.uva.nl/praat/) shows speech analysis of the end part of the speaker's utterance *it's just a…* which occurs immediately before a longer silence and the shoulder shrug. In the screenshot, red points show variation of the first format F0, the blue line shows the pitch contour, and the green continuous line shows intensity.

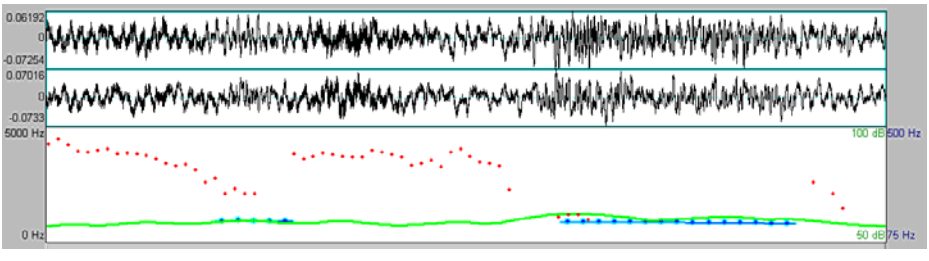


Fig. 3. Combined speech analysis in Praat (see text)

It can be seen that the utterance conforms to all the typical prosodic characteristics related to hesitations, and does not differ from the speaker's other hesitative utterances either. The speaker does not raise his voice (mean energy is about 60 dB), but lengthens the function word so that the speaking rate seems to slow down. The pitch contour stays flat, and shows no upward tendency immediately before the pause (and the shoulder shrug), which is expected since the pause does not occur at a major syntactic boundary.

Considering the correlation of prosody and hesitation related communicative body movements, an interesting question is thus if the two modalities, speech and gesturing support each other or if they operate independently of in the communicative situation. It seems that the shoulder shrugging gesture and prosodic marking have no special correlation in this particular case: the shrug occurs during the pause and it seems more natural to relate the prosody to the unfinished utterance and uncertainty in general,

rather than to the particular gesture. Following this line of thought, it seems safe to assume that the gesture is used to emphasize the speaker's uncertainty, or to add "don't care" –type meaning association – as some subjects described the gesture. In fact, it may be justified to conclude that the shoulder shrug in this particular case has a status of an independent communicative act itself, i.e. after the unfinished utterance *it's just a*, the speaker continues with a shoulder shrug to finish off the sentence with a gesture expression in order to indicate that it is not important to express, or that he is not able to express the actual content in words.

4 Conclusions and Future Work

The current work is based on an interview study of subjects with different cultural background, concerning their interpretation of a particular shoulder shrugging. We confirmed the previous hypothesis of the distinction between Eastern and Western cultures, and the differences in interpreting and using shrugging gestures. We also found out that in Middle-Eastern cultures the same gesturing tends to be interpreted in an opposite way: as a sign of a confident and certain interlocutor. Because this kind of behaviour may also imply arrogance, the gesture can also be considered socially inacceptable and impolite. As a result, we concluded that the underlying interpretation concerning shoulder shrugging may be related to "lack of knowledge" which can then be further analysed in terms of "lack of ability to continue" and "lack of willingness to continue", and that different cultures then may focus more on the uncertainty interpretation while the others focus more on the confidence interpretation.

However, in order to draw statistical conclusions of intercultural communication strategies and to make generalizations about hesitation related expressions, especially shoulder shrugging, in wider cultural contexts, it is important to collect more data on the hesitation expressions in different cultures, and also more examples of shoulder shrugging itself. We can then use a similar kind of methodology as applied in the current work, to collect information about the appropriate interpretations of these expressions. Interviews of people with different cultural background would allow us to collect data on the participants' subjective views and interpretations of the selected video examples, while a word test -type questionnaire would provide us with more objective information about non-verbal signals and their functions in communication. This way we can study how hesitation, uncertainty, doubt, and lack of knowledge are expressed in different cultures, and also how particular expressions, such as shoulder shrugs, are interpreted in different cultural contexts.

We can also include activity analysis into the study and observations of uncertainty in various types of social activities, and consequently, obtain further insights about different manifestations of hesitation related phenomena. The roles of the interlocutors in different activities support different behaviour patterns and thus also differences in the acceptable ways of expressing one's hesitation and ignorance. For instance, communication strategies used when talking to one's superior or a colleague or when chatting with a familiar or an unfamiliar partner, are likely to differ from each other concerning the type and frequency of expressions, the use of gesturing and body movements, and tolerance of silence. Moreover, the five dimensions of [5], i.e. hierarchy, individualism, gender, uncertainty, and orientation, can also be associated

with the roles and activities that the individuals in different cultures are involved in, and consequently, cultural differences can be studied from point of view of social activities that the individuals as members of various social groups take part in. Intercultural communication can thus be analysed in terms of social activities and interactions among individual agents, instead of contrasting stereotypical behaviours.

In ICT (Intercultural Communication Technology) virtual humans and embodied communicative agents form an important application area and the behaviour of such artificial agents is becoming more human-like. Such applications aim at recognizing and monitoring the user's behaviour – both verbal and non-verbal – and providing responses that are appropriate given the task (e.g. factual information exchange) and the interaction context. As the context also includes cultural context, adaptation of the agents to different languages and cultures is a relevant topic that needs to be modelled as well. For instance, in educational applications, training environments, and virtual companions such culture-specific factors affect the users' learning and enjoyment, and taking them into consideration can improve the efficiency and effectiveness of the applications. In these cases, the starting point has usually been an explicit indication of the level of understanding by the user, and it has been common to study conversational feedback and grounding processes in order to provide the users with appropriate responses and effective help in problematic situations. This kind of constructive feedback is useful if we consider the participants' understanding and interpretation, i.e. the intake of the information and its grounding in the existing background information. Hesitation related information is also important in this context since it allows the interlocutors to express the degree of understanding and the degree of commitment to the presented topic. In the smooth communication it is important to convey agreement and emotional stance to the partner, i.e. besides their cognitive understanding of the information, the speakers should also indicate their degree of commitment to the presented information. Studies of various hesitation related phenomena can therefore complement the grounding analysis in the analysis and construction of shared understanding and mutual bonds.

Acknowledgements. The first author thanks NICT and Doshisha University for the opportunity to collect the interview data and conduct the first analyses while a NICT Visiting Fellow at Doshisha University, Kyoto.

References

1. Allwood, J.: Tvärkulturell kommunikation in Allwood, J (Ed.) Tvärkulturell kommunikation, Papers in Anthropological Linguistics 12, University of Göteborg, Dept of Linguistics. Intercultural Communication, Also in English (1985)
2. Allwood, J.: Bodily Communication - Dimensions of Expression and Content. In: Granström, B., House, D., Karlsson, I. (eds.) Multimodality in Language and Speech Systems, pp. 7–26. Kluwer Academic Publishers, Dordrecht (2002)
3. Endrass, B., Rehm, M., André, E.: Culture-specific Communication Management for Virtual Agents. In: Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Budapest, Hungary, pp. 281–288 (2009)
4. Hall, E.T.: Beyond Culture. Anchro Press, Garden City NY (1976)
5. Hofstede, G.: Culture's consequences: International differences in work-related values. Sage, Newbury Park (1980)

6. Jan, D., Herrera, D., Martinovski, B., Novick, D., Traum, D.: A computational Model of Culture-specific Conversational Behavior. In: Proceedings of Intelligent Virtual Agents Conference, pp. 45–56 (2007)
7. Jokinen, K., Campbell, N.: Non-verbal Information Sources for Constructive Dialogue Management. In: Tutorial at LREC-2008, Marrakech, Morocco (2008)
8. Jokinen, K., Wilcock, G.: Contextual Inferences in Intercultural Communication. In: Arppe, A., et al. (eds.) A Man of Measure. Festschrift in the Honour of Fred Karlsson. Finnish Linguistic Association, pp. 289–298 (2005)
9. Pajupuu, H.: Cultural Context, Dialogue, Time. Estonian Academy of Sciences, Institute of the Estonian Language, Tallinn (1995)
10. Shimojima, A., Katagiri, Y., Koiso, H., Swerts, M.: An Experimental Study on the Informational and Grounding Functions of Prosodic Features of Japanese Echoic Responses. In: The Proceedings of ESCA Workshop on Dialogue and Prosody, pp. 187–192 (1999)
11. Witten, I.H., Frank, E.: Data Mining:Practical machine learning tools and techniques, 2nd edn. Morgan Kaufmann, San Francisco (2005)

Appendix

==========
QUESTIONNAIRE FOR THE VIDEO INTERPRETATION
==========

Please fill in the form and return it to the instructor.

AGE: _____ GENDER: _____

MOTHER TONGUE: _____

WHICH OTHER LANGUAGES DO YOU SPEAK?
Please indicate also the level of your knowledge:
(1) like native (2) good command
(3) sufficient (for daily life) (4) basic knowledge only

_____

_____

_____

NATIONALITY: _____

WHICH COUNTRIES HAVE YOU LIVED IN AND FOR HOW LONG?

_____

_____

_____

HOW WOULD YOU INTERPRET THE PERSON'S BODY MOVEMENT IN THE VIDEO?
Choose as many alternatives as you like:

hesitant ___	unhesitant ___	troubled ___	doubtful ___	distracted ___
confused ___	disturbed ___	puzzled ___	certain ___	uncertain ___
resolute ___	irresolute ___	perplexed ___	reserved ___	unconvinced ___
strong ___	wavering ___	weak ___	shy ___	incoherent ___
timid ___	bold ___	bashful ___	confused ___	unwavering ___
reluctant ___	sorry ___	afraid ___	baffled ___	ambivalent ___
sure ___	unsure ___	decisive ___	confident ___	indecisive ___
polite ___	impolite ___	apologetic ___	regretful ___	inconfident ___

PLEASE DESCRIBE THE MEANING OF THE BODY MOVEMENT IN YOUR OWN WORDS:

_____

_____

_____

WOULD YOU USE SIMILAR GESTURING IN A SIMILAR SITUATION YOURSELF?

YES ___ NO ___

IF YES, WHY?
IS IT COMMON IN YOUR CULTURE?

_____

_____

_____

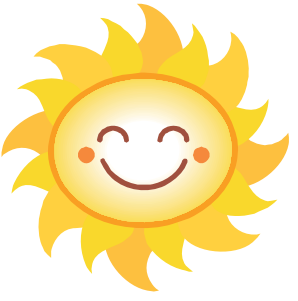
IF NOT, WHY NOT?
WHAT KIND OF GESTURING WOULD YOU USE IN A SIMILAR SITUATION IN YOUR CULTURE?

_____

_____

_____

==========
THANK YOU VERY MUCH FOR YOUR TIME!
WE APPRECIATE YOUR HELP!
==========



A Socio-Cultural Model Based on Empirical Data of Cultural and Social Relationships

Afia Akhter Lipi[1], Yukiko Nakano[1], and Matthias Rehm[2]

[1] Dept. of Computer and Information Science, Seikei University, Musashino, Tokyo 180-8633, Japan
afiaakhter@hotmail.com
[1] Dept. of Computer and Information Science, Seikei University, Musashino, Tokyo 180-8633, Japan
y.nakano@st.seikei.ac.jp
[2] Dept. of Media Technology, Aalborg University, Niels Jernes Vej 14, DK-9220 Aalborg, Denmark
matthias@imi.aau.dk

Abstract. The goal of this paper is to integrate culture and social relationship as a computational term in an embodied conversational agent system by employing empirical and theoretical approach. We propose a parameter-based model that predicts nonverbal expressions appropriate for specific cultures in different social relationship. So, first, we introduce the theories of social and cultural characteristics. Then, we did corpus analysis of human interaction of two cultures in two different social situations and extracted empirical data and finally, by integrating socio-cultural characteristics with empirical data, we establish a parameterized network model that generates culture specific non-verbal expressions in different social relationships.

Keywords: Culture, Social relationship, Posture, Bayesian network.

1 Introduction

When people interact, their interaction patterns and behavior, such as the distance they stand from each other, posture, facial expression, how much they gaze at one another, and many other attributes depend on the interpersonal relationship existing between them and the geographical boundary which is dictated by the culture to which they belong [6]. Behavior regarded as friendly in one culture may, thereby, be interpreted as inappropriate or impolite in another. Thus, in an era of globalization, there is an increasing need for a raised awareness of the culturally specific behaviors associated with various social relationships.

In order to identify the culture-specific behaviors involved in such interactions, this study set out to develop a socio-cultural model in relation to non-verbal behaviors for the purpose of creating appropriate nonverbal expressions specific to postures through utilizing Embodied Conversational Agents so that users would be able to experience culture-specific postures in situations involving different social relationships. In order

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 71–84, 2010.

to formulate such a socio-cultural model, two issues initially needed to be addressed: 1) describe culture and social relationships in a precise way and 2) identify reliable empirical data.

With regard to the first issue, theories that have examined culture and social relationships need to be evaluated. Hofstede's dimensions [7] pertaining to cultural characteristics were accordingly employed for the purpose of explaining culture that are defined using five criteria, each of which are quantitative, and consist of: hierarchy, gender, identity, uncertainty, and orientation. Importantly, Hofstede's dimensions rely on variables that are more directly linked to social and organizational processes that focus specifically on human values, rather than on general beliefs and practices. For the purpose of explaining social relationships, we carried out a general assessment of previous studies and from this, and also based on copious studies relating to social psychological research, four parameters were identified that were able to categorize social relationships: power, sense of like mindedness, frequency of contact, and length of relation.

With respect to the second issue, the identification of appropriate empirical data was initially required. Unfortunately, the material available in the literature was not specific enough and there was also little quantitative data that could be used for empirical purposes. As a consequence, data from human subjects from Germany and Japan were compared and analyzed in relation to two specific kinds of social relationship where a first meeting and interactions takes place with a higher status individual.

Through integrating theories of culture and social relationships with empirical data, we were thus able to update the socio-cultural model [1] that was employed in the present paper, which generates culture specific non-verbal expressions. Our ultimate aim is not only to construct a model that can specify embodied conversational agents, but also to propose a general model, which can assess nonverbal criteria for a variety of cultures in different social relationships.

This paper is organized in the following way. In Section 2, previous studies will be reviewed whereas socio-cultural characteristics will be briefly outlined in Section 3. Section 4 will describe the empirical data as extracted from the review of available research and Section 5 will present an updated version of the socio-cultural model which will be linked to the combined socio-cultural criteria and empirical data. Section 6 will present concluding remarks and recommend directions for further research in the field.

2 Relevant Studies

Many studies have shown that culture affects nonverbal behavior, such as gesture, posture, and proxemics [2, 3, 8, 9, 14]. In a study of avatar mediated communications, Nakanishi et al. [18] developed a virtual meeting space where human-human communications were mediated by avatars that allowed the cultural gap between communicators to be compensated for. In their study, a cross-cultural experiment was carried out where subjects from two cultural backgrounds were recruited and asked to interact through engaging in safe topics relating to, for example, the movies, music, sports, weather, and also unsafe topics concerning money, politics and religion. The

agent normally remained at some distance but intervened in the conversation regarding the topic under discussion by the subjects only when the latter had temporarily ceased to communicate. The researchers of the study were interested in the influence the agent exerted on the different interaction styles of subjects, who were from two different cultural groups, i.e., American and Japanese. It was found that, during the safe topic the agent exerted a more positive effect on the American compared to the Japanese subjects, whereas during the unsafe topic the presence of an agent influenced both groups equally and the subjects also felt the conversation was more interesting.

In addition to culture, social relationships are also an important for specifying nonverbal behaviors [4, 8, 6, 11]. [4] Demonstrates how non-verbal behaviors are associated with high and low social power, where different types of non-verbal attributes such as posture, gesture, head movement, gaze, proxemics, and other criteria were examined, and where significant results were also found for each category. In the case of postural behavior, it was established that individuals with greater authority were believed to have a more erect posture, lean forward more, display an open body position, and orient themselves towards others. An individual with great authority was also expected to shake, tilt upwards, and orient the head towards others more often.

Such findings in human communication studies have been exploited for the purpose of developing virtual agents for investigating behavior in foreign cultures. In this respect, the ORIENT educational application [19] consists of an agent oriented role-play system designed to create cultural empathy for 13-14 years old students. A culture-specific character was thus developed where 13-14 year olds were asked to carry out role play with an agent, who was designed as someone from another planet, in an imaginary artificial culture based on the cultural theory of Hofstede. The procedure involved adolescents initially being introduced to a foreign culture where they attempted to become accepted by that culture by means of an agent from the culture itself. This role based interaction thereby allowed adolescents to become empathic towards another person from a foreign country and, in this way, a link could be forged between behavior and culture through the creation of a believable agent.

In order to formulate such culturally adaptive behaviors using agent systems, computational models and mechanisms for the realization of nonverbal behaviors in agents were required. In order to achieve this, [2] provided a model that simulated culturally specific behaviors using virtual agents, where a set of parameters specifying proxemics, gaze, and overlap in turn-taking for Anglo American, Mexican, Spanish, and Arabic individuals. One of the shortcomings of this approach the researchers found, however, concerned the lack of data that reflected real cultural behavior, especially in relation to quantitative data for gaze and overlap in turn-taking with regard to the literature on Arab culture, which has led to approximations having to be employed based on qualitative descriptions. Bogdanovych et al. [21] has also put forward a model that relies on virtual culture in a 3D virtual world based on the environment, objects, and knowledge associated with a particular culture. In their study, the emphasis was placed on cultural knowledge, the dissemination of culture, and social relationship factors as a way to model culture.

The literature provides a great deal of information as to how communicative behavior is influenced by culture and social relationships yet precise quantitative data is virtually absent from such sources which could potentially be exploited in a technical way to model culture computationally. As a consequence, in the present paper, we first defined socio-cultural traits and then analyzed the human communicative behavior of two cultures involving two different kinds of social relationship. Following this, socio-cultural theories were integrated with both the empirical results obtained from the analysis and the proposed socio-cultural model, which were employed in the present study for generating culture specific to non-verbal expressions in ECA.

3 Cultural and Social Characteristics

From a theoretical approach, we employ Hofstede theory to describe cultural characteristics and take social psychologist description to explain social characteristics. In this section, we described the socio-cultural characteristics in detail.

3.1 Cultural Characteristics

We start with introducing Hofstede theory [7]. Hofstede defines culture as a "the collective programming of the mind that distinguishes the members of one group or category of people from others."The theory consists of the following five dimensions, which are based on a broad empirical survey.

1. Hierarchy *(Small/Large)*: Hierarchy is the extent to which the members of society accept unequal distribution of power. This affects the behavior of both less powerful and more powerful members. The fundamental issue addressed by this dimension is how a society handles inequalities among people. This has consequences for building institutions and organizations.
2. Identity *(Individualism/Collectivism)*: This is the degree to which individuals are integrated into a group. On the individualist side, ties between individuals are loose, and everybody is expected to take care for herself/himself. On the collectivist side, people are integrated into strong and cohesive groups.
3. Gender (*Masculinity/Femininity)*: The gender dimension describes the distribution of roles between the genders. In feminine cultures the roles differ less than in masculine cultures, where competition is rather accepted and status symbols are of importance.
4. Uncertainty *(Weak/strong):* The tolerance for uncertainty and ambiguity is defined in this dimension. It indicates to what extent the members of a culture feel either uncomfortable or comfortable in unstructured situations which are novel, unknown, surprising, or different from usual.
5. Orientation *(short/Long)*: This dimension distinguishes long and short-term orientation. Short-term orientation stands for a society fostering virtues oriented towards persistence and perseverance, thrift, ordering relationships by status and observing this order by having a sense of shame. Long-term orientation stands for a society fostering virtues of personal steadiness and stability, protecting face, respect for tradition and reciprocation of greetings, favors and gifts.

Since cultural characteristics in Hofstede theory are synthetic, a set of parameter values indicates the cultural profile. Fig. 1 gives Hofstede's ratings [2] for Germany and Japan. For example, in Gender dimension, Japan (95) is more masculine society than Germany (66).

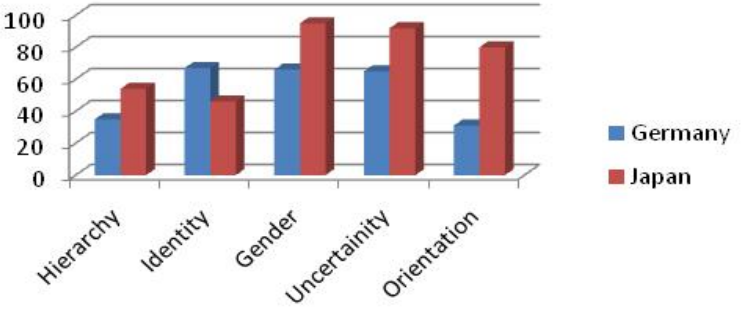


Fig. 1. Hofstede ratings for Germany and Japan

3.2 Characterizing Social Relationship

Wierzbica [22] claimed that the concepts and ways of describing social relationships have not been exact enough thus leading to the term being employed differently by various investigators. Wierzbica also emphasized how the terminology used by researches in the area of cross-cultural pragmatics also tended to be too imprecise. For example, the term "intimacy" is used as "directness," "spontaneity," "solidarity," "self expression," "frequency of contact," and so on, by researchers from different cultures without explaining what exactly is meant by the concepts. In comparing the way in which researchers have employed such terms, it is obvious that the same concept is not being used uniformly. Moreover, Spencer-Oatey [11] claimed that the way social relationships are referred to remains imprecise and, secondly, people from different cultures regard social relationships in various ways. For example, in western culture the vertical relationship associated with "power" is seen in a negative way, whereas in both China and Japan this is regarded as positive. Thus, a person who is "superior" does not simply dominate or control a person of lower status as the individuals involved are regarded as being bound together in a relationship, which involves extensive mutual ties and responsibilities.

In this study, we employed the following four criteria which have been previously defined by [11] for the purpose of measuring social relationships.

1. Power *(High/Low)*: By power we mean the power of the speaker over the hearer in a given role involving a relationship. One person is thereby regarded as holding power over another in the sense that such an individual is able to control the behavior of another. Power can also be based on several factors, such as physical strength, wealth, age, institutional role [11].

2. Sense of like mindedness *(High/Low)***:** This concept refers to the situation where similar views are held that leads to a strong relationship between two individuals [11].
3. Frequency of contact *(High/Low)***:** This refers to a social dimension that is horizontal in structure that determines the format of a relationship. Frequent contact [11] refers to distance/intimacy.
4. Length of relation *(Long/Short)***:** This term also refers to how roles in a relationship are measured. The duration of the relation hence influences the amount of closeness in a relation [11]. In cross-cultural research it has been established that individuals from different cultures regard this factor differently. For example, friends are regarded as close by Lim and Bowers but as intermediate in terms of closeness/distance by Blum-Kulka et al [11].

4 Comparative Corpus Analyses

Evidence derived from the literature indicates that both verbal and non-verbal behavior consists of an assimilation of social and cultural factors. In the present study, we examined posture as one kind of non-verbal behavior. Fig. 2 presents an example of this, where it can clearly be seen that, in the first situation, where the individual is interacting with someone for the first time, the posture of the person is more relaxed than in the second scenario where the interaction involves someone of a higher status. Posture can, therefore, be exploited as a way of determining the relationship between individuals.

Fig. 2. Interaction in two social relationships in a Japanese person

In the present study we carried out an empirical study of posture. Based on previous studies [1], posture was therefore categorized according to five criteria as follows: frequency, duration, mirroring, spatial extent, and rigidness. Brief definitions of the characteristics appertaining to posture are provided below.

1. Frequency: Frequency of change in posture
2. Duration: Duration till which a person remains in the same posture
3. Spatial Extent: Amount of space used in a posture
4. Rigidness: Rigidness or relaxation apparent from the posture
5. Mirroring: Frequency of instances when an individual unconsciously imitates a partner's posture during a conversation.

We subsequently analyzed the influence of these factors in terms of variations in culture and social relationships. We thus initially collected data on posture from an international German-Japanese project referred to as CUBE-G. Details of the data from the CUBE-G project are provided in [12].

4.1 Analysis of Corpus Data in Relation to Cultures

In this section, we analyzed the characteristics of posture studying CUBE-G data based on two cultures: German and Japanese. We annotated the head, leg and arm posture using Bull's coding scheme [15] in relation to first time conversations involving eight German and nine Japanese cases. For the head posture, we took 6 categories, for arm posture we took 16 categories and for leg posture we took 4 categories of posture shapes/types from Bull's coding scheme [15].

As cultural differences were only found for the arm postures, and not for the leg and head postures, these were therefore examined. Table 1 sets out the arm posture changes that were extracted from studying the data of German and Japanese cultures involving first time interactions. The values for "Frequency," "Duration," and "Mirroring," were derived by calculating the average number of posture shifts observed in the main data of the present study. In order to specify the value for "Spatial extent" and "Rigidness" we took the frequently occurring arm postures. Details as to how we formulated the value of each of the posture traits in relation to culture is provided in [1], [17].

Table 1. Posture characteristics affected by cultural variation

Culture	Frequency	Duration	Mirroring	Spatial Extent	Relax
Japanese	4.60	12.26	5.50	1.97	2.38
German	8.08	7.79	0.88	4.04	5.56

Table 1 indicates that the Japanese tend to engage in less frequent postures, remain in the same posture longer, engage in more frequent mirroring, take up less space, and display a more rigid posture in comparison to Germans, which represents findings which have been supported by previous studies. Sanchez-Burks [8], for example, has pointed out that members of cultural groups that are collectivistic in nature tend to display more behavioral mirroring than those from cultural groups that are more individualistic. The findings from the present study therefore support the notion that a collectivistic culture (Japanese) displays a higher incidence of mirroring than in an individualistic culture (German).

4.2 Analysis of Corpus Data in Relation to Social Relationships

Following the above procedure, we then studied the main data from the CUBE-G project in relation to two different social relationships namely, a first time encounter and an interaction with someone of a higher status. Details from the main analysis for two different social relationships are given in [17]. The way in which posture is affected by variation in social relationships is indicated in Table 2.

Table 2. Posture as affected by two kinds of social relationship

Culture	Social Relationship	Frequency	Duration	Mirroring	Spatial extent	Relax
Japanese	First Time(FT)	4.60	12.26	5.50	1.97	2.38
	Higher Status(HS)	1.32	43.52	0.00	0.54	0.73
	Proportion(FT/HS)	3.48	0.28	0.00	3.63	3.29
German	First time	8.08	7.79	0.88	4.04	5.56
	Higher status	2.38	20.43	0.00	0.98	1.44
	Proportion(FT/HS)	3.40	0.38	0.00	4.10	3.85

Table 2 shows that, in both cultures, German and Japanese, frequency is decreased, duration is longer, no mirroring occurs, posture made is smaller and more rigid when interaction is going on with high credential person than with someone for the first time. The evidence from previous studies supports these results. In the case of mirroring, for example, our results indicate that no mirroring occurs with regard to interactions involving high status individuals which provides support for study [8] that, as low power (relative to one's interaction partner) generally prime people to be relationally vigilant, and thus make the less or absence of behavioral mirroring The fourth and seventh rows in Table 2 also indicate that the proportions (FT/HS) for each variable are similar for the two cultures. For example, the proportion relating to Frequency in Japanese culture comes to 3.48 and that for German culture 3.4, which suggests that social relationships affect non-verbal behavior in both cultures in a uniform way.

5 Establishing a Socio Cultural Model Based on Empirical Data

By combining the empirical data extracted from the main analysis involving socio-cultural characteristics, a network model was thereby able to be created. In creating this network model, many questions arose such as how a social, cultural, and behavioral network might first arise; what comes first, social criteria, cultural factors, or non-verbal behavior?; can culture and social factors help in promoting non-verbal norms or, conversely, do non-verbal norms underpin culture or social relationships.

5.1 The Relationship between Social Factors, Cultural Factors, and Non-verbal Behavior

Evidence deriving from previous studies has shown how culture influences the way individuals interpret and evaluate social interactions [14]. In South Korea and Japan, for example, rigid posture indicates that a person holds an influential position whereas in America a relaxed posture gives the impression that a person is credible [14]. In addition, and as explained in section 3, the vertical concept of "power" is regarded as positive by the Japanese and Chinese whereas this is seen as negative by westerners [11]. Thus, cultures tend to vary in attitudes towards social roles and relationships.

The question therefore arises as to the way in which social and cultural factors may be linked to non-verbal behavior. In fact, socio-cultural attributes and non-verbal behavior may continuously influence one another and interact in dynamic and complex ways.

5.2 Reason for Employing a Bayesian Network

We employed a Bayesian network to represent the model used in the current study as this is capable of being exploited in a bi-directional way. Therefore, the model can be used to set or modify the nonverbal behavior of an agent by setting the evidence for a given culture and social relationship as well as to infer the cultural background and social situation from given nonverbal behavior. .A further reason for employing a Bayesian network as the basis for modeling relates to its ability to deal with incomplete and unpredictable information as well as uncertainties at any stage of operation. This capability is crucial for present purposes as the link between social, cultural, and nonverbal behavior involves a many to many mapping procedure.

5.3 The Parameter-Based Socio-Cultural Model

In order to create a Bayesian network that is able to predict non-verbal behavior, the GeNie [13] modeling system was employed. Fig. 3 illustrates the updated version of the Bayesian network as employed in the present model. To maintain consistency with the previous model we had created [1], a social relationship layer was added. In designing this new model, we carefully investigated the empirical results obtained from the present analysis with regard to the two social relationships outlined i.e., a first time encounter and an interaction with someone of a higher status. Based on the

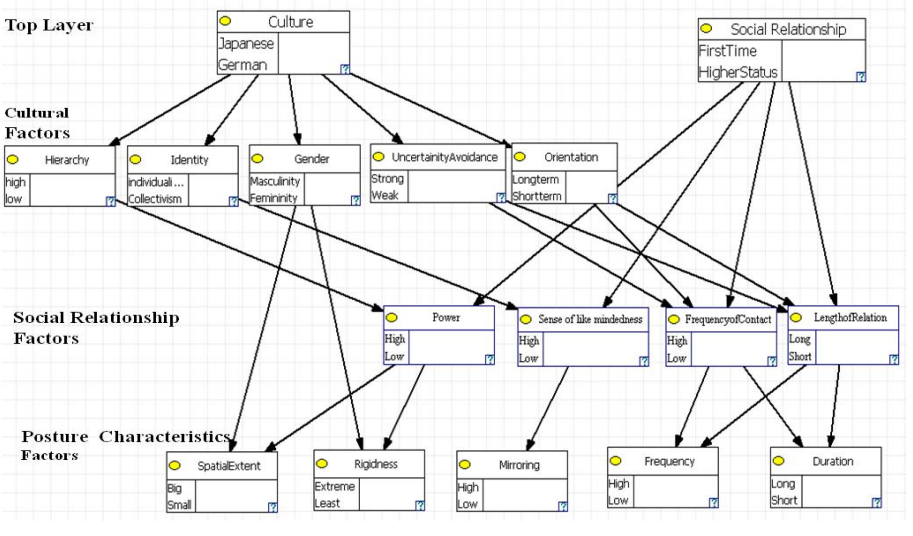


Fig. 3. Bayesian network model for predicting posture characteristics

analysis, we determined which factor or node should be linked to which node in the next layer. Details regarding the design network can be found in [20]. A summary of the network design employed in the present study is shown below.

Top layer: The entry nodes for the Bayesian network consist of a culture and a social relationship node. Culture is connected with Hofstede's dimensions and social relationship is connected with social characteristics. In the existing study, German and Japanese cultures have also been included, as well as the two social relationships with regard to a first meeting and an interaction with a high status individual.

Cultural factors: The layer relating to cultural factors defines Hofstede's five dimensions. As shown in Fig. 1, Hofstede's ratings for each culture have been used for the probabilities in each node for each culture.

Hofsted's cultural factors were linked to the factors for social relationship, as culture influences how various individuals behave in different social relationships. The details regarding the connections for these nodes are provided in [20].

Social relationship factors: The next layer concerns social relationship criteria. The probability of each node being affected as a result of cultural factors and social relationships is set out accordingly. The probability assigned was based on the cultural factors of Hofsted's dimensions as given in Fig. 1, as well as the data obtained from the main study regarding interactions in a first meeting scenario including those with a high status person, which are set out in Table 2.

Posture characteristic factors: The lowest layer consists of the posture characteristics. To assign the probability for each node, we took the empirical data extracted from the main analysis as set out in section 4. For example, for the node Frequency in the Bayesian network, we calculated probability to be 0.66 in the German case and 0.34 for Japan, since the posture shift frequency of the German data (8.075) is 1.75 times that of the Japanese data (4.6).

5.4 Examples of Predicting Postures

When a culture and a social relationship are chosen at the top level as evidence, posture characteristics are estimated. For example, as shown in Fig. 4, when *German is* selected as evidence for Culture and *First time Meeting* as evidence for Social Relationship, the results for spatial extent is large (51%), rigidness small (51%), mirroring high (52%), frequency high (51%), and duration short (51%).

As shown in Fig 5, by retaining *German* as an evidence for culture and selecting *Higher Status* as evidence for Social Relationship, the results for spatial extent are small (51%), rigidness extreme (52%), mirroring small (79%), frequency low (52%), and duration long (51%). Note that the posture of the German subject in Fig 4 is changed to one that is smaller and more rigid in Fig 5 when interaction takes place with someone of a higher status.

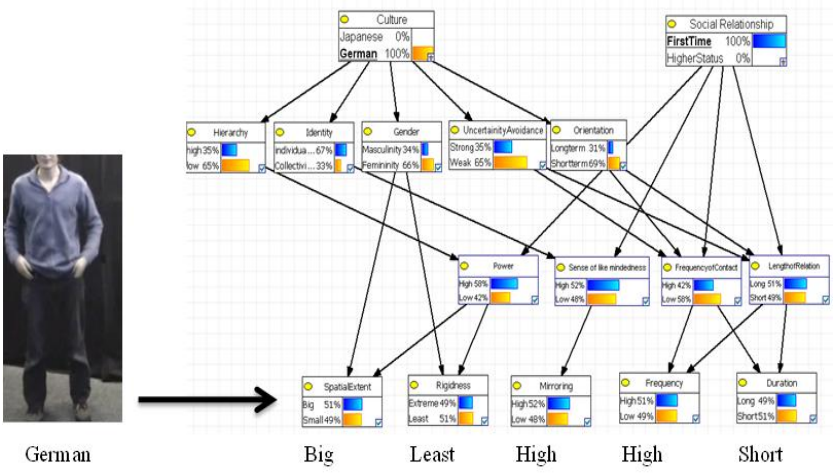


Fig. 4. A German with big and relaxed arm posture interacting with someone for the first time, and the Bayesian Network predicting posture characteristics for German culture in a first time meeting

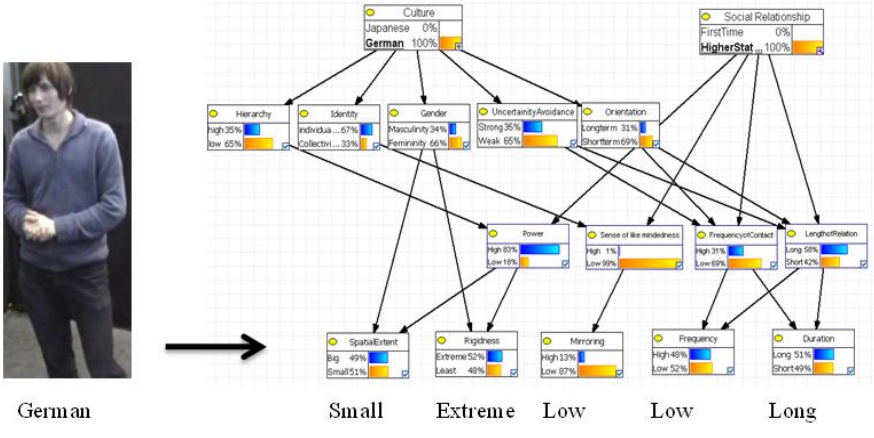


Fig. 5. A German with small and rigid arm posture interacting with someone of higher status, and the Bayesian Network predicting posture characteristics in German culture for higher status

As shown in Fig. 6 taking *Japan* as an evidence for culture and *First time Meeting* as an evidence for Social Relationship, the results for spatial extent are small (66%), rigidness the smallest (67%), mirroring high (57%), frequency high (55%), and duration short (54%).

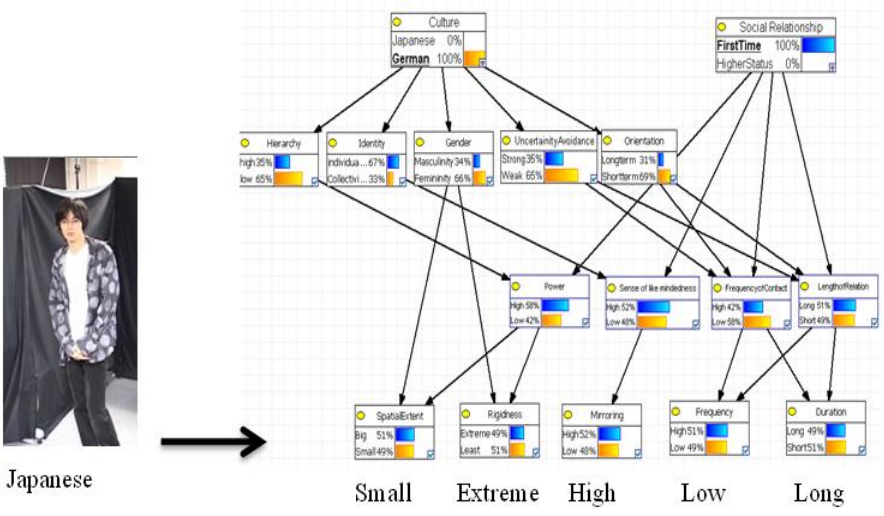


Fig. 6. A Japanese with small and rigid posture interacting with someone for the first time, and the network model predicting the posture characteristic in Japanese culture during a first time meeting

As shown in Fig.7, keeping *Japan* is as an evidence for culture and selecting *Higher Status* as evidence for Social Relationship, the results for spatial extent are smaller (68%), rigidness more extreme (69%), mirroring the least (87%), frequency lower (56%), and duration longer (55%). In a similar way as in the German example shown in Fig. 4 and 5, the posture for the Japanese subject in Fig. 6 changes to a smaller and more rigid one in Fig. 7 when interacting with someone with a higher status.

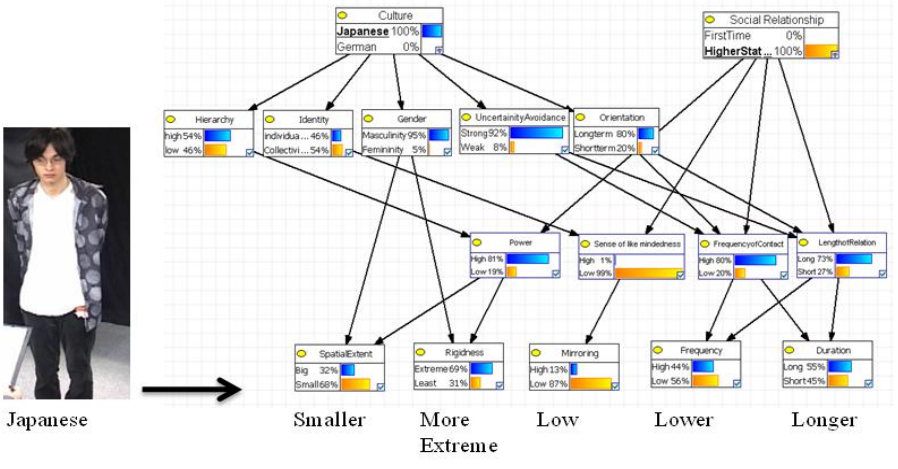


Fig. 7. A Japanese with smaller and more rigid posture interacting with someone of a higher status, and the network model predicting the posture characteristics for Japanese culture with regard to higher status

By providing empirical data for different cultures and social relationships, the model presented here is able to estimate the nonverbal criteria for any culture in different social relationships, and therefore provides a model for identifying the non-verbal factors in any culture or social relationship.

6 Conclusion

In the present study, we initially attempted to identify theories that could be used to explain culture and social relationship in a quantitative manner. We then looked at two cultures and extracted statistical information with regard to two different kinds of social relationship. The statistical information was then assimilated with existing culture and social relationship theories where we employed a parameterized socio-cultural model, which estimated proper posture based on culture and social relationships. This model thereby provides a means of identifying the non-verbal attributes for any culture or social relationship.

In future work, we intend to employ this model in our distance-learning system on the web where two users from different countries can log on to the service and teach their own language to a partner, as well as learn a foreign language from another person. This system will thereby enable users to experience exchanges by employing cultural specific postures through exploiting a human-computer interaction. In creating such a facility, it will also, however, be necessary to evaluate how people perceive the postures generated by our model in terms of naturalness.

Acknowledgment. The work described in this paper was partially supported by the German Research Foundation (DFG) with research grant RE2619/2-1, the Japan Society for the Promotion of Science (JSPS) with a grant-in-aid for scientific research (C) (19500104), and by the European Community (EC) in the eCIRCUS project IST-4-027656-STP.

References

1. Lipi, A.A., Nakano, Y.: A Parameter-based Model for Generating Culturally Adaptive Nonverbal Behaviors in Embodied Conversational Agents. In: HCII 2009, San Diego, pp. 631–640 (2009)
2. Jan, D., et al.: A Computational model of culture-specific conversational behavior. In: Proceedings of Intelligent Virtual Agents, Paris, France, pp. 45–56 (September 2007)
3. O'Neill-Brown, P.: Setting the stage for the culturally adaptive agent. In: Proceedings of the 1997 AAAI Fall Symposium on Socially Intelligent Agents, pp. 93–97 (1997)
4. Carney, D.R., Hall, J.A., LeBeau, L.S.: Beliefs About The Nonverbal Expression Of Social Power. Journal of Nonverbal Behavior 29(2), 105–123 (2005), Summer 2005 a 2005 Springer Science Business Media, Inc. (June 2005) doi:10.1007/s10919-005-2743-z
5. Triandis, C.H., et al.: Three cross-cultural studies of subjective culture. Journal of Personality and Social Psychology, Monograph Supplement 8(4), 1–42 (1983)
6. Krauss, R.M., Chiu, C.Y.: Language and Social Behavior. In: Gilbert, D., Fiske, S., Lindsey, G. (eds.) Handbook of social psychology, 4th edn., vol. 2, pp. 41–88. McGraw-Hill, Boston

7. Gaspay, A., Dardan, S., Legorreta, L.: Software of the Mind – A Review of Applications of Hofstedes Theory to IT Research. Journal of Information Technology Theory and Application (JITTA) 9(3), 1–37 (2008)
8. Sanchez-Burks, J., Bartel, C., Blount, S.: Fluidity and performance in intercultural workplace interactions: The role of behavioral mirroring and social sensitivity. Journal of Applied Psychology 94(1), 216–223 (2009)
9. Ting-Toomey, S.: Communication across Culture. The Guildford Press, New York (1999)
10. Hofstede, G., Bond, M.H.: The Confucius connection: from cultural roots to economic growth. Organisational Dynamics 16(4), 4–21 (1988)
11. Spencer-Oatey, H.: Reconsidering power and distance. Journal of Pragmatics 26, 1–24 (1996)
12. Rehm, M., Nakano, Y., André, E., Nishida, T., Bee, N., Endrass, B., Wissner, M., Lipi, A.A., Hung-Hsuan, H.: From observation to simulation: generating culture-specific behavior for interactive systems. AI & Society, Special Issue on Enculturating Human Computer Interaction 24(3) (2009)
13. GeeNIe and SMILE, http://genie.sis.pitt.edu/
14. Allbeck, B.: Creating Embodied Agents with cultural Context. In: Payr, S., Trappl, R. (eds.) Agent Culture:Human –Agent Interaction in a Multicultural World, pp. 107–125. Lawrence Erlbaum Associates, London (1981)
15. Bull, P.E.: Posture and Gesture. Pergamon Press, Oxford (1987)
16. Rehm, M., Endrass, B., Wissner, M.: Integrating the User in the Social Group Dynamics of Agents. In: Workshop on Social Intelligence Design, SID (2007)
17. Lipi, A.A., Nakano, Y., Rehm, M.: Culture and Social Relationship as factors of affecting communicative non-verbal behaviors. In: HAI Symposium 2009, Tokyo (2009)
18. Nakanishi, H., Ishida, T., Isbister, K., Nass, C.: Designing a Social Agent for Virtual Meeting Space. In: Payr, S., Trappl, R. (eds.) Agent Culture:Human –Agent Interaction in a Multicultural World, pp. 245–266. Lawrence Erlbaum Associates, London
19. Aylett, R., Paiva, A., Vannini, N., Enz, S., André, E., Hall, L.: But that was in another country: agents and intercultural empathy. In: Decker, K.S., Sichman, J.S., Sierra, C., Castelfranchi, C. (eds.) Proceedings of AAMAS (2009), Budapest (May 2009)
20. Lipi, A.A., Nakano, Y., Rehm, M.: Culture and Social Relationship as Factors of Affecting Communicative Non-verbal Behaviors. JSAI journal (under review)
21. Bogdanovych, A., et al.: Virtual Agents and 3D Virtual Worlds for Peserving and Simulaing Culture. In: Proceedings of IVA, Amsterdam,The Netherlands (September 2009)
22. Wierzbica, A.: Cross-cultural pragmatics.The semantics of human interaction. Mouton de Gruyter, Berlin and New York (1991)

Dynamic Term Suggestion for Searching Multilingual School Documents

Kohei Sawa[1], Yusuke Okano[2], Masahiro Hori[1], and Chigusa Kita[1]

[1] Graduate School of Informatics, Kansai University
2-1-1 Ryozenji-cho, Takatsuki-shi, Osaka 569-1095 Japan
{horim,ckita}@res.kutc.kansai-u.ac.jp
[2] Faculty of Informatics, Kansai University
2-1-1 Ryozenji-cho, Takatsuki-shi, Osaka 569-1095 Japan

Abstract. The number of school children having ties overseas is on the rise year after year in Japan. In order to support these children and their parents, we developed a multilingual school document portal site, and made it open to the public. The portal site allows easy-to-use document retrieval by faceted classification as well as keyword search. However, it is not necessarily easy for the users to express their information needs in query terms, and they often come up with poor results from short and generic query terms. In pursuit of the formulation of better query statements with a few search terms even in the initial query, we realized a dynamic term suggestion or auto-suggest interface in the portal site. In the auto-suggest interface, suggested terms are ranked according to domain relevance, rather than merely on the basis of term occurrence frequency in the document collection. In this paper, we explain an overview of the school document portal site, the auto-suggest interface realized in the portal, and the domain-dependent term-weighting scheme, followed by the results of a user study conducted to demonstrate the effectiveness of the auto-suggest function.

Keywords: Auto-suggest, dynamic term suggestion, faceted search, domain vocabulary, multilingual school document, portal site.

1 Introduction

The number of residents having ties overseas is on the rise in Japan, and currently, more than twenty-five thousand children, who do not speak Japanese as their mother language, live in Japan. According to a survey by Ministry of Education, Culture, Sports, Science and Technology in 2008 [1], the top three foreign languages (i.e., Portuguese, Chinese, and Spanish) to be supported in Japanese school constitute 72.9% of the total number of school children whose mother language are not Japanese. In addition to these top three, the top seven languages, which include Filipino, Korean, Vietnamese, and English, constitute 94.1% of the total number [2]. In schools that accept pupils with international background, it is widely known that to overcome the difficulty of communication

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 85–99, 2010.

between teachers and guardians is crucial, and to solve the shortage of appropriate support for specialist subjects for pupils without enough command of Japanese are needed.

Multilingual school documents are parallel text written in multiple languages, and used by school teachers for pupils and their guardians. Such school documents are made for meeting various educational needs: notification of school activities, inquiry about health condition, financial notices, immigration procedures, and so on. A number of school documents are made available as downloadable files through Web sites of institutions such as extra-government organizations, prefectural boards of education, and nonprofit organizations. Those school documents are, however, maintained individually and separately in each Web site. Therefore, it is not easy for the teachers to obtain a desired document from the results of general-purpose search engines, which include all the related pages as well as document links relevant to the search terms.

In order to facilitate sharing and distribution of such school documents, we have developed a multilingual school document portal site [3], which collects over 1,200 links to existing multilingual school documents of twelve languages. The school document portal[1] has been made open to the public since Oct. 2009. The development of the portal site has been conducted as part of the Project MUSE (Multilingual Support for Education) that aims at the deployment of multilingual educational support tools and resources (e.g., educational materials, vocabularies, and documents) through the Web [4].

The school document portal site is realized as a faceted navigation system that allows users to explore a collection of information by using a number of orthogonal attributes or facets [5,6]. Since it is often not easy for the users to fully narrow down a scope of information solely with predefined facet values, this portal site allows the users to perform keyword search as well. However, it is not necessarily straightforward for the users to express their information needs in query terms, and they often come up with poor results from short and generic query terms [7]. It is pointed out even searchers well-acquainted with a target domain relies on objective information such as names of authors, projects, and article titles [8]. It is also reported that the mean number of search terms in a Web search query is marginally above two for both general-purpose search [9] and domain-dependent Web site search [10].

Therefore, in order to help searchers formulate better query statements with a few search terms, it is helpful for the searchers to be suggested query terms as early as possible even during the initial query, rather than after they received the search results. Such dynamic term suggestion is sometimes referred to as auto-suggest [11]. Variations in the auto-suggest feature can be categorized roughly into four types according to their domain-dependence of search range (i.e., general purpose vs. domain dependent) and granularity of term completion (word level vs. character level). By the word-level suggestion, it means that an incomplete word cannot be expanded to be a full spelled word by the system assistance, but the users need to type a word in full before the system

[1] http://www.tagengo-gakko.jp/bunsho/ (in Japanese).

suggests additional words or phrases. Auto-suggest interfaces are already available in practice for general-purpose search engines with character-level completion (e.g., Google Suggest and Yahoo Search Assist), and a log-based longitudinal study provides evidence that auto-suggest interfaces are rapidly prevailing [12].

Despite the increasing presence, there has been little user study on the effectiveness of auto-suggest interfaces [11]. There is a case for the word-level completion with a general-purpose search engine [15]. In the experiment, the participants made positive comments on the interface in the questionnaire, but there was no statistically significant difference between the cases with and without auto-suggest in terms of the users' search time[2] and quality of results [15].

The auto-suggest interface we developed for the document portal falls into the type of the character-level suggestion for domain-dependent search. In this interface, the suggested terms are ranked on the basis of a term-weighting method that takes into account the domain vocabulary as well as entropy of topic in the topic classification scheme created for the school document portal. In this paper, we describe a user study to demonstrate the effectiveness of the domain-dependent auto-suggest interface considering the cases without auto-suggest and domain vocabulary. The remainder of this paper is organized as follows. The school document portal site is briefly introduced in the next section, and the section 3 explains the auto-suggest interface integrated with the search box of the portal. The section 4 presents the evaluation method, and finally the results of evaluation and our findings are given in the section 5.

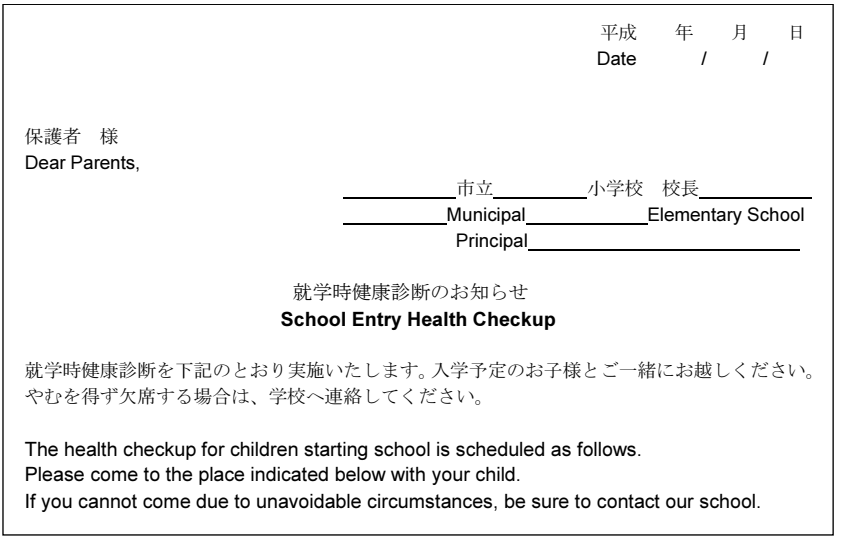
2 Multilingual School Document Portal Site

Multilingual school documents are composed mostly of parallel texts line-by-line in both Japanese and a target language. Figure 1 shows an example of a school document. Most school documents are created separately for each language by translation staff members, and handed out from a class teacher to foreign children to notify school activities and schedules to their parents/guardians. The school document portal site as a faceted navigation system allows users to find items on the basis of multiple aspects or facets, instead of forcing the users to view items from a single classification hierarchy.

As shown in Figure 2, the top page of the portal site[3] displays several facets that characterize multilingual school documents: topic, language, file type, and area of document owning institution. For example, if a school teacher needs to prepare a document of health checkup prior to a school activity, which is handed out to Spanish children and expected to be available in a PDF format. The teacher as a portal user may start by selecting "Spanish" from the language

[2] Since the user interface was not responsive taking 1.8 seconds on an average until the suggested terms were displayed [15], the benefit of the auto-suggest interface was not be fully utilized in the experiment.

[3] The school document portal is currently afforded to Japanese school teachers, and the site in itself is available only in Japanese. The screen copies in this paper are translated in English for the explanation purpose.



平成　年　月　日
Date　/　/

保護者　様
Dear Parents,

＿＿＿＿＿市立＿＿＿＿＿小学校　校長＿＿＿＿＿
＿＿＿＿＿Municipal＿＿＿＿＿Elementary School
Principal＿＿＿＿＿＿＿＿＿＿＿＿＿＿

就学時健康診断のお知らせ
School Entry Health Checkup

就学時健康診断を下記のとおり実施いたします。入学予定のお子様とご一緒にお越しください。
やむを得ず欠席する場合は、学校へ連絡してください。

The health checkup for children starting school is scheduled as follows.
Please come to the place indicated below with your child.
If you cannot come due to unavoidable circumstances, be sure to contact our school.

Fig. 1. Portion of a multilingual school document

facet, and will be presented immediately with Japanese titles of documents with translation in Spanish. Further selection of "Pre-event health checkup" from the topic facet eliminates options in unspecified facets. The file-type facet in this situation only presents file types available under the current conditions, namely, the ones for Spanish parallel documents classified in the selected topic.

In the course of facet-value selection, the user can see a list of documents that meet the specified conditions. If the user further selects "PDF" from the file-type facet, the portal shows titles of documents available in PDF format, whose topic is "Pre-event health checkup" (Figure 3). Note here that the retrieved documents are links collected from different institutions across the country. The area-of-owning-institution facet in the lower left of Figure 3 indicates that the currently listed ten documents come from the three prefectures. In this way, the faceted navigation allows the users to elaborate the facet-value selection progressively, making sure the effect of choice in a facet on the available options in the other facets. Furthermore, dynamically updating the number of available items for each facet value, faceted navigation avoids coming up with empty list or dead ends resulting from unsatisfiable combination of facet values [6].

3 Dynamic Term Suggestion

Since most Web users are accustomed to keyword search, it is helpful to seamlessly integrate the category-based faceted search with keyword search in the same interface [13]. In this school document portal, faceted and keyword search functions are provided in the same interface, and users can make use of them in an arbitrary order to specify requirements for desired documents. Although

Fig. 2. Top page of the multilingual school document portal site

the keyword search is easy to use and helpful, it is not necessarily obvious for the users to select terms effectively navigate the search space. By looking into the search log of the portal site, there were cases where the users typed in generic terms (e.g., *checkup, school*, and *communication*), which did not contribute to narrowing down the search results.

In order to help searchers improve query statements, query expansion techniques have been investigated [14]. Query expansion is to build a new or modified query by supplementing with additional terms, and can be performed manually, automatically or interactively [14]. Interactive query expansion (IQE) is a potentially useful technique to improve queries by incorporating users' decision upon the term selection. However, IQE has not often been used in operational systems due to insufficient integration with search interfaces without substantial help during the initial query formulation [15]. For instance, a real-time query expansion [15] is proposed as an extension of IQE, which allows dynamic term expansion even during the initial query formulation, but it is still less interactive because the expansion is performed only at the word level. The dynamic term suggestion we adopted here is also a sort of real-time query expansion, and more responsive due to the expansion at the character level rather than the word level.

Moreover, it is reported that subject knowledge facilitates increasing the extent and specificity of the queries made by searchers familiar with the subject, and contributes to the improvement of search effectiveness [16]. Therefore,

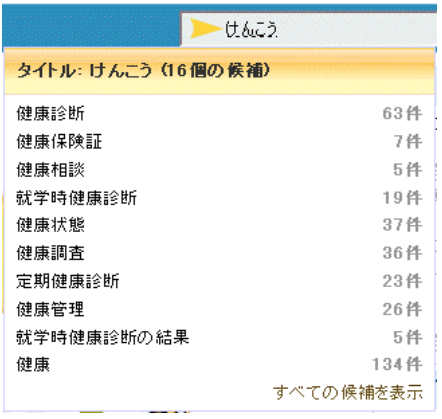
Fig. 3. Results of faceted search with the multilingual school document portal

domain knowledge is an important source of improving query statements. The dynamic term suggestion realized in the school document portal exploits domain knowledge, so that suggested terms can be ranked taking account of domain relevance, rather than merely on the basis of term occurrence frequency in the document collection. In the following subsections, we present an overview of the auto-suggest interface, and then briefly explain the domain-dependent term-weighting scheme.

3.1 Auto-Suggest Interface

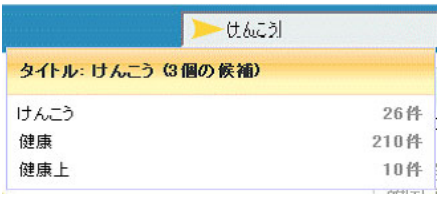
Figure 4 shows a prototype[4] of auto-suggest interface realized in the search box of the school document portal. While the user is typing a term, the interface presents a list of suggested terms after every keystroke. The term expansion is automatically executed once the user starts keystroking for every 500 milliseconds. The suggested terms are shown in the drop-down list, and include compound terms expanded by partial matching with what has been typed so

[4] The interface is currently integrated only for the experimental use, and is not available in the public release.



1. medical examination
2. health insurance card
3. health counseling
4. pre-school medical checkup
5. health condition
6. medical checkup
7. regular health checks
8. health management
9. pre-school medical checkup results
10. health

(a) Terms suggested with domain vocabulary



1. health (in phonetic characters)
2. health
3. health-related

(b) Terms suggested without domain vocabulary

Fig. 4. Drop-down lists with suggested terms on the left in Japanese, and the terms in English on the right for the explanation purpose

far. Each suggested term is displayed with the number of documents relevant to the term. When the user clicks one of the suggested terms, a query is executed and a list of document titles are given in the content area as shown in Figure 3.

The drop-down lists in Figure 4 show the suggested terms came up when the user typed *kenkou*[5] in hiragana or Japanese phonetic characters. If there exist more than ten suggested terms, "show all candidates" link is displayed in the lower right of the drop-down list as in Figure 4(a). All the candidate terms can be listed by vertically extending the drop-down list when the user clicks the show-all-candidate link. The terms in Figure 4(a) are suggested and ranked taking account of the domain vocabulary. They include not only narrower terms such as the first candidate *kenkou-shindan* that means 'medical examination,' but also related terms such as the second candidate *kenkou-hokenshou* that means 'health insurance card.' The dynamic term suggestion here helps the users to follow the Web search practice that is to modify queries from broad to narrow formulation via adding terms for more precision [9].

In contrast, Figure 4(b) shows suggested terms came up simply relying on generic vocabulary without regard to any domain-specific terminology. In this

[5] One of its possible meanings is 'health.'

Table 1. Number of documents classified in each topic

Topic	Number of documents
Achievement/career path	72
Announcement/request	275
Expense/benefits	85
Healthcare	351
School activities	201
Notification/certificate	87
Disaster/crime-prevention	51
Schooling system	77
Miscellaneous translations	73

case, the candidate in the list include only generic terms such as *kenkou* in phonetic characters and kanzi characters, respectively in the first and the second candidate in Figure 4(b). Differences in the variation and ranking of suggested terms come from the domain-dependent term-weighting scheme behind the auto-suggest interface, which is further explained as follows.

3.2 Domain Knowledge

The domain knowledge used for the auto-suggest feature has been developed as part of the project MUSE, and consists of a subject classification scheme of school documents, and vocabularies in school-education domain. The school documents in the portal site are categorized by the scheme and displayed in the topic facet shown in the upper left of Figure 2. The classification scheme for school documents was created by means of card sorting and hierarchical clustering methods, and consisted of the nine of top-level topics [3]. Table 1 shows the top-level topics and the number of documents classified into each topic in the portal site. Note that some of the topic names are composed of two (e.g., *Announcement/request*). This is because they were aggregated as a single topic name as results of hierarchical clustering, so that it can be avoided to create isolated clusters that may include only a few documents in a topic.

Besides the document classification scheme, terms in school-education domain have been collected and refined by school teachers and translation staff members participating in the project. The domain vocabulary used in this study contains 2679 terms, which includes terms related to educational activities (361 terms), school calendar (918 terms), children's healthcare (542 terms), and so on. With reference to the domain vocabulary as well as a generic Japanese dictionary, indexing terms are extracted from Japanese sentences in the parallel translation of all the school documents, by using a morphological analyzer[6]. All the nouns are extracted as indexing terms from the documents, but the terms simply consist of a single (hiragana or kanzi) character are excluded as stop words. On

[6] https://sen.dev.java.net/

the other hand, compound terms in the domain vocabulary may be suggested as narrower or related terms as explained in the term suggestion with domain vocabulary in Figure 4(a). To put it another way, compound nouns in the domain vocabulary facilitate implicit use of semantic relations during the dynamic term suggestions, and provides the users with the semantic context entailed in the school documents in the portal.

3.3 Term Weighting

In the auto-suggest interface, the suggested terms are ranked in the order of significance. The equation (1) below is a domain-dependent metrics that exploits domain vocabulary and topic classification scheme as domain knowledge. Weight of a term i (W_i) is calculated as arithmetic average of X_i and Y_i, which is multiplied by Z_i.

$$W_i = \frac{(X_i + Y_i)\, Z_i}{2} \tag{1}$$

X_i here considers if a term is included in the domain vocabulary, and Y_i measures the infrequency or rarity of a term's occurrence in a subset of documents that are classified into the same topic. This is based on the assumption that the less likely a term is to be associated with topics through documents the better is it likely to be at discriminating relevant from irrelevant a subset of documents in the same topic. Moreover, the arithmetic average is multiplied by Z_i in order to reduce significance of terms that are likely to be weak in semantics and occur frequently. Further details of each term in the equation (1) are given as follows.

$$X_i = \begin{cases} A_N/I_N, & \text{if term } i \text{ is included in domain vocabulary} \\ 0, & \text{otherwise} \end{cases}$$

where A_N is the number of terms in domain vocabulary; and I_N is the number of terms in documents

$$Y_i = -\frac{1}{S_N} \sum_{k=1}^{S_N} \frac{D_{ki}}{D_i} \log \frac{D_{ki}}{D_i}$$

where S_N is the number of topics; D_i is the number of documents containing term i; and D_{ki} is the number of documents containing term i and classified in topic k

$$Z_i = \begin{cases} 0.5, & \text{if the number of hiragana characters in term } i \leq 2 \\ 1, & \text{otherwise} \end{cases}$$

4 Evaluation

The aim of this study is to investigate how effectively users can retrieve school documents with the support of auto-suggest interface as well as the domain-dependent term-weighting method. The study used a 2x2 mixed factorial design with one between-participants variable and one within-participants variable.

The between-participants variable was type of search interfaces (with or without auto-suggest). The within-participants variable was inclusion of the domain vocabulary (with or without domain vocabulary). In this section, we describe aspects of the evaluation method, beginning with the experimental system.

4.1 System

In order to focus on the auto-suggest interface and control system conditions, the feature of faceted search was deactivated and only the keyword search was made available in the school document portal site. The portal site was customized into four ways in accordance with the 2x2 factorial design. The baseline system provided only the keyword search function without the auto-suggest interface, where the domain vocabulary was not included in the indexing terms and the weight of indexing terms was calculated without considering the domain vocabulary. That is, the effect of X_i in the equation (1) is nullified while Y_i and Z_i is effective. On the other hand, the full-featured system provided the auto-suggest interface, and the domain vocabulary was considered for the indexing terms and their weighting function. In the other two systems, either of the auto-suggest feature or the domain vocabulary was taken into account in addition to the baseline system.

4.2 Method

Participants. A total of 32 undergraduate students were recruited from Faculty of Informatics, Kansai University. The participants were without prior experience of using the school document portal site, and paid to participate in the experiment.

Apparatus. We tested all the participants in a lab setting on the campus, using Mozilla Firefox 3 browser on Windows Vista workstations with 19-inch monitors set at 1280 by 1024 pixels in 24-bit color. The system ran on a server located in the campus, and data was recorded in the server log.

Experimental Tasks. Some search tasks are exploratory and requires searchers to gather information on a subject to allow them to perform some action [15]. In contrast to such an open-ended search, this portal is primarily used for seeking particular pieces of information or school documents in a target language. Therefore, a type of tasks to be investigated here is not exploratory search but known-item search [17]. The known-item search is to seek for a particular information item under the searchers' anticipation that such an item exists, and sometimes referred to as navigational queries in Web site search [18].

In this study, we suppose a situation where the school document portal is used to find out and download document files of a desired topic in a target language. Participants are asked to assume they are a teacher in an elementary school, and a situation looking for a Portuguese document that meets specified needs. Table 2 shows the descriptions of the experimental tasks presented to the participants.

Table 2. Descriptions of the search tasks

T1:	In your class, some pupils need detailed tuberculosis (TB) examination. Find a document to notify that the examination is TB reaction test and direct chest X-ray.
T2:	The schedule of the open class day has been fixed. There also will be a parent/guardian meeting after classes. Find a document to notify the date and plan of the day.
T3:	Currently mass cold infections are going around the schools in the city. Find a document informing on temporary closing of class and schedule changes.
T4:	For educational purposes, school operates recycling. Find a document to bring newspapers, magazines, cardboard of no use to school.

Procedure. One half of the participants were assigned to the system without the auto-suggest feature whereas another half were assigned to the system with the auto-suggest feature. In each system condition, the participants attempted the four search tasks in Table 2, where two tasks were assigned exclusively to each of the two term weighting conditions (with or without the domain vocabulary). In both systems, the task presentation order was carefully counterbalanced across participants.

The experiment was conducted in the following sequence. After answered a short demographic questionnaire, the participants were given an overview of the study and a brief explanation of the system. The participants were then asked to perform one practice task prior to the experimental tasks. For each of the experimental tasks, the participants were presented a task instruction and a sample copy of a school document that meet the requirements of the task at hand. There was no feedback from the system if the participants selected a correct document or not. This situation reflects the nature of the known-item task in the sense that the searchers need to determine whether an expected information item is actually found. In addition, the search time is limited to at most 5 minutes for each task because it is not realistic for Web users to spend more time for a know-item search task. Finally, the participants were asked to complete a short questionnaire about their experience.

5 Results and Discussions

In the following analyses, the level of significance was set to $p < 0.05$ for all comparisons.

5.1 Search Success Rate

Table 3 shows average rates of search success achieved in the limited search time within 1, 2, or 3 minutes. In each of the search time limit, a two-way ANOVA

Table 3. Average rate of search success (SD in parentheses)

	NAS		AS	
Time (min)	NDV	DV	NDV	DV
1	0.09 (0.20)	0.28 (0.31)	0.22 (0.31)	0.31 (0.24)
2	0.38 (0.38)	0.63 (0.33)	0.59 (0.40)	0.69 (0.24)
3	0.56 (0.39)	0.84 (0.23)	0.84 (0.23)	0.81 (0.24)

NAS = no auto-suggest, AS = with auto-suggest,
NDV = no domain vocabulary, DV = with domain vocabulary.

was used to determine if there was a significant difference in the success rate between the search interface type and the inclusion of the domain vocabulary. There were significant main effects of the inclusion of the domain vocabulary in the search time limit of 1 minute $[F(1, 30) = 4.175, p < 0.05]$ and 2 minutes $[F(1, 30) = 4.291, p < 0.05]$. In each case, the success rate was higher when the domain vocabulary was included. There was no significant interaction between the factors in both cases.

On the other hand, when the search time was limited within 3 minutes, a significant main effect was found for the search interface type $[F(1, 30) = 4.342, p < 0.05]$, and the success rate was higher when the auto-suggest interface was used. There was a tendency of interaction $[F(1, 30) = 3.488, 0.05 < p < 0.10]$ in this case. This tendency was revealed due to the considerably low success rate (0.56) in the baseline system, in which neither the auto-suggest interface nor the domain vocabulary was provided, while the success rates in the other three conditions were greater than 0.80.

According to the above results, it was demonstrated that the domain vocabulary contributed to the improvement of the success rate during the first two minutes. In contrast, the effect of the auto-suggest interface became more influential when the search time elapsed had exceeded two minutes. The rates of search success in Table 3, however, consider only the success or failure of the search. In order to further investigate the effectiveness of the search assist feature, we focus on the number of queries in each task as well as precision and recall.

It is important to note here that when the search time limit was three minutes, mean plus standard deviation of the success rate went beyond 1.0 or the maximum value in the three of the four conditions (Table 3). This situation suggests a ceiling effect that may depress averages below the true values. This effect might be present due to the relatively decreased complexity of the search task when performed for more than three minutes. In addition, when the search time was limited to one minute, the maximum success rate was 0.31. Thus the one-minute search time is too short to be considered for the assessment of the search effectiveness. In the following analyses, therefore, we take the case of the two-minute search time into account.

Table 4. Mean values in the case of two-minute search time (SD in parentheses)

	NAS		AS	
	NDV	DV	NDV	DV
Number of queries	3.16 (1.34)	2.16 (0.72)	1.66 (0.58)	1.41 (0.44)
Precision	0.05 (0.04)	0.13 (0.09)	0.13 (0.05)	0.19 (0.07)
Recall	0.48 (0.20)	0.66 (0.22)	0.95 (0.09)	0.72 (0.16)

NAS = no auto-suggest, AS = with auto-suggest,
NDV = no domain vocabulary, DV = with domain vocabulary.

5.2 Number of Queries

The number of queries formulated in each task is shown in Table 4. As will be noted from the table, the number of queries was reduced to less than 2.0 by the auto-suggest interface, regardless of the inclusion of the domain vocabulary. A two-way ANOVA was conducted to determine if there was a significant difference in the average number of queries between the search interface type and the inclusion of the domain vocabulary. Significant main effects were found for the search interface type $[F(1, 30) = 27.380, p < 0.001]$ and the inclusion of the domain vocabulary $[F(1, 30) = 7.797, p < 0.01]$. There was no significant interaction between the factors. In each case, the number of queries was smaller when either the auto-suggest interface or the domain vocabulary was adopted.

It is reported that the mean number of queries[7] in a user session is 1.73 (with a median of 1) for Web site search in a local government site [10]. This value is much lower than the mean value 4.86 (with a median of 8) observed in the case of a general-purpose search engine [9]. Since the type of search dealt in this study is not a general-purpose search but a Web site search, the number of queries improved by the auto-suggest seems to be a moderate level with regard to the case of the local government site. Needless to say, the number of queries alone would not be a good indicator of the search effectiveness. However, if we consider the high success rate as well as the reduced number of queries, it is reasonable to acknowledge the effect of the auto-suggest interface.

5.3 Precision and Recall

Table 4 shows precision and recall averaged across queries formulated by each participant within the two-minute search time. A two-way ANOVA was performed to examine differences in the precision and recall. ANOVA for recall revealed that there was a significant interaction $[F(1, 30) = 19.486, p < 0.001]$ between the search interface type and the inclusion of the domain vocabulary. As for the precision, significant main effects were found for the search interface type $[F(1, 30) = 26.348, p < 0.001]$ and the inclusion of the domain vocabulary $[F(1, 30) = 10.942, p < 0.005]$. In each case, the precision was improved, and

[7] Not as unique queries but queries including repeat queries.

there was no significant interaction between the factors. In recent studies, the search performance is often assessed on the basis of the highly ranked items such as the top 10 results, and high recall is not always the standard assumption in the assessment of search results [11]. Therefore, regardless of no significant improvement in the recall, it is noteworthy that the precision has been significantly improved in this study.

In our auto-suggest interface, when a term is selected from the suggestions, a list of school documents are displayed as candidates (Figure 3). Currently, such candidates are listed in the order of document titles[8], and sometimes it is not straightforward for the users to find out a desired item from a lot of candidates. A previous study investigated the ways of organizing items suggested by auto-completion in a known-item search task [19]. It was concluded that alphabetical order was suitable for a domain-independent lexical thesaurus, while domain-dependent grouping meaningful for users should be chosen carefully because every grouping strategy did not improve the search performance [19]. Further research will be needed for the organization of the suggested items, so that the quality of the search results can be elaborated.

6 Concluding Remark

In this paper we demonstrated the potential effectiveness of the auto-suggest interface, which exhibited better performance when used with domain vocabulary. The domain vocabulary utilized in this study was developed manually by school teachers and translation staff members. The effectiveness of the auto-suggest interface, therefore, may be dependent on the quality of the domain vocabulary. It is important to further investigate ways of developing vocabulary with semantic relations, so that the dynamic term suggestion technique can be applicable to the broader context other than the school education domain.

It would also be necessary to continue the evaluation with real users of multilingual school documents because the experiment presented in this paper was conducted with undergraduate students. In addition, currently the auto-suggest interface is integrated with the search box of the portal, but not fully integrated with the faceted search. We are planning to further investigate the effectiveness of the auto-suggest interface in cooperation with real users by using the full-featured faceted search system.

References

1. http://www.mext.go.jp/b_menu/shingi/chousa/shotou/042/houkoku/08070301.htm (in Japanese, accessed 2010-03-18)
2. http://www.mext.go.jp/b_menu/houdou/21/07/1279262.htm (in Japanese, accessed 2010-03-18)

[8] In the order of Japanese phonetic characters, which corresponds to the alphabetical sequence in English.

3. Hori, M., Ohnishi, N., Kita, C.: Development of a portal site for sharing multilingual school documents: Design and evaluation of classification scheme structuralized by card sorting. Information Processing Society of Japan (IPSJ) Journal 52(2), 590–603 (2010) (in Japanese)
4. Hori, M., Kita, C.: Toward multilingual support of educational document sharing based on the language grid. In: Proceeding of the 2009 International Workshop on Intercultural Collaboration, Palo Alto, CA, pp. 237–240 (2009)
5. Broughton, V.: The need for a faceted classification as the basis of all methods of information retrieval. Aslib Proceedings 58(1/2), 49–72 (2006)
6. Tunkelang, D.: Faceted Search. Morgan and Claypool Publishers (2009)
7. Sutcliffe, A.G., Ennis, M., Watkinson, S.J.: Empirical studies of end-user information searching. Journal of the American Society for Information Science and Technology 51(13), 1211–1231 (2000)
8. Markey, K.: Twenty-five years of end-user searching, Part 2: Future research directions. Journal of the American Society for Information Science and Technology 58(8), 1123–1130 (2007)
9. Spink, A., Wolfram, D., Jansen, B.J., Saracevic, T.: Searching the Web: The public and their queries. Journal of the American Society for Information Science and Technology 52(3), 226–234 (2001)
10. Chau, M., Fang, X., Sheng, O.R.L.: Analysis of the query logs of a Web site search engine. Journal of the American Society for Information Science and Technology 56(13), 1363–1376 (2005)
11. Hearst, M.A.: Search User Interfaces. Cambridge University Press, Cambridge (2009)
12. Anick, P., Kantamneni, R.G.P.: A longitudinal study of real-time search assistance adoption. In: Proceedings of SIGIR 2008, New York, NY, pp. 701–702 (2008)
13. Yee, K.-P., Swearingen, K., Li, K., Hearst, M.: Faceted metadata for image search and browsing. In: Proceedings of CHI 2003, Ft. Lauderdale, Florida, pp. 401–408 (2003)
14. Efthimiadis, E.N.: Query expansion. Annual Review of Information Science and Technology 31, 121–187 (1996)
15. White, R.W., Marchionini, G.: Examining the effectiveness of real-time query expansion. Information Processing and Management 43(3), 685–704 (2007)
16. Sihvonen, A., Vakkari, P.: Subject knowledge improves interactive query expansion assisted by a thesaurus. Journal of Documentation 60(6), 673–690 (2004)
17. Allen, B.: Recall cues in known-item retrieval. Journal of the American Society for Information Science and Technology 40(4), 246–252 (1989)
18. Broder, A.: A taxonomy of Web search. SIGIR Forum 36(2), 3–10 (2002)
19. Amin, A., Hildebrand, M., van Ossenbruggen, J., Evers, V., Hardman, L.: Organizing suggestions in autocompletion interfaces. In: Boughanem, M., et al. (eds.) ECIR 2009. LNCS, vol. 5478, pp. 521–529 (2009)

Intercultural Collaboration Support System Using Disaster Safety Map and Machine Translation

Yoshiyasu Ikeda, Yosuke Yoshioka, and Yasuhiko Kitamura

School of Science and Technology, Kwansei Gakuin University
2-1 Gakuen, Sanda-shi, Hyogo 669-1337, Japan
ykitamura@kwansei.ac.jp

Abstract. Natural Disaster Youth Summit (NDYS) is an intercultural collaboration project promoted by a NPO called JEARN (Japan Education And Resource Network). They are working on collaborative learning about disaster prevention by exchanging disaster safety maps created by students over the world and by discussing over BBS's and video conference systems on the Internet. Language barrier is the most difficult issue in this activity, and in order to communicate freely in their own native language, linguistic support, such as using machine translation systems, is necessary. We develop CoSMOS, an intercultural collaboration support system using disaster safety map and machine translation, and discuss collaboration supports using CoSMOS from four viewpoints; handling high-definition digital images, linking disaster safety maps to the world map, supporting collaborative learning, and supporting multiple languages.

Keywords: Intercultural collaboration, Machine translation.

1 Introduction

Collaboration is an interactive process where two or more people work together to achieve a common goal. The development and proliferation of the Internet technology enable us to communicate with people in distant places and have potential to facilitate intercultural collaboration where people in different culture and language collaborate with each other [1].

Natural Disaster Youth Summit (NDYS)[1] is an intercultural collaboration project run by JEARN(Japan Education And Resource Network). JEARN is the Japanese branch of iEARN which is the largest non-profit global education network in the world. The NDYS project tries to establish an education network where students and their teachers over the world can collaborate with each other to learn disaster prevention by creating and exchanging disaster safety maps on which dangerous places and shelters around school are located [2].

The project so far utilizes conventional BBS's (Bulletin Board Systems) and video conference systems to support the collaboration work over the Internet. These tools can facilitate communication among them in distant places, but the difference of language is a most formidable obstacle to hinder intimate collaboration among them in different

[1] http://ndys.jearn.jp/ja/index.html

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 100–112, 2010.

culture. At present, English is used as the common language for communication, however, all the participants, who are middle or high school students, are not good enough to speak fluently in English and their teachers have to assist them a lot. In order to assist the students and their teachers to communicate in their native language, multi-linguistic supports using machine translation are expected [3].

We are developing a new intercultural collaboration support system called CoSMOS (Collaborative Safety Maps on Open System), which supports collaborative work among students over the world on disaster safety maps utilizing the Internet and machine translation technologies.

In this paper, we show design goals of CoSMOS to support intercultural collaboration on disaster safety maps in Section 2. We discuss the implementation issues of CoSMOS in Section 3 and intercultural collaboration using CoSMOS in Section 4. We especially discuss and compare two ways of multilingual translation, which is a key technology to support intercultural collaboration, in Section 5. We describe related works and compare CoSMOS with them in Section 6. Finally, we summarize this paper with our future work in Section 7.

2 Design Goals of CoSMOS

We need to achieve design goals as follows to support intercultural collaboration in the NDYS project.

1. Handling high definition digital images
2. Linking disaster safety maps to the world map
3. Supporting collaborative learning
4. Supporting multiple languages

2.1 Handling High Definition Digital Images

The main purpose of creating disaster safety maps is to make students locate dangerous places, safe places, and shelters around school and to let them learn how to prevent disasters from happening. Currently, disaster safety maps are created as posters by students, so transforming the maps into digital images is required to upload them onto the Web system available on the Internet. The map is not just a picture to see but it contains various disaster information. The messages and symbols depicted on the map to show the disaster information should be clearly readable on Web browsers, so it is required to transform the maps into high definition digital images to be readable in detail. The system should assist the students to handle the images easily.

2.2 Linking Disaster Safety Maps to the World Map

One of the goals of the NDYS project is to create a “global” disaster safety map by students over the world collaborating with each other. We need to create it by linking individual disaster safety maps to a world map. The students can learn the geographical relation among the maps from the global map, and understand the cultural background of each disaster safety map.

2.3 Supporting Collaborative Learning

Photo sharing Web site such as Flickr may be satisfactory only to share disaster safety maps over the Internet. However, in order to support collaborative learning about disaster prevention, we need to provide a way for the students to discuss with each other over disaster safety maps. Consequently, we need to incorporate a function to post comments on the disaster safety maps. Some students post comments to a map and the others reply to them so that they can learn collaboratively.

2.4 Supporting Multiple Languages

The students from various countries, such as Japan, Taiwan, Canada, Turkey, and so on are participating in the NDYS project, and various languages are used in each school. If the students can communicate with others in their native language, it is comfortable for them to work together with others. To make it possible, the system should support multiple languages by using machine translation. It is requested that comments submitted by the students should be translated automatically into other languages.

3 Intercultural Collaboration Support System CoSMOS

We are developing an intercultural collaboration support system CoSMOS as a Web system in which the students in different countries collaborate with each other using a Web browser. CoSMOS is developed based on WordPress[2], which is an open source blog software that is written in PHP and uses MySQL database. Four design goals, which are mentioned in the previous section, are achieved by implementing four functions: handling disaster safety maps, linking disaster safety maps to the world map, posting comments on disaster safety maps, and translating comments into multiple languages.

3.1 Handling Disaster Safety Maps

Students in each school create a disaster safety map in their own poster format style, so the map is taken as a picture by a digital camera, and is uploaded onto CoSMOS. To consider the intent of the students who create the disaster safety map, not only the drawing and/or painting but also the messages in the map should be clearly readable. To this end, the picture is taken as a high definition digital image over 10M pixels. However, we have to deal with two problems in order to display the image.

1. Loading time:
   As the number of pixels increases, the file size of the image becomes large. It takes a long time to load the entire image on a browser.
2. Operability of image browsing:
   When the size of image is over 10M pixels, the image usually overflows the window size of Web browser. It is not easy for the students, who are novice users of computer, to handle a large image.

[2] http://ja.wordpress.org/

To cope with the first problem, we refer to how Google Maps[3] loads the map image onto a browser. In Google Maps, the entire map is divided into multiple images with different resolutions, each of which corresponds to a zoom level and a range, and are aligned as a grid in the database. When we view a map of a specific range and resolution on a Web browser, only the corresponding image depending on the resolution and the range is selectively loaded from the database. The loading time is much less than that of loading the entire image. Google Maps Image Cutter [7] is a software that divides an image into multiple ones complying with the Google Maps format and generates HTML codes to display the images through the Google Maps interface. We use it to display disaster safety maps on CoSMOS.

To cope with the second problem, we use the zooming and dragging functions which are compatible with those of Google Maps because disaster safety maps are displayed using the Google Maps interface. Students can easily zoom out/in to view the entire/detail part of a disaster safety map by using a mouse.

3.2 Linking Disaster Safety Maps to the World Map

One of the goals of the NDYS project is to create a "global" disaster safety map by students over the world collaborating with each other. We create such a global map by linking disaster safety maps to a world map. Each disaster safety map is uploaded on Cosmos as an individual page, and a link to the the world map is attached using the Google Maps API[4]. The reason why we choose the Google Maps API is because Google Maps covers the entire world, and it can be customized by utilizing its functions.

In order to embed the Google Maps in CoSMOS, we use the marker function to create a link from a disaster safety map to the world map. The metadata of the disaster safety map is stored in the Disaster Safety Map database, which is implemented by using MySQL, in the following format.

[ID], [Latitude], [Longitude], [Page ID]

"Latitude" and "Longitude" represent the coordinate of the disaster safety map on the world map. "Page ID" is the URL of the page on which the disaster safety map is uploaded.

3.3 Posting Comments on Disaster Safety Maps

CoSMOS has a function to paste a note on a disaster safety map to support collaborative learning over the maps.

Since a student can paste a note anywhere on the map, he/she can pinpoint a specific part of the map in which he/she has a comment or a question. The other student who created the map can reply to the pasted notes, so they can communicate with each other in a Q&A style.

We utilize the marker function of Google Maps to indicate the position of note on a disaster safety map by specifying the coordinate of the note on the map. Note data are stored in the Note database implemented by MySQL in the format below.

[3] http://www.google.co.jp/maps
[4] http://code.google.com/intl/en/apis/maps/

[Note ID], [Map name], [X-coordinate], [Y-coordinate], [Date posted], [Content], [Parent note ID], [Input language], [User ID]

X and Y coordinates are the latitude and the longitude on Google Maps to paste the note. A note and the replies create a thread and "Parent note ID" specifies the parent note of the note in the thread.

3.4 Translating Comments into Multiple Languages

CoSMOS supports five languages: Japanese, English, Turkish, French, and Chinese (Traditional), so students, who speak one of the languages, can use their language to collaborate with others.

A messages submitted to CoSMOS as a note is automatically translated into four languages by using machine translation services provided by the Language Grid [9], which is a platform where we can integrate a number of language services and resources to produce various multi-lingual applications on the Internet. When an original messages is written in a language but English, the message is translated into English first, and then translated into other languages. This means that English is used as an intermediate language in CoSMOS. Fig. 1 shows the translation flow and the actual translation services used in CoSMOS.

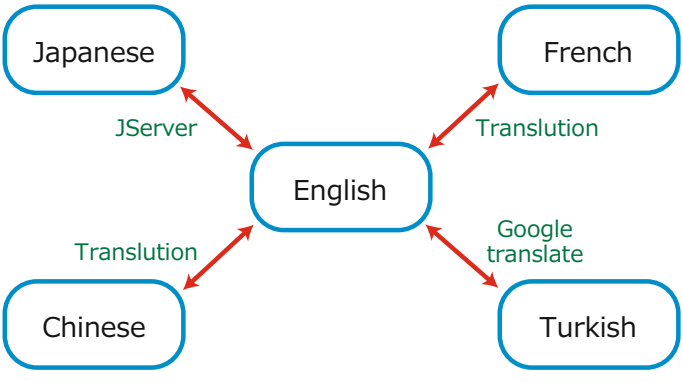


Fig. 1. Translation flow and the name of translation services

For example, when a Japanese message is translated into Chinese, the message is translated into English by using JServer and is translated again from English into Chinese by using Translution.

In addition, the Language Grid provides a Japanese-English glossary on natural disasters and dictionaries of academic terms, and an English-French glossary on natural disasters. When a translation service between these languages are activated, the system automatically consult the glossaries and dictionaries to improve the quality of translation.

A messages input from a student is translated into multiple languages, so multiple messages in different languages need to be stored. To this end, we utilize a qTranslate function[5], which is a plug-in of WordPress. We can handle multi-lingual messages by composing a message in the following format.

<!−−: Language code of language 1 −−>a message in language 1
<!−−:−−><!−−: Language code of language 2 −−>a message in language 2
<!−−:−−>c<!−−: Language code of language n −−>a message in language n
<!−−:−−>

For example, a message for greeting is stored as follows.

<!−−: en −−>Hello. <!−−:−−><!−−: ja −−>B <!−−:−−><!−−: fr −−>Bon jour. <!−−:−−>

When a message is displayed in CoSMOS, the specific message is extract from the multi-lingual one depending on the user language. If the code of the user language is French (fr), “Bon jour.” is extracted and shown on the page by using the qTranslate function.

As shown in Fig. 2, a note is actually displayed depending on the user language in CoSMOS.

Fig. 2. Note shown in multiple languages

Machine translation services often produce incorrect translations. To cope with this problem, the system displays back-translation [10], so the student can check the quality of translation, and fix his/her input message if necessary. Back-translation is to reversely translate the translated result in the target language into one in the input language. If the input text and its back-translation are similar in meaning, the translation is estimated to

[5] http://www.qianqin.de/qtranslate/

be correct. On the other hand, if the back-translation does not make sense, the translation is estimated to be incorrect. The student fixes his/her input until the input text and its back-translation look similar.

4 Intercultural Collaboration Support Using CoSMOS

CoSMOS[6] has been available on the Internet. Currently, 15 disaster safety maps from 12 countries are uploaded on CoSMOS. Anyone can view disaster safety maps, but only registered users can post notes to prevent from spams.

CoSMOS can display the contents in one of five languages; Japanese, English, Turkish, French, and Chinese (Traditional) and there are buttons to switch the user language on the top bar of CoSMOS.

Fig. 3. Top page of CoSMOS

4.1 Viewing Disaster Safety Maps

The top page of CoSMOS is the world map shown in Fig. 3. Individual disaster safety maps created by students over the world are linked from the world map.

A red marker indicates the location of a disaster safety map. When we move a mouse cursor over a marker, a balloon to describe a summary of the corresponding disaster

[6] http://www.kitamura-lab.jp/cosmos/

Fig. 4. Viewing an entire disaster safety map

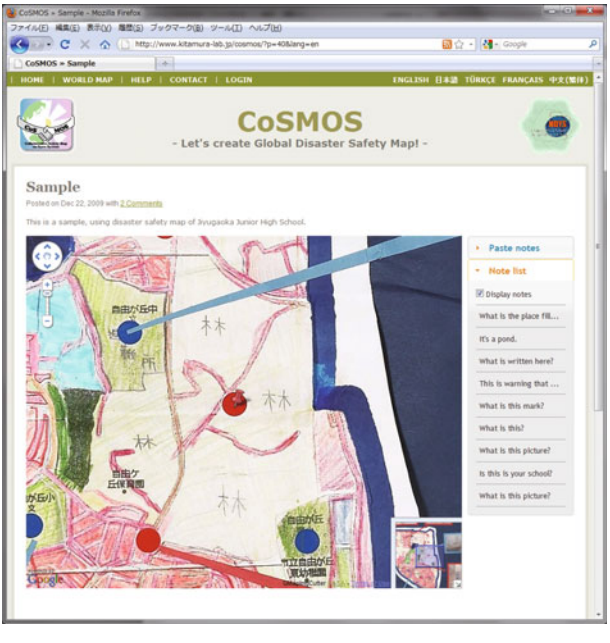


Fig. 5. Zooming in a disaster safety map

safety map appears with a link to the map. If we click the link, the corresponding map, which is embedded in the Google Map interface, appears as in Fig. 4.

We can use the Google Maps Controls to move around and zoom in/out the map. If we zoom in the map as shown in Fig. 5, we can read the detail of the map. A red pin represents the location of a note pasted on the map. If we put the mouse cursor over the red pin, the corresponding note pops up and we can read the comments posted to the map as shown in Fig. 2.

4.2 Pasting Notes to Disaster Safety Maps

As a style of collaborative learning, some students can paste a note on a disaster safety map and others can reply to the pasted note. Because a lot of information are contained in a disaster safety map, it is easy to understand what the message or question means by pasting a note at a specific point on the map. In addition, the note is translated into multiple languages.

In order to paste a note, we first press the “Paste notes” button on the left bar and point the location to paste a note. We then write a message in the dialog shown in Fig. 6 following the process below.

First, we select a language to write from a list of “Input language.” Second, we input a message in the “Message” text box, and press the “Translate” button. If the

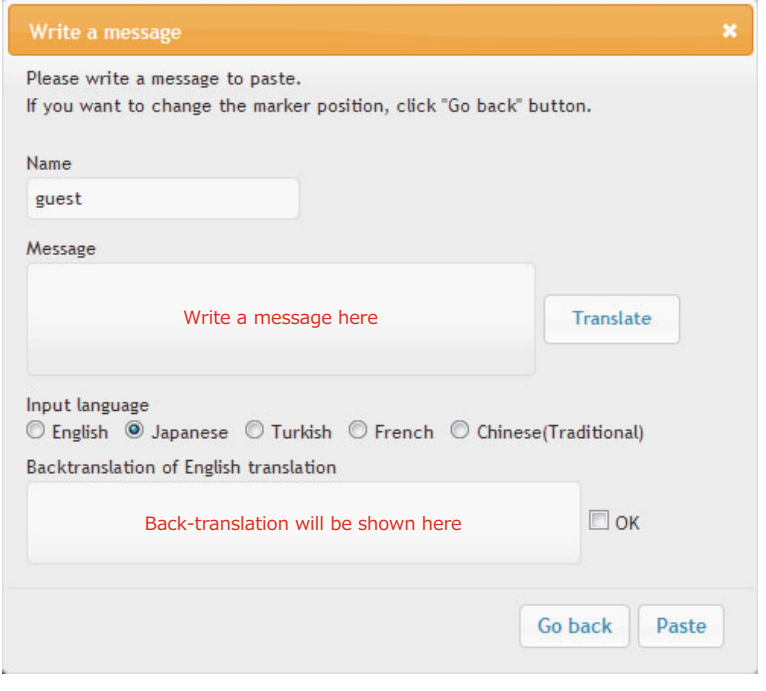


Fig. 6. Message input dialog

input language is not English, the back-translation of the input message is shown in the "Back-translation of English translation" text box. Comparing the original message and the back-translation, if they are similar in meaning, we check the "OK" check-box to confirm the translation.

However, if the back-translation does not make sense, we fix the original message and retranslate it until the back-translation makes sense. Finally, after confirming the back-translation, we press the "Paste" button. The message is translated into three languages except English and the input language, and the set of messages in five languages is pasted on the disaster safety map as a note, as mentioned in the previous section.

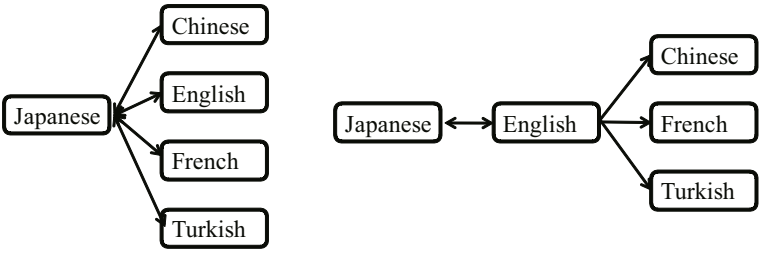
When we write a message in English, we just press the "Paste" button. The message is translated into four languages except English and is pasted on the map. In this case, there is no way to confirm the translation quality. Discussion on how to translate a message into multiple languages is given in the next section.

5 Multilingual Translation

When we initially designed the translation process of CoSMOS, we displayed the back-translation for four languages except the input language, so the user has to check the quality of translation in each language referring to the back-translation, as shown in Fig. 7(a). However, this method takes a long time to translate the input message and retrieve its back-translations. It needs 8 translations (one regular translation and one back-translation for each of 4 languages). In addition, if some back-translations are incorrect, the user has to repeat to fix and retranslate the message until the back-translations make sense. These process increases the burden of the user to post messages.

As a remedy to cope with this problem, we use only English to confirm the translation quality, as shown in Fig. 7(b). When the input message is not English, the system translates it into English once and back-translate the English translation to confirm the translation quality. If we confirm the quality, the English translation is further translated into the other languages with no confirmation.

This method takes a shorter time to translate a message than the initial method. It needs only 5 translations (four regular translations and one back-translation). To



(a) Direct translation (b) English-mediated translation

Fig. 7. Two methods of multilingual translation (from Japanese)

Table 1. Average processing time of translation (in sec.)

	All languages	English only
Sentence A (30 letters)	7.56	1.16
Sentence B (19 letters)	7.11	1.19
Sentence C (12 letters)	9.71	1.13
Sentence D (25 letters)	9.81	1.42
Sentence E (44 letters)	8.12	1.24
Overall average	8.46	1.23

evaluate an actual performance, we translated 5 Japanese sentences for 10 times, and measured the average processing time of each translation in both methods. Table 1 shows the result.

The processing time of the English-mediated translation is 1.23 seconds in average and that of the direct translation is 8.46 seconds in average. If the number of languages to be translated increases, the processing time increases significantly.

6 Related Works

We discuss the advantages of CoSMOS by comparing with conventional systems, as shown below, to support intercultural collaboration.

- TransBBS
  TransBBS[6] is a multilingual bulletin board system. The system is the most basic type of intercultural collaboration support system and supports text based collaborations. By using machine translation services, submitted messages are automatically translated into other languages on the BBS.
- AnnoChat
  AnnoChat[5] is a multilingual chat system. The system has an annotation function, in which an image or text message can be attached to a term as a note, to assist the users in different cultural background to understand the meaning of the term.
- Flickr
  Flickr[7] is a representative photo-sharing community system. The users can easily share photos on the Web system by uploading them. They can collaborate with each other by submitting comments to photos.
- Google Maps
  Google Maps[8] is a most widely used GIS (Geographical Information System) in the world. The system stores, manages, and presents various types of geographical

[7] http://www.flickr.com/
[8] http://maps.google.com/

and map data all over the world. The users can append messages on the map by using the My Maps function.

Table 2 shows how each system achieves four goals that are required to support the NDYS project, mentioned in Section 2. Only CoSMOS can achieve all the goals.

Table 2. Average processing time of translation (in sec.)

	High-definition image	Link to the world map	Collaborative learning	Multiple language support
TransBBS	No	No	Partly	Yes
AnnoChat	Partly	No	Partly	Yes
Flickr	Yes	No	Partly	No
Google Map	Yes	Yes	No	No
CoSMOS	Yes	Yes	Yes	Yes

6.1 Handling High Definition Digital Images

TransBBS cannot handle images at all. AnnoChat can attach an image, but it is difficult to attach a high-definition one to display a disaster safety map. An image in AnnoChat is just an annotation to explain the meaning of term. Flickr,Google Maps and CoSMOS can handle high definition images.

6.2 Linking Disaster Safety Maps to the World Map

There is no way to link disaster safety maps to the world map in TransBBS, AnnoChat and Flickr. Google Maps and CoSMOS can provide a world map to locate disaster safety maps.

6.3 Supporting Collaborative Learning

TransBBS and Flickr provide a BBS to support collaboration over a disaster safety map, but the function is limited because the users can submit a message to an entire map but not to a specific part of the map by pasting a note. In AnnoChat, an image is used as an annotation to a term and it is difficult for the users to collaborate with each other over a disaster safety map. Google Maps has no way to collaborate over the map. CoSMOS has a function to paste a note on a specific part of a disaster safety map.

6.4 Supporting Multiple Languages

TransBBS, AnnoChat, and CoSMOS utilize machine translation services, so messages are translated into multiple languages. Flickr and Google Maps support no machine translation, so the users have to collaborate with each other by using a specific language such as English.

7 Summary

We are developing a new intercultural collaboration support system CoSMOS, which uses disaster safety maps and machine translation to support the Natural Disaster Youth Summit project. High-definition image of disaster safety maps created by students are posted on CoSMOS, and students can view the map in detail on a Web browser. We create a global disaster safety map by linked individual maps to a world map. As a support for collaborative learning over disaster safety maps, we implemented a function to paste notes on the map. Students can post notes to the maps and reply to the notes. In addition, CoSMOS supports five languages by using machine translation services provided by the Language Grid and this makes the students collaborate with each other by using their native language.

We introduced CoSMOS at "NDYS Forum 2010," which was held in Kobe, Japan on January 24, 2010, and it has been available on the Internet. After gathering the log data, we are going to evaluate the system from the viewpoint of collaboration support and what kind of intercultural collaboration can be done over CoSMOS.

References

1. Ishida, T., Fussell, S.R., Vossen, P.T.J.M. (eds.): IWIC 2007. LNCS, vol. 4568. Springer, Heidelberg (2007)
2. Naya, Y.: How Intercultural Disaster Reduction Education Change Students: A Case Study of an Evening Course Senior High School in Hyogo. In: Ishida, T., Fussell, S.R., Vossen, P.T.J.M. (eds.) IWIC 2007. LNCS, vol. 4568, pp. 368–381. Springer, Heidelberg (2007)
3. Naya, Y.: Utilizing 'Langrid Input' for Intercultural Communication in Senior High School in Japan. In: Proc. of International Workshop on Intercultural Collaboration, pp. 281–284 (2009)
4. Kido, T.: Information Sharing for Diversity Collaboration. IPSJ SIG Notes ICS-99(47), 25–30 (1999)
5. Fujii, K., Shigenobu, T., Yoshino, T.: Evaluation of Annotation in Intercultural Communication Chat Tool Using Machine Translation. Trans. Information Processing Society of Japan 48(1), 63–71 (2007)
6. Nomura, S., Ishida, T., Funakoshi, K., Yasuoka, M., Yamashita, N.: Intercultural Collaboration Experiment 2002 in Asia: Software Development Using Machine Translation. Trans. Information Processing Society of Japan 44(5), 503–511 (2003)
7. Hudson-Smith, A., Milton, R., Batty, M., Gibin, M., Longley, P., Singleton, A.: Public Domain GIS, Mapping & Imaging Using Web-based Services. In: CASA Working Paper 120, Centre for Advanced Spatial Analysis, London (2007)
8. Shigenobu, T., Fujii, K., Yoshino, T., Nadamoto, A.: Language Grid: Development of Support Environment for Intercultural Collaboration. In: Annual Conference of the Japanese Society for Artificial Intelligence, pp. 3I1-6 (2007)
9. Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration. In: IEEE/IPSJ Symposium on Applications and the Internet (SAINT 2006), pp. 96–100 (2006)
10. Shigenobu, T.: Evaluation and Usability of Back Translation for Intercultural Communication. In: Aykin, N. (ed.) HCII 2007. LNCS, vol. 4560, pp. 259–265. Springer, Heidelberg (2007)

Development of a Multilingual Translation Service for Interpretations and Usage Examples of Mobile Phone Pictograms

Seiya Tanaka, Aika Nakakubo, Momoki Kimura, Kazuya Takeda, and Tomoko Koda

Faculty of Information Science and Technology, Osaka Institute of Technology
1-79-1 Kitayama, Hirakata-shi, Osaka, 573-0196 Japan
+ 81 72 866 5182
koda@is.oit.ac.jp

Abstract. To find mobile phone pictograms that have age-specific or gender-specific differences of interpretations, we surveyed 276 Japanese mobile phone users and compiled over 6,600 of their interpretations and usage examples for these pictograms. We used these reported pictogram interpretations and usage examples to develop a prototype of a web-based age-specific pictogram dictionary application that can be searched and viewed by specifying the pictogram, age group, and/or gender as search conditions. By applying Language Grid machine translation services to this application, we developed a web-based service that translates these pictogram interpretations and usage examples into several languages. This multilingual translation service for our age-specific pictogram dictionary enables the pictogram interpretations and usage examples reported by the Japanese survey respondents to be presented to foreign residents of Japan. It should help expedite and facilitate communication on mobile phones across different languages and cultures.

Keywords: pictogram, cell phone, mobile communication, intercultural communication, machine translation, web service computing, Language Grid.

1 Introduction

The rapid growth of telecommunications technology in recent years has resulted in a rapid rise in popularity of mobile media such as mobile phones and PHS terminals. The Japanese now consider the mobile phone an indispensable tool for daily life. A MyVoice Communications survey found that 95.0% of Japan's population use mobile phones, 94.5% of mobile phone users send and receive email by mobile phone, and 69.6% of these email users use pictograms in their email [1].

Users conveying information by means of communication tools such as email or chat often tend to misunderstand the emotional content of messages when they rely on text to express emotions or subtle nuances that in a face-to-face conversation would be expressed and conveyed nonverbally by means such as facial expressions, body language and tone of voice [2, 3]. So users often use pictograms as a way to convey

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 113–126, 2010.

emotions and subtle nuances difficult to convey with just text. Pictograms are used to convey emotions, and are also sometimes used in place of words. Figure 1 shows the use of a pictogram on a mobile phone to express an emotion, while Figure 2 shows the same pictogram used in place of a word. The musical note pictogram in Figure 1 is being used to mean‘fun’, while the same pictogram in Figure 2 is being used to mean ‘music’.

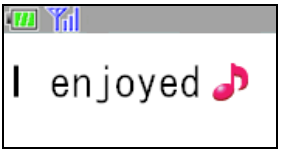


Fig. 1. Example of pictogram used to express emotion

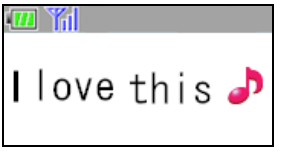


Fig. 2. Example of pictogram used in place of word

But interpretations of pictograms vary from individual to individual—not everyone interprets them the same way. A study of cultural differences designed to investigate whether children in Japan and the US interpret pictograms in the same way was carried out by giving the children a questionnaire survey asking for their interpretations of various pictograms. It found that the children in each country had different interpretations of pictograms related to six areas—gestures, color and gender, time, space, familiar objects and faces/facial expressions [4]. And as shown in Figure 3, the authors of this paper have found that during routine exchanges of mobile phone emails with their mothers, their mothers’ interpretations of the pictograms used can sometimes differ from their own interpretations. These findings suggest that the interpretation of some pictograms may be not only culture-specific but also age-specific.

So a way to minimize miscommunication caused by age-specific and culture-specific differences of interpretation of pictogram meaning would help mobile phone pictograms carry out their intended function of conveying emotions and subtle nuances. The aims of this research were (1) to investigate age-specific pictogram interpretations by collecting pictogram interpretations from various age-groups (2) to develop a prototype of an age-specific pictogram dictionary application that lets users search and view pictogram interpretations and usage examples, and (3) to develop a prototype of a web-based service providing multilingual translations of the interpretations and usage examples in the search results returned by the dictionary application.

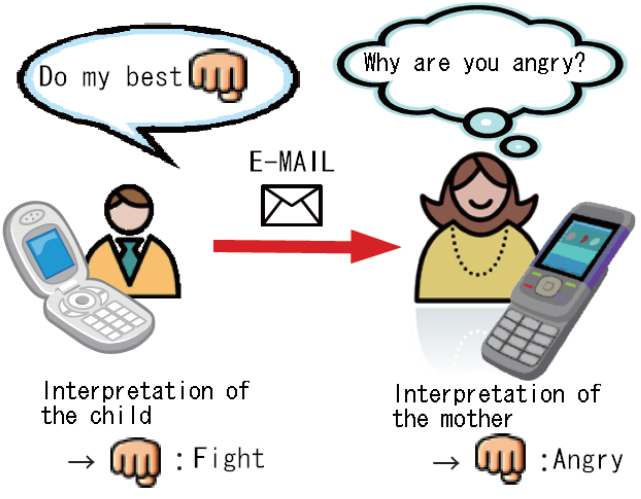


Fig. 3. Example of misunderstanding caused by different interpretations of same pictogram by age

The research overview is as follows: (1) We collected questionnaire responses from 276 Japanese mobile phone users in five age groups ranging from teenagers to those 50 or older. We compiled total of more than 6,600 interpretations and usage examples for mobile phone pictograms, then analyzed the compiled questionnaire data to find pictograms with age-specific or gender-specific differences of interpretation. (2) We then developed a web-based prototype of a pictogram dictionary application that lets users search and view its more than 6,600 data items by specifying an age group and/or gender as search conditions. (3) We used web-based machine translation services provided by Language Grid to create a multilingual translation service that provides translations of the pictogram interpretations and usage examples into several languages. To enable this service, we first created a standard-language correspondence table to convert colloquialisms into language that could be understood by a machine translation service. Language Grid is a system that lets users turn existing language resources on the internet into web services they can freely combine to create new language services [5].

Section 2 describes the questionnaire survey we conducted to investigate age-specific and gender-specific differences of interpretation of mobile phone pictograms, and describes the pictograms that had age-specific and gender-specific differences of interpretation. Section 3 provides an overview of the age-specific pictogram dictionary application we developed for this research. Section 4 describes the multilingual translation service we developed for this research. Section 5 describes the feedback and comments we received. Section 6 is the conclusion.

2 Pictogram Interpretations and Usage Examples

This section provides an overview of our questionnaire survey on pictogram interpretations, and describes our analysis of the compiled questionnaire results.

2.1 Pictogram Selection by Preliminary Survey

Since it would have been difficult for our questionnaire survey of pictogram interpretations to cover every one of the large number of pictograms used on mobile phones, we conducted a preliminary survey for each carrier (au, DoCoMo and Softbank) to select the pictograms to be covered.

The respondents were 38 male and female university students. We asked the respondents representing each carrier which pictograms they frequently use. Based on their responses, we selected 24 pictograms for each carrier, each with a usage frequency of at least 50%. Figure 4 shows the 24 pictograms selected for each carrier. Our pictogram interpretation questionnaires covered these 24 pictograms for each carrier.

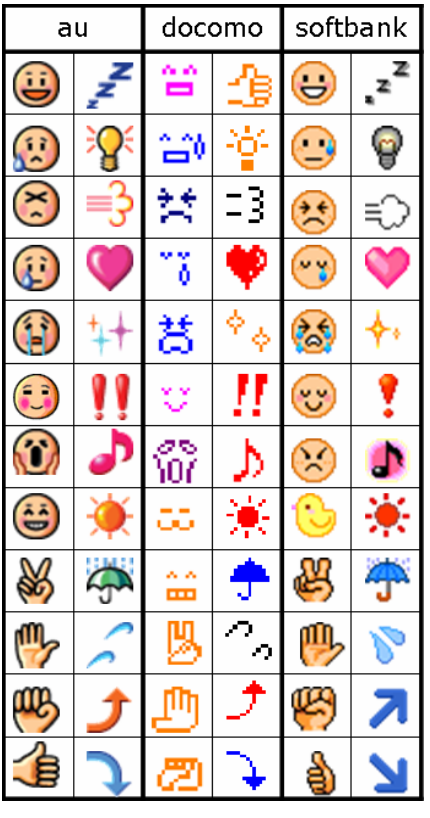


Fig. 4. Pictograms covered by questionnaires

2.2 Pictogram Interpretation Questionnaires

Our questionnaires were designed to compile pictogram interpretations and usage examples. Before we distributed the questionnaires, we asked respondents which mobile phone carrier they use and gave them the questionnaire created for their

carrier. To specify the demographic attributes of the respondents, we asked each respondent's gender and age group (teens/twenties/thirties/forties/fifties or older). We then asked the following questions for each of the 24 pictograms selected by the preliminary survey:

- Do you use this pictogram? (Yes/No)
- How often do you use this pictogram? (1 = Hardly ever use, through 4 = Often use), or a usage frequency of 0 if the respondent answered 'No' to the previous question.
- What is your interpretation of this pictogram? (Free response)
- Provide a usage example of this pictogram. (Free response)

We received responses from 276 respondents of both genders and all five age groups, from questionnaires given between October 1 and November 20, 2008, and between July 1 and October 31, 2009.

2.3 Questionnaire Results

We tallied the interpretations of all 72 (24 × 3 mobile phone carriers) pictograms, and after standardizing any variations in the presentation of identical interpretations among the 6,624 interpretations (such as '[Laughter]' and 'Laughing') to forms agreed on unanimously by the three authors of this paper, we tallied the number of respondents who reported using each interpretation. We then calculated the usage frequency of each interpretation as a percentage of the respondents using that pictogram, and used these percentages to find pictograms with age-specific or gender-specific differences of interpretation.

We defined a pictogram with an age-specific interpretation when one of its interpretations accounts for more than 20% of its all interpretation responses in the same age group, and the percentages of the interpretation have wider than 20% among the ones in other age groups. Gender-specific pictograms are also defined in the same way. We found 13 pictograms that show tendency of age-specific or gender-specific differences of interpretation in the following pictograms—pictograms showing (1) gestures, (2) emotions, and (3) faces/facial expressions. We defined a pictogram with no difference of interpretation to be a pictogram with the same interpretation reported by all the demographic groups, with the interpretation accounting for more than 50% of the responses from each group. We found eight pictograms with a tendency of having no differences of interpretation. Pictogram interpretations and examples are described in detail in the sections below.

2.3.1 Pictograms with a Tendency of Age-Specific Differences of Interpretation

This section describes pictograms with a tendency of age-specific differences of interpretation. In the sections below, 'AG1' denotes respondents aged 10 to 19, 'AG2' respondents aged 20 to 29, 'AG3' respondents aged 30 to 39, 'AG4' respondents aged 40 to 49, and 'AG5' respondents aged 50 or older.

- Pictograms showing gestures
  The number of answers for the pictogram shown on the left of Figure 5 is as follows: AG1 (n=16), AG2 (n=25), AG3 (n=10), AG4 (n=5), AG5 (n=6). Interpretation as 'Happy' in the pictogram accounts for more than 20% in

AG1-AG3 (AG1 (n=5), AG2 (n=6), AG3 (n=4)), while this interpretation is not shared by AG4 and AG5.

- Pictograms showing emotions
  The number of answers for the pictogram shown on the middle of Figure 5 is as follows: AG1 (n=9), AG2 (n=12), AG3 (n=10), AG4 (n=9), AG5 (n=3). AG1 through AG4 respondents' interpretation as 'Surprise' for the pictogram accounts for more than 20% (AG1 (n=5), AG2 (n=4), AG3 (n=2), AG4 (n=3)), while the same interpretation was not found in AG5 respondents. AG2 through AG5 respondents' interpretation as 'Emphasis' for the pictogram accounts for more than 30% (AG2 (n=8), AG3 (n=3), AG4 (n=3), AG5 (n=1)), while the same interpretation is not shared by AG1 respondents. AG5 respondents reported a unique interpretation of 'Moral support' (n=2) that was not shared by the other age groups. AG1 respondents used the phrases 'Seriously' and 'OK' as usage examples for the pictogram, while AG5 respondents used the phrases 'Fight' and 'Go for it'.
- Pictograms showing face/facial expressions
  The number of answers for the pictogram shown on the right of Figure 5 is as follows: AG1 (n=27), AG2 (n=35), AG3 (n=16), AG4 (n=18), AG5 (n=11). AG1 and AG2's interpretation as 'Sorry' for the pictogram accounts for more than 20% (AG1 (n=6), AG2 (n=7)), while this interpretation is not shared by other age-groups. AG4 and AG5's interpretation as 'Patience' accounts for 20% (AG4 (n=4), AG5 (n=3)), while AG3 has less than 10% of account, no account for AG1 and AG2.

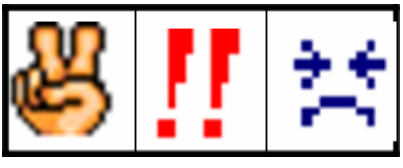


Fig. 5. Pictograms with a tendency of age-specific differences of interpretation

2.3.2 Pictograms with a Tendency of Gender-Specific Differences of Interpretation

This section describes pictograms with a tendency of gender-specific differences of interpretation across all the age groups. In the sections below, 'M' denotes male respondents, and 'F' denotes female respondents.

- Pictograms showing gestures
  The number of answers for the pictogram shown on the left of Figure 6 is as follows: M (n=49), F (n=48). Male's interpretation of the pictogram that account for more than 20% are 'Hitting' (M (n=10), F (n=0)), while female respondents' interpretation of 'Fighting spirit' accounts for more than 20% (M (n=0) F (n=10)).
- Pictograms showing emotions
  The number of answers for the pictogram shown on the middle of Figure 6 is as follows: M (n=49), F (n=48). Male's interpretation of 'Love' accounts for 59% with a difference of 27% from the accounts of female (M (n=29), F (n=16)).

Female respondents reported an interpretation of the same pictogram as 'Happy' (M (n=0) F (n=24)) or 'Fun' (M (n=0), F (n=4)). This finding suggests that men use this pictogram only for 'Love', while women often use it to express enjoyment in a general sense without a specified referent.

- Pictograms showing face/facial expressions
  The number of answers for the pictogram shown on the right of Figure 6 is as follows: M (n=19), F (n=33). Male respondents reported an interpretation of 'Failure' (n=5). This interpretation is not shared by female respondents, who reported an interpretation of 'Problem' (n=5).

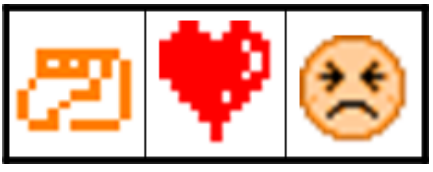


Fig. 6. Pictograms with a tendency of gender-specific differences of interpretation

2.3.3 Pictograms with a Tendency of Common Interpretations among Age and Gender

Figure 7 shows the pictograms for which almost no age-specific or gender-specific differences of interpretation were found. The number of answers for the pictogram shown on the top-left of Figure 7 is as follows: AG1 (n=16), AG2 (n=25), AG3 (n=10), AG4 (n=5), AG5 (n=6), M (n=19), F (n=33). All the age groups and both genders have the common interpretation of the pictogram as 'Anger' that accounts for more than 50% (AG1 (n=8), AG2 (n=15), AG3 (n=7), AG4 (n=3), AG5 (n=3), M (n=17), F (n=17)).

The number of answers for the pictogram shown at the bottom-left of Figure 7 is as follows: AG1 (n=27), AG2 (n=35), AG3 (n=16), AG4 (n=18), AG5 (n=11), M (n=49), F (n=48). All the age groups and both genders have the common interpretation of the pictogram as 'Rain' that accounts for more than 50% (AG1 (n=15), AG2 (n=24), AG3 (n=8), AG4 (n=16), AG5 (n=9), M (n=24), F (n=43)).

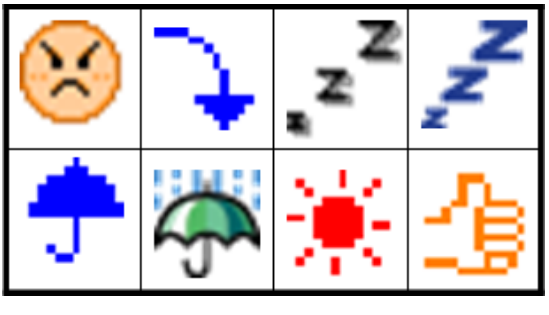


Fig. 7. Pictograms with a tendency of common interpretations

Other common interpretations of the bottom-left pictogram are 'Umbrella' and 'Feeling gloomy'. The interpretations these pictograms had in common across all the age groups and genders suggest they are unlikely to cause misunderstandings due to differences of interpretation.

3 Development of Age-Specific Pictogram Dictionary Application

This section provides an overview of the prototype of an age-specific pictogram dictionary application, and describes its functions.

3.1 Age-Specific Pictogram Dictionary Application

Our age-specific pictogram dictionary is a web-based application that lets users enter a mobile phone carrier, gender and/or age group (AG1 to AG5) as search conditions (multiple conditions can be entered) to search and view pictogram interpretations and usage examples. It was developed to help prevent miscommunications caused by different interpretations of pictograms.

Although we collected more than 6,600 questionnaire data to create a table of pictogram interpretations to achieve this objective, we still need more interpretations from wider variety of participants in order to fully investigate differences in pictogram interpretations by age groups. We developed the dictionary application as a prototype.

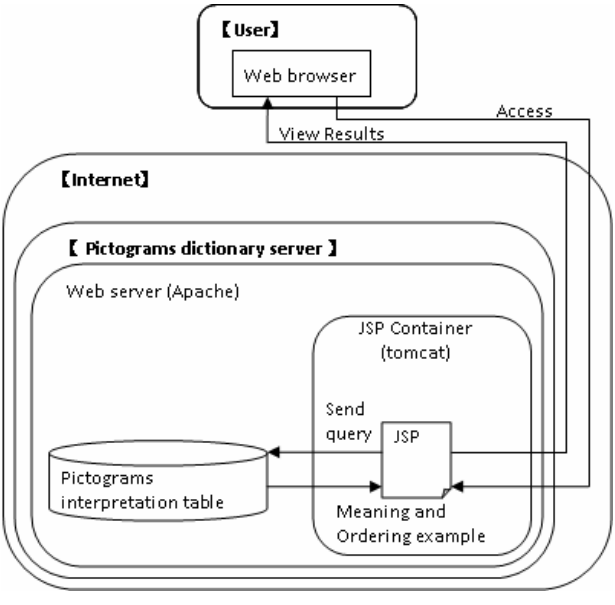


Fig. 8. System configuration of age-specific pictogram dictionary application

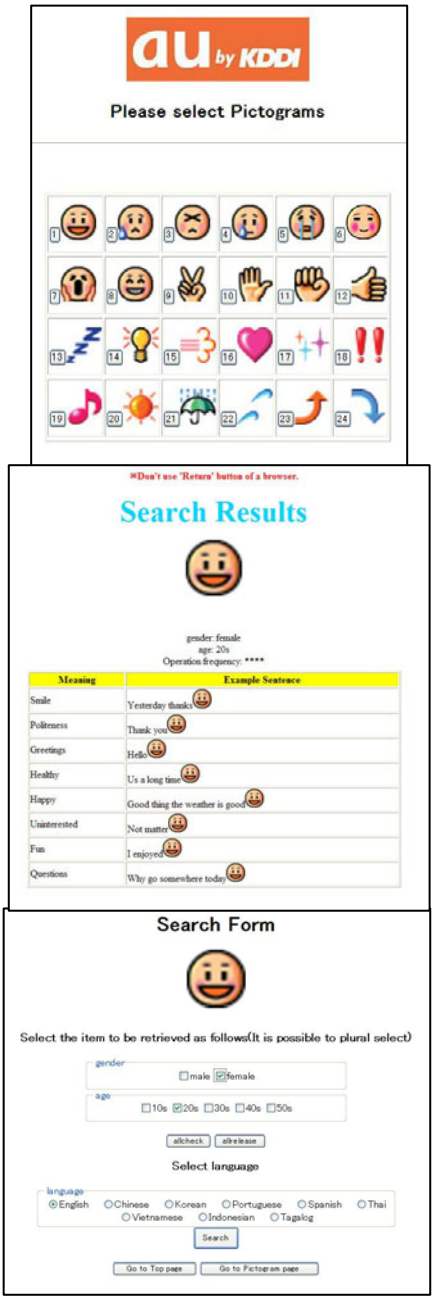


Fig. 9. Flow of operation screens of age-specific pictogram dictionary application

3.2 Functions of Age-Specific Pictogram Dictionary Application

The functions of each screen in the age-specific pictogram dictionary application are described below. Figure 9 shows the flow of operation screens.

- Carrier selection screen
  Used to select the carrier (au, DoCoMo or Softbank).
- Pictogram selection screen (top of Figure 9)
  Used to select any of 24 pictograms. Positioning the mouse pointer over a pictogram displays the most common interpretations given to it by each age group and gender, on the right side of the screen.
- Search screen (center of Figure 9)
  Used to select an age group and/or gender as search conditions, and display the results matching the entered conditions. Multiple search conditions can be selected.
- Search results display screen (bottom of Figure 9)
  Displays the search results (pictogram interpretations, usage examples, and usage frequencies).

The system configuration of the age-specific pictogram dictionary application is as follows (refer to Figure 8): The user starts by accessing a website written by JSP (JavaServer Pages), to select their mobile phone carrier, the pictogram to query, and the target age group and gender. The pictogram interpretation table within the system receives the query, and outputs the result (the interpretations and usage examples for the pictogram).

4 Development of Multilingual Translation Service for Age-Specific Pictogram Dictionary

We developed a service that uses Language Grid machine translation services to provide multilingual translations of the age-specific pictogram dictionary application described in the previous section.

4.1 Objective

Japan now has nearly two million foreign residents with a fairly high (84%) rate of mobile phone ownership [6]. Since pictograms have culture-specific differences of interpretation as described in Section 1, we suspect that foreign residents communicating with Japanese by means of mobile phone pictograms may have difficulty understanding the intended interpretations. The multilingual translation service we developed for the age-specific pictogram dictionary described in the previous section translates the interpretations and usage examples into several languages, for presentation to foreign residents. Its aim is to help expedite and facilitate communication on mobile phones across different languages and cultures.

4.2 System Overview

Using the Japanese mobile phone pictogram interpretations and usage examples as a language resource, we developed a web service that translates them into several

languages by combining machine translation services offered by Language Grid. The intended users are foreign residents of Japan from the ten most commonly represented countries—North and South Korea, China, Brazil, the Philippines, Peru, USA, Thailand, Vietnam, Indonesia, and the UK [7]. (Nine languages were actually used due to repetition of the primary language.)

The system configuration of the multilingual translation service is as follows (refer to Figure 10): The user starts by accessing a website written by JSP, to select their mobile phone carrier, the pictogram to query, the target age group/gender, and the language to translate into. The pictogram interpretation table and standard-language correspondence table (described later) receive the query, and output the result (the

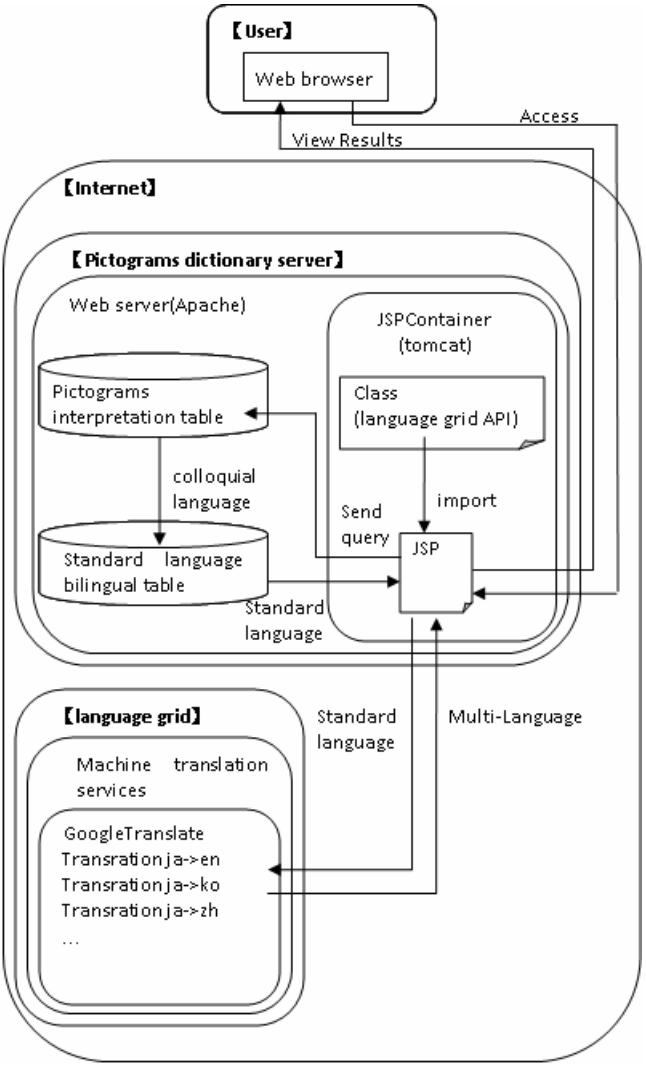


Fig. 10. System configuration of multilingual translation service

interpretations and usage examples for the pictogram, written in Japanese). Language Grid machine translation services then translate this Japanese result, and the translation is output to the user as the search result. The service uses Google Translate provided by Language Grid as its machine translation engine.

4.3 Standard-Language Correspondence Table

The interpretations and usage examples of pictograms given in the pictogram interpretation table are colloquial forms peculiar to mobile phone usage, so attempting to machine-translate them without preprocessing generates incorrect results. We therefore created a standard-language correspondence table to convert these colloquial forms to machine translatable standard-language forms. The standard-language correspondence table has two fields, one for the colloquial form of the entry and one for its machine-translatable standard-language form. It converts entries from colloquial to standard-language forms by linking to the pictogram interpretation table and calling queries.

When selecting the terms to enter in the standard-language correspondence table, we sought to improve translation precision by using a back-translation service provided by Language Grid to check that the entered text matched the back-translated result, before storing the data. Of the 6,624 pictogram interpretations/usage examples we compiled in Japanese from the questionnaires described in Section 2, we linked the original colloquial forms to standard-language versions for a total of 1,337 items. The remaining 5,287 items yielded correct results when machine-translated without modification. Table 1 shows examples of colloquial forms and their standard-language equivalents, and the compares the machine translation results for each.

Table 1. Example colloquial and standard-language forms, and machine-translation results of each

Colloquial language (in Japanese)	Standard language (in Japanese)	Translation results in a colloquial language	Translation results in a standard language
Ohaaa	Ohayo	--- us	Morning
(wara)	(warai)	(Laughs)	Laughter
Kono TV omoshiroi	Kono TVbangumi wa omoshiroi	Interesting in this TV	This TV show interesting
Suteki	Suteki (in Kanji)	Suteki	Nice
Asobo	Asobou.	Fun	Let's play.

5 Feedback and Comments

This section presents feedback and comments we received from users who tested the prototype of the age-specific pictogram dictionary application and multilingual translation service.

Eight Japanese users tested our age-specific pictogram dictionary application. It received positive feedback such as “It should come in handy, I’d like to use it.” (AG2 man) and “It will show my children and husband how to use pictograms.” (AG4 woman). One of questions raised was “Isn’t it difficult to judge which interpretation the writer intends from among the many choices available in the search results?” (AG5 man). In response to this question, the pop-up window that displays the most common interpretation of a pictogram (see 3.2) is useful. However, we believe the application should list an excessive number of meanings when needed so that users can learn from large number of interpretations and usage examples, and enable them to see age-specific and gender-specific differences of interpretations and relevant usage examples. Another comment we received was “I’d like the pictogram dictionary to be accessible from mobile phones too.” (AG5 man).

We received positive feedback from three foreign users who tested the multilingual translation service. One comment was “I’ve always used emoticons instead of pictograms when sending email on mobile phones to Japanese people. From now on I want to use this service so I can send lots of pictograms too.” (AG2 Korean man). One of the requests we received was “I’d like the translation service to support my language.” (AG2 Laotian man). Laotian is not supported by the translation services provided by Language Grid, so will have to await a new translation service.

The above comments and feedback indicated features and research issues to be improved and added in the future research. Firstly on the application, two features could be added to improve our age-specific pictogram dictionary application and multilingual translation service: (1) a service that lets users view pictogram interpretations and usage examples from their mobile phone, and (2) a pictogram interpretation/usage example email response service that returns a text email containing age-specific interpretations and usage examples for the pictograms it receives in query emails sent from user mobile phones to the pictogram dictionary server. To make the search interface offer more practical benefit, it could also be improved by adding a reverse search function that outputs pictograms in response to user input of pictogram interpretations. Emoticons could also be added to the pictogram interpretations and usage examples, by creating a correspondence table linking Japanese mobile phone pictograms to the emoticons used in Japan and abroad. This feature would make it easier for Japanese to understand the interpretations of emoticons used by foreigners, and give foreigners a better understanding of the interpretations of Japanese mobile phone pictograms.

Secondly, processes for collecting data and detecting age-specific or gender-specific interpretations need improvements and bring up several research issues. We used paper questionnaire, analyzed the data by hand, and developed a prototype application. In future research, web survey is more useful for collecting interpretations and usage examples for more pictograms from more participants. Developing algorithms for automatic detection of age-specific or gender-specific interpretations is also needed by utilizing services on Language Grid, such as morphological analyzers and parallel translation services.

6 Conclusion

This research developed a prototype of an age-specific pictogram dictionary application and multilingual translation service to reduce miscommunications caused

by age-specific and gender-specific differences of interpretations in mobile phone pictograms.

We analyzed 72 pictograms using questionnaire data compiled from the responses of Japanese mobile phone users in five age groups ranging from teenagers to those 50 or older. We found that thirteen pictograms depicting faces/facial expressions, gestures, and emotions have tendencies of age-specific or gender-specific differences of interpretation.

We plan to continue gathering data periodically to expand the pictogram dictionary with additional everyday interpretations and usage examples for mobile phone pictograms. Improving the features of our age-specific pictogram dictionary application and multilingual translation service will expedite and facilitate communication by means of mobile phone pictograms, between users of different age groups and cultures.

Acknowledgements. This research was conducted as a research activity in Special Interest Group on Graphical Communication in the Language Grid Association. We thank to the Language Grid Project at National Institute of Information and Communications Technology (NICT) for this research to come true by providing frameworks and web services for machine translations.

References

1. MyVoice.com: Frequently Used Functions in Japanese Cell phones, http://www.myvoice.co.jp/biz/surveys/11601/index.htm (acquired in 2009/12) (in Japanese)
2. Eshita, M.: Deep structure f networked society, Chuko Shinsy (2000) (in Japanese)
3. Wallace, P.: The psychology of the internet. Cambridge University Press, Cambridge (1999)
4. Cho, H., Inaba, E., Ishida, T., Takasaki, T., Mori, Y.: Semantics in Pictogram Communication. IPSJ SIG Notes 2006(110), 1–8 (2006) (in Japanese)
5. Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration. In: IEEE/IPSJ Symposium on Applications and the Internet (SAINT 2006), pp. 96–100 (2006)
6. Kushchu, I.: Foreigners acceptance of mobile phones in Japan, http://www.moba-ken.jp/wp-content/pdf/final_Ibrahim_w.pdf (acquired in 2009/12) (in Japanese)
7. United Nations, Demographic Yearbook system, Demographic Yearbook Special Census Topics, http://unstats.un.org/ (acquired in 2009/12)

Cultural Computing – Creative Power Integrating Culture, Unconsciousness and Software

Naoko Tosa

Kyoto University,
Yoshida Nihonmatsu-cho, Sakyo-ku, Kyoto, 606-8501, Japan
tosa@media.kyoto-u.ac.jp

Abstract. The author is carrying out technology studies to explore and expand human emotions, sensibility, and consciousness by making innovative use of artistic creativity. We develop interfaces for experiencing and expressing the "essence of culture" such as human feelings, ethnicity, and story. History has shown that human cultures have common and unique forms such as behavior and grammar. We suggest a computer model for that process and a method of interactive expression and experiencing cultural understanding using IT called "cultural computing". We particularly examine Japanese culture, although it is only a small subject of computing.

Keywords: Communication, Software, Information System, Cultural Computing, Interactive Art.

1 Introduction

“The ethnic crisis, the urban crisis, and the education crisis are interrelated. If viewed comprehensively all three can be seen as different facets of a larger crisis, a natural outgrowth of man’s having developed a new dimension – the cultural dimension – most of which is hidden from view. The question is, How long can man afford to consciously ignore his own dimension?” This phrase is from “The Hidden Dimension” by cultural anthropologist Edward T. Hall, 1966.

Nowadays, computers play an important roll in various ways in our life. Cellular phones, e-mails, websites, games as well as PCs are almost parts of our life, and they became the daily items or media. Computers were only ‘machines to calculate something’ at first, but now they are ‘media for thinking and memorization support.’

Let us see the relation between traditional customs and computers. Computers are typically used for calculation to restore something or for historical simulation. Archiving the fading cultures with using computers are barely in the use of thinking and memorization support, but it isn’t an effective application of the ability of computers which now treats multimedia and are connected in network. The present ages often communicate with someone who has another cultural background, so they are needed to understand the history of their culture and other cultures. Because the typical way to understand them is to read the books or to go to the museum, understanding another culture with picking appropriate information is not so easy.

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 127–136, 2010.

Can we understand the culture using computers as 'media for thinking and memorization support'which became more suitable for network, mobile, and two-way communication by the development of information technology? This paper describes basic methodology of cultural computing, that is, to treat the essences of the deep-inside culture like sensitivity, national traits, or narratives. And it integrates them into verbal and nonverbal information, proposing the prosperity of the field that treats the experience of exchange between cultural experiences and culture models by using computer. The cultural computing, which is essential for the communication ability of future computer, introduces you to this new field which defines what humans have stored in each culture and it's history in forms of actions or grammars are sharing common or peculiar forms by showing some concrete methodology and some examples.

2 Media for Thinking and Memorization Support

2.1 The Transcending Artist

What is the aim of arts? This is very huge and heavy question, but if I dare to answer, it is 'the visualization of the status of someone's heart' in a word.

The value of arts is in contrary of the one of technologies. Great art works are having universal values that have not faded yet. On the other hand, old technologies are often selected and surpassed by other new ones. Media arts does transcend them, making new relationship between art and technology, and affecting to other various fields. In other words, arts transcend technologies and technologies transcend arts.

I was conscious of my way as a transcending artist, and determined to obtain global viewpoint instead of local, Japanese culture as the next step. I moved my base of research activity to Massachusetts Institute of Technology, USA. However, I faced the difference between cultures there. I noticed that there were the differences not only in our daily life, but also in the feelings, memories, sign, unconscious communications that are strongly related to their cultures. I noticed that extremely Japanese-characterized expressions as well as global express had also been used in my works, although I had tried to express my artworks generally.

I have tried to assimilate myself into America at first, but it was difficult and I strongly realized that I am Japanese. I paid attention to actions and grammars that are seemed as 'Japanese culture' in America. I tried to expression the difference between American behaviors or grammars and Japanese ones in the arts.

At that time, I have met Sansui ink paintings (Japanese landscape paintings) of Sesshu. Sansui paintings are not landscape painting, but they are imagined scenery. This is in relation to the unconsciousness of occidental Jung's psychology. I discovered the way of unconscious communication to understand other cultures by using the components of unconsciousness in Sansui paintings.

I succeeded to create an interactive artwork, 《ZENetic Computer》 by modeling the structure of Zen and Sansui, which are thought the very Japanese culture. After that, I created the Kanji inspiration 《i.plot》 which gives relations between psychological associations and graphical images of ideograms. Furthermore, I

produced《Hitch Haiku》system which supports creating the Haiku from several Kanji (Chinese character) input, using the template of 5 – 7 – 5 characters.

Surprisingly, when I exhibited these works in overseas, for example MIT, many Americans understood these works. This shows that these arts, transcending the time as history or culture, and the space as nations, became the media which people in other culture can understand, by picking up the structure of peculiar traditional Japanese culture through the computer.

2.2 Cultural Poiesis

By continuing these studies, I discovered that I could precede the methods of art and technology. That is, interactive works with computer modeling of Japanese culture can be brand new media that enable everyone to understand other cultures by touching them. Thus, obstacles between Japanese traditional culture that was thought as peculiar and other cultures were slowly disappeared.

To express a core of a culture based on the traditional culture model, by using computer. Many people from around the world can create their own Haiku or Zen, and send them to the world by the computer using traditional cultural model through the metaphor. Of course, computers can model not only Japanese cultural metaphor but also the world theater of Shakespeare, and it is possible to create Kabuki composed by the metaphor of the Globe Theatre. I think this can be called as the creation of the culture, the cultural computing.

I noticed that computers have a feature appropriate to create the new culture. Computer processes are divided into algorithms and data. We can seem them as types and contents! Handling the culture by the computer would lead to the creation of the new culture. The culture in that context is a poiesis (that means creation in Greek) of communication between different cultures.

3 From Occidental Unconsciousness to Eastern Sansui Paintings

3.1 《Neuro-baby》 Connects *Here* and *There*

I am a postwar generation with plenty of American cultures for TVs available when I was born. The effect is great that unconsciously I had an interest in surrealism of adolescence and Jung's unconsciousness of psychoanalysis. After that, I began to express something invisible, like consciousness or feeling. That is why I started to create computer characters, computer characters exist "there." On the other hand in virtual reality, and human exist in this world "here" Sometimes through "there." We can communicate each other.

At that time, computers were in evolution from workstation to the personal computer. When I saw a personal computer at first, I knew by intuition that this item has functionality like human's one. I wanted to create a grown-up human, and make him/her talk through the feelings, by visualizing the internal consciousness as my theme of expression.

In the fields of AI, many researchers studied computers that can talk with humans as one field of technology. But their conversations are almost conventional, or a

verbal exchange which is programmed in advance. So we can't help feeling that we are forced to talk along some patterns. Conversations should be more fresh, free and enjoyable.

I created a work 《Neuro-baby》 based on these ideas. The baby computer character cries, laughs, or does some actions from the user's voice expression of talk.

When I announced it in the international conference, it became the center of the attention of researchers in the fields of AI and robots. I did not know why, but now I think my work hit a blind spot in the point that it had an aim to exchange the feelings by the conversation, instead of their aim, to exchange the information.

The exhibition titled 'Artificial Life' took place at Als Electronica in 1993. Studies and works with computers and robots aiming at simulation of Lifelike evolution or humanize by using the theorem of biology or neural networks. There, I met an strange foreigner who was watching the 《Neurobaby》 many times, and talking to it in funny tone. After a casual greeting, I knew he was Rodney Brooks, a worldwide authority of MIT in the fields of AI and robots.

AI and robot technology met with Interactive art. I felt that technologies were approaching to arts in a new way. I also met Thomas Ray who were studying artificial life at ATR in the position of biology. At that time, few people created interactive arts, and there're no concrete way in this field.

I met Yoichi Tokura, who was the director of ATR (Advanced Telecommunications Research Institute International) Human Information Communication Research Laboratories and studied a sound using baby at ATR, and got the position of a guest researcher in new laboratory, Media Integration & Communications Research Laboratories of ATR. And among many technological researchers, I had studied the mechanism of the communication of feeling from 1995 to 2001.

Is it the technology using the method of art? Or is it an art using the method of technology? Looking back upon it, I focused on the essential component which is inherent in the human communications, rather than their technologies. I think I spend more time to visualize in artistic method using the technology. Communication is an action deeply related in our instinct, and includes many interesting phenomena. First, I focused on handling nonverbal information like feeling started in 《Neurobaby》, and second I focused on handling the story.

A typical example of nonverbal information is feeling, and it contains much information widely spread from simple feelings like happiness, anger, sadness and comfort, sensitivity to the unconscious sense. So tackling these problems can make us enable the interactive visualization of feelings, sensitivity, and unconsciousness.

3.2 MIT CAVS

After I had quit ATR Media Integration and Communications Research Laboratories, I stayed Boston as the fellow of MIT Center for Advanced Visual Studies from 2002 to 2004.

The history of art & technology started from the avant-garde art group, EAT which was active in later half of 1960s. The main members of the group were engineers who engaged in Bell Telephone Laboratory of AT&T. Famous artists, Robert Rauschenberg, Robert Whitman and John Cage participated in, and carried out activities crossing the borders of arts, dances, music, videos, pursuit of the borderline

between art and technologies by setting the foothold to New York. EAT is the mother group of Center for Advanced Visual Studies (CAVS) established in MIT, 1967. Gyorgy Kepes, who had defected to America and became a professor of the department of architecture, was there. He invented large movements in various fields like urban life, environment and life, the fusion of art and science, as well as the architecture. Kepes founded the CAVS, and became as a director of it. It was 1968, when whole America had hoped for new arts. CAVS has the longest history as a laboratory of art & technology (and it is now at 3F above the main gallery of MIT Museum). It was an pioneer of performance and collaboration, and its effect to post generations is so large.

Artists like Namu June Paik, Charlotte Moorman who is cellist and collaborated with Paik, Scott Fisher who invented 'Head Mount Display', which created virtual reality (VR) technology are typical members of it.

I worked as a fellow from 2002 to 2004, invited by Prof. Steven Benton who was the 4th director of it. He was a recognized authority on a study of holography, and he invented the rainbow holography. He is also one of those founded MIT Media Laboratories. He took me in so that CAVS can change its motto from "art and science" to "art and computer", that means, from analog to digital. My expected role was to blow Asian wind of digital arts on CAVS, which had been deep-rooted in European culture.

3.3 We Cannot Take Our Culture Off

Have you had impatience about communication and thought, "Communications usually succeed, but why do they occasionally fail?" It fails because you communicate with someone supposing that they should be able to understand you. Communications are needed, however, for they cannot understand you. If you get it, you will find how to communicate with them calmly.

If we have understood all of us already, we do not need to communicate with each other. Japanese people has a strong pride of homogeneous race, so we apt to have relatively the same feelings, impressions and opinions about someone's actions, phenomena and a course of our society. As they typically call it a 'direct communication from mind to mind (Ishin-denshin in Japanese)', we are prone to think something supposing that they should understand us.

The characteristic of Japanese becomes obvious when we go abroad. That's because we face the condition that we cannot make ourselves understood even if we have thought they should be able to understand. Many other Japanese looks hard to be normal, like herbivorous animal living with carnivorous animals.

In America, however, especially Anglo-Americans often communicate with considering that a person cannot understand them, I think. Actually, I had colleagues from Germany, Greece, Lebanon, Japanese, Chinese, French and Anglo-America. Southern Europeans usually use a nonverbal communication like Japanese, which makes me feel a sense of closeness. However, Anglo-Americans keep a distance when they communicate. Somehow, they do not open their hearts, or they tend to hesitate at showing their feelings. This fact shows that they communicate with considering that a person cannot understand them.

The same situation sometimes occurs when we communicate with a computer as well as a person.

In Boston, audio response systems often answer when we call to enter into contracts with a telephone or a gas service. Computers ask us, “My name is Alice (e.g.), please answer me to register.” “May I ask your name, please?” “May I ask your address?” Those questions continue. And what is worse, the speech recognition sometimes fail. Then the computer says, “I couldn’t recognize, please repeat that again?” Answering it three times will be the limits of our patience. One of what are good about American frontier spirit is this big-hearted and trendy disposition.

“Do people here really patient about computer operators?” I asked my friend and she advised me, “In that case, you should wait while human operator appears.” She says that various things (computers or humans) meddle in phenomena, and extended the time we decide something. This analysis also shows the difference between Japanese who accustomed to communicate tacitly and Americans who accustomed not to communicate tacitly.

While I stayed at Boston, I realized by my experience that we should think communications occur only when people can’t understand each other, considering the difference of their culture. Many times I experienced satisfaction of sympathy with someone whom I thought was not understandable, which is beyond expression. We cannot communicate freshly without this moment. If we communicate not only with discovering our errors, differences or sympathies but also with exchanging and amplifing our knowledge and feelings, the communication will transchend each cultures.

We obtained global comunication by adopting medica technologies in face-to-face communication which had been limited to a small community. E-mails, social networks, blogs enabled us to communicate more easily with people from around the world, beyond barriers of distance and culture. On the other hand, many people feel communications being more and more shallow these days. Rather, these shallow communications brought a recent typical face-to-face conversation, “Did you read my message?” Communications may be turning into extremely superficial communications, with taking off the tastes of cultures.

Ignoring this tendency will cause the decline in our communication ability which we have had as the basic instinct from ancient times. We immediately need the new communication media which can convey one’s depth of feeling crossing the border of cultures. I knew it is realizable during my 2 years stay in Boston.

I wanted to create the communication media with which we can communicate deep feelings transcending the culture. As the result of my stay in Boston, I had this strong idea.

3.4 Technologies Combined with the Spirit

I visited Western China for 10 days, in the later July 2002. The aim of this trip was to discuss with Tibetan doctors and philosophers, and to complete the fieldwork for my research theme that looks for the problems about arts, technologies and hearts.

First, I visited Xining, the capital of Qinghai province. Gelug who consider the religious precepts highly important prosper there, and the founder, Tsongkhapa, and 14th Dalai Lama are also from there.

Many tulkus, including 14th Dalai Lama, practiced asceticism in the Kumbum Monastery (Ta'er), one of the 6 biggest temples of Gelug. I inspected the temple in where 4,000 monks practiced asceticism, and felt great energy or a sense of many people's mind flowing inside the temple where many pagodas (stupas) standing.

Tibetans naturally have an idea that the medicine and the philosophy is the same. Doctors are philosophers, and are Buddhist priests at the same time. I was impressed by the nature that doctors see what is wrong in patient's heart at first.

Tibetan philosophy has an faith to give freedom to all the afflicted lives. Their thought is deeply related to consciousness, feelings, the space and lives, centering the bowels of mercy and wisdom. The cosmic view has to do with our essential problems, the wheel of life and the existence. Many cultures in Tibet include Buddhist Tantrism as the appreciation of the idea that life cycles.

South of Xining is Guide (貴徳), which is famous for its hot springs. A 3,800 meters highland is nearby it. We visited a Tibetan tent (yurt), and they offered us butter named 'tsampa', milk of yak and dishes using barleys. We entered Tongren (Regon in Tibetan) and crossed a huge dam with a ship.

We went back to Xining and visited Arura Tibetan Medical Center. We discussed with five people including famous Dr. Denchi, the hierach and Buddhist philosopher, and Dr. Tanjinja who is nyingma (specialists for sadhana in Tibetan cabala). When we referred our spirituality of arts and the possibility of fusion between art and technology, they identified with us and said, "It is a possible idea as one of the figure of future religion." This encouraged us so much.

Tibetan Buddhists set the sprit of bodhisattva very important. This means to put off their tenacity to themselves and self-love, to have altruistic love. I noticed that this spirit is deeply related into interactive art.

Interactivity in art is shallow and the value is low if the purpose is up to self-assertive or communication of feeling. What is important is the interaction with having the sprit of bodhisattva and altruistic love. That is, if computer systems succeed to interact with having the sprit of bodhisattva, the interactivity of the system can deeply resonate the high spirit with other people.

I knew in my visit to Tibet that the Buddhism, which was born in India, has stayed in Tibet adjusting to the climate there and that the global consciousness is remaining there. I was impressed with the fact that we Japanese and Tibetan Buddhist are able to understand each other at a deep side of our spirit. I hoped to create something that Westerners can also understand the Buddhism, as a media expression using technologies. I just met the Sesshu, which critically affected my artworks after that.

3.5 Meeting the Sansui

I met the Sesshu at the exhibition "Sesshu – special exhibition at 500th anniversary of his death" (Kyoto National Museum, 2002). I was fascinated with the world Sesshu had created. I did not have a special interest in Japanese culture before. For some reason, the Sansui world of Sesshu in that exhibition seemed a virtual reality which expressed his heart!

In old China, a Sansui picture was once about a landscape we wanted to watch forever, a place we wanted to go to play, a place we wanted to live, and a hometown of our heart in which we wanted to pass away. The Sansui picture is imagined scenery

like that. Its bleeding, cracked, feathering lines of ink brush draw the movement of the heart. It makes us feel the color even if it is monochrome.

I had an inspiration to compute the Sansui picture typically by Sesshu and the world of Zen which was expressed in Sansui picture. Zen makes us feel Japanese culture by its absence of absolutes, beauty sense of "Wabi-Sabi", and getting rid of the water from the Chinese garden with taking the Asian culture in. Many elements in Japanese culture are gathered in Sansui, like the Ume-Sansui (Sansui with Japanese aprlicot). I wanted to express the Sansui picture and Zen culture, centering the Japanese Zen.

Cultures consist of 'the God, the Buddhist image, a view of Life, a view of world'. They have been created, changed, opposed and fused with each other, and are irrational and rational. They, which have both irrationality and rationality, have seemed difficult to handle.

Almost all the existent media art works featured traditional contents like Noh and Kabuki, and added interactive function to them. These were seemed as only superficial digital expressions, though these are extensions of texts, images, videos, and their combinational multimedia that explain the traditional culture. They are only a explanations and not new arts. They are on the stage only to treat the surface, than to approach directly to the culture.

There was no research to make use of the hierarchy of Japanese culture to the computer logic in the existent computer technology. This is the reason why there are no arts expressing the deep historical 'culture' in large scale yet. Another reason is that everyone have paid attention to the uniqueness of Japanese culture and seemed the Japanese culture as Japanese superficial expressions.

On the other hand, technologies finds the mechanism of phenomena, and analyzes the elements of them with the structure. And the study finds new relationship between different things and constructs them, by reconstructing them, trying some combination of them and comparing them.

What we can make use of in creating new media arts are extracting the basic structure or thoughts of Japanese culture, modeling them or using them as tools with using the technology. Fresh media works or art works would likely be created by that. This method will bring about a great possibility to the advancement in media arts and interactive arts hereafter.

Thanks to the cooperation of Seigo Matsuoka, who is a researcher of Japanese Culture in Editorial Engineering Laboratory, we took a little advantage in this difficult challenge that we reconstruct the world of Zen which was expressed in Sansui picture on the computer. Though we needed three years, we reached to the unique system named《ZENetic Computer》 as the result.

We succeeded to construct the extremely futuristic interactive system by projecting a part of allegory or symbol in Sansui pictures, Yamato-e (Japanese traditional paintings), Haikus, Kimonos that reminds of the Japanese culture – the structure of the oriental thought, the structure of Buddhist philosophy and the mechanism of Japanese traditional culture – which rarely have featured by the computers before.

This system uses various symbols and allegories that are included in Buddhism, an oriental thought and the Japanese culture. This is because they include a plenty of implications, and they have extraordinary terms, figures or colors. There are many rules in Sansui pictures and the world of Zen. We discovered that computers can

handle them, if we can select and extract them. For example, there are 'San-En' which is an expression of Sansui pictures, and 'Go-Un' (five elements which form a self-existence) which is a function to recognize the human in Buddhism and so on.

The first exhibition of this system was in MIT museum. I wondered whether Westerners understand it or not. As a result, however, it was accepted by many Westerners and won the great popularity. Westerners had felt that Sansui pictures and Zen is extremely oriental and hard to approach, but they gave me impressions that they could understand them through the interaction with this system. I myself had an impression that we could achieved the initial goal to express Japanese culture in media, when I saw a American child interacting joyfully with this system.

After that, it was exhibited in SIGGRAPH, the international conference of CG, and Kodaiji, a Zen temple in Kyoto. Each exhibition won great popularity. This success of the experiment using this system made me certain that the 'cultural computing' which computes the culture is reasonable to set to my research goal.

4 Structure of the Culture Becomes a Communication Technology

4.1 An Interaction to Reach the Racial Memory

I was encouraged by the success of 《ZENetic Computer》, and felt that interaction that reaches the deep-inside racial memory was the research I wanted to realize as the next stage. From the dry interaction of computers to the friendly and impressive interaction. How to realize this challenge?

I tried to classify the types, structures and relationship of what supports racial memories in Japanese culture with my co-researcher, Seigo Matsuoka. Below is the detail:

1. Japanese natural climate
   Japanese transient weather and nature, thought of transience like 'Monono-Aware', beauty senses like 'Wabi-Sabi', existential thought that loves present situation.

2. Relationships between Japanese culture and Asian one (Japanese own method to take over the Asian culture)
   Transformation from Chinese Sansui pictures to Japanese ones, Chinese gardens and grove gardens to the Japanese Rock Garden.

3. The syncretization of Shinto and Buddhism
   The cultural structure that was reconstructed as a belief system, mixing the native faith and the Buddhism.

4. Characteristics of Japanese language
   Waka poem, Haiku poem, Noh thater, and the script of Kabuki. And as applications, Honka-Dori, Uta-Makura, Kakari-Musubi, etc.

5. Japanese Design
   Japanese designs are the most popular. Two-dimensional designs are Mon (armorital bearings), Ori (pattern of textiles), colors, paper patterns, lines for example. Three-dimensional dynamic designs are the design of Noh, Kabuki, etc.

On these bases, we can consult on various racial types of Japanese culture and the rear communication.

4.2 Computers Do Not have a Cultural Information Hierarchy

Scientific technologies have developed Web2.0 like Google, Youtube, Wikipedia and SNS by which we can send our information more easily. Robot technologies are also developing new basic techniques to realize the superior functions that living things have. More specifically, they are global communication technologies including the movement functionality, manipulation system, distributed autonomous system for upgrading the intelligence.

There is, however, no cultural information hierarchy that is needed to live with humans. Adding the local cultural information by cultural computing to here may contribute to create higher-level communication systems.

5 Conclusion

These methods of 'cultural computing' enables us to model and structure the deep-inside essentials of culture like sensitivity, intuition, racial characteristics and narratives that we have not able to quantification. I have set my goal to realize the communication that moves one's racial characteristic expanding the present computer's communication ability to have an ability to reflect the difference in feelings, consciousness and memories, based on the culture. If these systems are realized, social practical and cultural information expression systems through the languages, voices and movies will be realized in various fields.

References

1. Tosa, N.: Cultural Computing. NTT Publishing (2009) (in Japanese)
2. Stafford, B.M.: Visual Analogy: Consciousness as the Art of Connecting. The MIT Press, Cambridge (2001)
3. Tosa, N., Nakatsu, R.: Life-like Communication Agent -Emotion Sensing Character MI and Feeling Session Character MUSE. In: The Third IEEE International Conference on Multimedia Computing and Systems (ICMCS), pp. 12–19 (1996)
4. Minsky, M.: Society Of Mind. Simon & Schuster, New York (1988)
5. Hall, E.: The Hidden Dimension. Anchor, New York (1990)
6. Nakatsu, R.: Future of Communications: How can technology Contribute the Spiritual Aspect of our Communications. Ohm-Sha Publishing (2010) (in Japanese)
7. Tosa, N., Matsuoka, S.: ZENetic Computer: Exploring Japanese Culture. The Journal of the International Society for Arts, Sciences and Technology 39(3), 205–211 (2006)

Logos, Pathos, and Entertainment

Ryohei Nakatsu

Interactive & Digital Media Instutite, National University of Singapore
21 Heng Mui Keng Terrace, I-Cube Building Level 2,
Singapore 119613
idmdir@nus.edu.sg

Abstract. Various new entertainments using information and media technologies have emerged and have been accepted among people all over the world. Casual and heavy games, communications using mobile phones, blogs, and twitters are such kinds of new entertainments. It is important to discuss the basic characteristics of such entertainment and to which direction they would direct our societies. In this paper, a comparative study of entertainment in developing countries vs. developed countries, in old days and in nowadays will be carried out. Also future relationship between entertainment and our society will be described.

Keywords: Entertainment, new media, logos, pathos, Plato, The Tale of Genji, Illiad.

1 Introduction

New technologies such as network technologies, information technologies, and media technologies are rapidly changing our society; human relationship, our life style, our communications, and so on. Entertainment is one of the areas to which these new technologies are influencing strongly. One good example is games. Playing computer games and video games are one of the common daily activities of people especially for young generation. Also online-games are becoming very popular among young generation in US and Asian countries. Another good example is communications. People communicate with their families and friends using e-mails, mobile phone based mails, SNS, twitters, etc. Until 1980s communication media such as telephones were mainly used for business communications. However nowadays communications are something more that business conversations and became everyday entertainment.

There are lots of discussions on these phenomena but most of them are observing only what is happening in our society and reporting the fact that more and more people are spending time for these new entertainments. However unfortunately there are only a few considerations on the basic reason why these new entertainments have been accepted by people all over the world. Some of the fundamental questions are; Is this a totally new phenomenon in our long history or not? Which aspect of our society is changing based on the introduction of these new entertainments and how much. Also more difficult question to answer is that to which direction these entertainment is leading our society. In other words what would be the future of our society where it is expected that people would spend more time for enjoying entertainment than now.

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 137–146, 2010.

In this paper, the author will try to answer these questions starting from the question of what is entertainment. It will be clarified that in the long human history people have tried to clearly separate logical aspect of our behaviors and emotional aspect, in other words "Logos" and "Pathos." Consequently we have succeeded in hiding the great power of emotion only to our private aspect our life.

However, new media such as games and mobile phones have strongly affected this relationship between logos and pathos and have partly destroyed it. What is happening nowadays is that people show emotional aspect of their behaviors even in formal situations. More distinctive fact is that this tendency is more obvious in Western countries.

Finally based on these considerations it will be anticipated to which direction our society is going forward. Also it will be discussed what we should do to keep that basic identity as human being.

2 Positioning of Entertainment

2.1 What is Entertainment?

It is well known that there are more than several billon people facing starvation even now all over the world. For these people it is crucial to obtain food to survive for tomorrow or even today. In the global level conferences/meetings such as World Economic Forum [1] how to cope with this problem has been one of the most serious topics.

On the other hand in developed countries new types of entertainment have emerged and are emerging such as chat on mobile phones, games on game machines and PCs, etc. People tend to spend more and more time by enjoying these entertainments. There have been big concerns and complaints against such a trend. The basic logic of such complains is that compared with other human activities such as learning, trading, industrial production and so on, entertainment is not productive. In other words it would be complained that entertainment is only a waste of time.

But is this true? Then why there are already established huge entertainment industries including movie industries, game industries, sports industries, and so on. And why there are huge needs and markets for such entertainment. Probably we should be careful when we discuss this topic.

2.2 Origin of Entertainment

What is happening now is not a totally new phenomenon. In the old era human life was simple. Human carried out firming or hunting to survive. And when they are not occupied with these works, they entertained themselves by various means. In other words we could say that food is strongly related to the physical sustainability of ourselves. At the same time we could point out that entertainment is related to the mental/spiritual sustainability of ourselves.

Then the era of civilization came. Human introduced various types of activities such as art creation, business activities, learning/teaching, religious activities, etc. Because of these activities entertainment became considered as a secondary activity in our life. Although entertainment retained a certain part of our everyday life, it has not been considered as essential part of our life (Fig. 1).

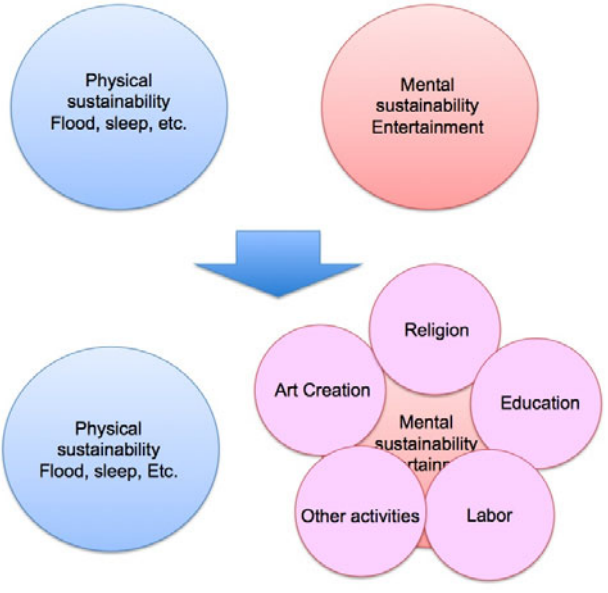


Fig. 1. Physical sustainability and mental sustainability of human being and their changes

2.3 Our Life and Entertainment

Even nowadays sometimes that fact that entertainment is an important part of our life exposes itself. Suppose that people are flying on an airplane. During the flight what people would do is to sleep, to eat, or to entertain themselves by watching movies, reading novels, etc. Only a few people do their business work during flight. Thins means that in a simple situation, our life style consists of three basic activities such as sleeping, eating, and entertaining ourselves.

What is happening now is in one sense "renaissance of entertainment." The introduction of new technologies, especially interaction technologies into traditional forms of entertainment has totally reformed and strengthened it. People re-noticed the basic strength and meaning of entertainment recognizing that entertainment is substantial part of their life. This is a key point when we try to understand such substantial issues as the role of entertainment in developed and developing countries, the future direction of entertainment, etc.

3 Entertainment and Developing / Developed Countries

3.1 Entertainment in Developing Countries

The importance of entertainment in developing countries has been underestimated. As is described above, entertainment has been from its origin an essential part of our life. In once sense, its importance is almost next to survival. Even though people in developing countries are not using PCs, mobile phones, game machines, etc., they know the importance of entertainment intuitively. Here is the key point.

We have been wondering and struggling how to introduce civilization into developing countries. It was considered difficult to teach the importance of civilized activities such as education, religion, business, etc. to people facing starvation. However, what is happening in developed countries would make this simpler. By observing the phenomena in developed countries we could say that the merge of entertainment with other activities such as education, business, etc. is emerging. In other words, many of the civilized activities are becoming one form of entertainment. This is in one sense a fundamental change of business models for various human behaviors and activities. Also important thing is that these new business model works even without utilizing cutting edge technologies.

3.2 Renaissance of Entertainment

In other words the border between entertainment and other activities such as education, business, training, etc is going to disappear. This is the reason why the phenomena happening now could be stressed as "renaissance of entertainment."

This means that many of the civilized human activities contain essence of entertainment and could be enjoyed. It would be better for us to introduce such serious activities as various types of entertainment to people in developing countries. This would make it far easier for people in developing countries to accept civilized human activities such as education, business, etc., because these activities are now interpreted as one form of entertainment and thus are familiar with them.

On the other hand, we should carefully look at the new entertainment enthusiastically accepted by young generation in developed countries. The question is that whether these are actually new types of entertainment or not. When we look into the details of these new entertainments, we found that most of them have their origin in the old forms of entertainment. For example what we experience by playing Role Playing Games is almost same as what we experience by readying fantasies. Another good example is new type of communications. New type of communications such as chats or text mails using mobile phones, twitters are actually reshaped form of old types of everyday conversation among our family members and friends.

This means that what information technologies and media technologies have done is to reshape the traditional entertainment keeping the same basic concept. Because of this, although in the early days of video/PC games the game market expanded with surprisingly great speed, it looks that the game market now has reached to a certain saturated situation and various new entertainments are facing the problem of sustainability. Probably we should learn the core concept of entertainment and the directions game industries should go by watching how people in developing countries entertain themselves.

4 Logos, Pathos and Entertainment

4.1 Logos and Pathos

What is happening in developed and civilized countries is actually deep and substantial. By the emergence of new entertainment our society and life style is experiencing a fundamental change.

Let's look back to the old Greek era in which we find the origin of long history of western philosophy. Plato, one of the most famous philosophers not only in Greek era but also in the whole human history, compared human spirit to a carriage with two horses and one driver in his famous book "the Republic" [2]. Here the driver is a metaphor of a rational aspect of human spirit, in other word "Logos." On the other hands two horses are a metaphor of a emotional aspect, in other word "Pathos" (Fig. 2). The former could be linked to the formal part of our life and the latter being the private part. And one of the horse means passion in emotion and the other is instinctive aspect of it.

Fig. 2. Logos and Pathos

By Plato logos was admired as the basis of human rational behaviors. For pathos, passion was admired as the source of human creative behaviors. But instinct was despised as undeveloped and dark side of human behaviors. Also he expected that logos would controls the dark side of pathos, instinct, with a help of bright side of pathos, passion.

This definition and statement by Plato determined the direction of philosophy and morality in Western world. Since the old Greek era, people have been trying to separate logos and pathos in their life. In other words, they have been trying to separate logical way and emotional way of living.

4.2 Formal and Informal Aspects of Our Life

Based on the philosophical considerations by Plato, in the long history of Western society, logos has been considered to be related to the formal aspect of our behaviors; behaviors in social situations, in business scenes, etc. On the other hand, pathos has been related to the private aspect of our activities. In the long history of Western world, people have been trying eagerly to clearly separate these two aspects. In addition to this people have tried to emphasize the importance of logos neglecting another aspect of pathos.

Also in Asian word we have such distinction. For example in Japan the formal aspect of our behaviors is called "Honne" and informal aspect as "Tatemae." Japanese people have been accused of having Honne and Tatemae by being said that Japanese people have double principle. However, based on the above observation this is not correct. All the human have formal aspect and private aspect, in other words

Honne and Tatemae. The problem is that as in Asian societies people have not been well conscious of these two aspects, they have not been good at clearly separating these two behaviors. In other words what is to be accused for Japanese people is the confusion of these two behaviors.

One good example is that former prime minster, Mr. Asoh. Once he declared at the parliament that actually he was against the privatization of Post Office when he was a member of Koizumi cabinet. The problem with his behavior is that as he was a member of the cabinet and this formal decision was made within the cabinet, it is not appropriate to express his private opinion at a formal situation. What was worse was that he thought he was honest by expressing his private opinion and could not understand why people accused him of his expressing Honne.

4.3 Approach between Logos and Pathos

However the introduction of new media and entertainment has invaded deeply into our life and has changed our behaviors. It was long considered as civilized and sophisticated behaviors to act logically hiding emotional aspect of behaviors. However, now people tend to expose their emotional behavior even in the formal part of their life. A good example is the way of communications. Nowadays it is common that people communicate with their intimate ones such as friends or family members using mobile phones or PCs even during meetings and dinners. Surprising thing is that communication with our intimates have been a typical private and emotional behaviors.

This means that in our everyday life again emotional behaviors became influential and play a major role, after long years of separation between logos and pathos and priority of logos toward pathos in formal situations. In one sense our behaviors are going back to those in the old era. Also this phenomenon has another important aspect. When observing this phenomenon we notice that there are two fundamental and distinctive features. One is that human behaviors in Western and Eastern world are approaching. Another is that human behaviors in developed countries and developing countries are approaching. In other words, this means that the difference of human behavior styles is disappearing between Western and Asian countries, developed and developing countries. People share the same thinking ways, principle, rules together. In one sense this is good as this is to lower the wall among different people and countries. But at the same time we are losing the local features of cultures we used to preserve in our long history.

5 Media and Entertainment

5.1 History of Media

It is important to consider why we are losing the separation between formal aspect of our behaviors and informal aspect. As is indicated above the invasion of new media have a strong influence on this.

To explain this let's look back our history of great inventions in media history. The two most impactful great inventions in media history is the invention of written characters and the invention of printing typography. These great invention made

people to think, memorize, discuss, describe, etc. using languages. In other words these inventions made people literal brain dependent. It is noteworthy to say that this mainly happed in Western countries. And based on this the long history of Western philosophy represented by Plato occurred. What happened in Asian counties is somewhat different. Somehow people did not try to clearly separate rational actions and emotional actions. The reason why this happened is an interesting research topic and will be described in other occasions. But anyway for Asian people the concept of separation between logos and pathos has not been so clear.

5.2 Influence of Movies and Telephones

Another two most impactful inventions in the recent history of media are telephones and movies. Nowadays we tend to focus only to recent inventions such as video games, mobile phones, e-mails, blogs, twitters, and so on. Unfortunately we almost forgot the big impact these two media have had on the basic change of our behaviors.

Before the invention of telephone, the formal aspect and emotional aspect of our behaviors was clearly separated. For emotional behaviors, mental distance and special distance were closely related. When we are together with our familiar people such as family members and friends we expose our private/emotional behaviors. We have a strong instinct to be connected with our familiar people. But before the invention of telephone when we were spatially separated from them, as we did not have methods to communicate with them, we had to hide the emotional aspect of us and behaved formally like in business scenes. But after telephone was invented, telephone made it possible for us to be connected to our familiar people even though spatially separated (Fig. 2). And gradually people tend to mix the formal and emotional aspect of our life. This is the fundamental reason why nowadays even during meetings, dinners, etc. people want to communicate with their familiars using mobile phones and smart phones. In other words the telephones has initiated the confusion between rationale and emotional aspects of our behaviors. The role of mobile phones and other recent media is only to accelerate this trend.

The invention of movies has the similar effect. Before the invention of movies people were trained as literal brain dependent and reading and writing were the major intellectual and communication behaviors of people. But the invention of movies introduced “images” as important communication medium. Images have a strong power to influence directly to the emotional part of our brain, or right brain. Therefore after long years of training to rationally use left brain, people began to depend on the usage of right brain. This means that people gradually became emotionally dependent instead of rationally dependent. The recent trend of the excessive use of computer graphics and animations have their origin to the invention of movie. What computer graphics and animation are doing is only to accelerate the trends.

5.3 Future of Entertainment and Media

As is described again after long years of separation between logos and pathos in human behaviors, these two aspects are approaching together and it could be expected that in the future it would merge again (Fig. 3).

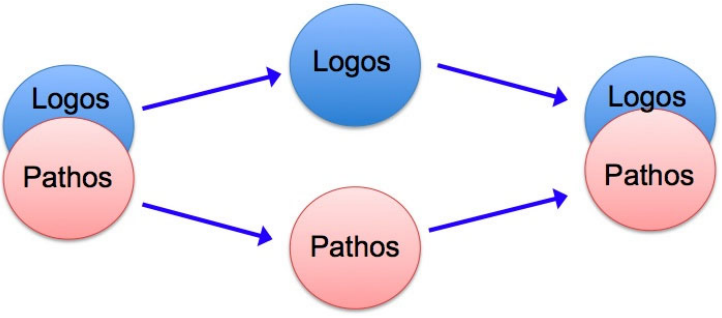


Fig. 3. Transition of the relationship between logos and pathos

What does this mean? Probably there are two aspects. One is that the behaviors in Western and Asians would approach and become identical. As is described above somehow in Asian countries the clear separation of logos and pathos has not consciously been aware nor discussed. Because of this Asian people have been accused as underdeveloped as human beings. But what is happening in Western countries is that their behaviors are approaching toward the behaviors of Asian people [3]. Good examples are their communication behaviors. Nowadays even during such formal evens as classes, meetings, discussion, dinners, etc. people tend often check e-mails on their mobile phones and try to communicate with their familiars. When a plane lands and arrives at a gate, the first thing people would do is to take out their mobile phones, turn on the switch, and check e-mails or start calling their family members. There is no distinctive difference between the behaviors of Westerners and Asians.

5.4 Future of Human Being

In one sense this is good as various gaps between the West and Asia have been long time headache for us. Probably we can call this a bright side of globalization. But on the other side there is a danger for this phenomenon. Each country and each race has preserved its own culture. Because of this there is a rich variety of cultures all over the world. This is one of the things what human being can proud of. But the merge of behaviors between West and Asia would mean that this would destroy this rich variety of cultural differences among different countries [4]. It would be a bad dream if we anticipate our future where there is no cultural difference and the people all over the world would be connected to the network all the time and repeat the receiving/sending of shallow text messages. This might be the bad dream shown in the movie Matrix.

There is another interpretation for the merge of logos and pathos. In the case of animals there is no such distinction as logos and pathos. These two are tightly merged together in the case of animals. This means that the merge of logos and pathos might mean that human being would go back to the state of animals. Therefore another bad dream is that we would go back to our very ancient time when human was not well developed and their behaviors were almost those of other animals.

Probably what we have to do is to learn from our history. Both in Western world and Asian world there were eras when logos and pathos were not so clearly separated and still people used to live the life of human. For example in the era of Iliad [5] described by Homer logos and pathos were not clearly separated but the people behaved honestly, bravely, and heroically. We should compare the behaviors of heroes/heroines in the classic with those of present people and learn from them on what our behaviors should be in the future.

On the other hand, in Japan we have The Tale of Genji [6], the oldest novel all over the world. In once sense the behaviors of heroes/heroines in this novel are somewhat similar to those of present day young generation. In the classic novel men and women frequently exchanged short poem called Waka. This was a method of communications for them. This is very similar to the behaviors of young generation nowadays. They frequently exchange short messages using mobile phones, twitters, etc. Then what is the difference? Probably the biggest difference is that exchange of poems is art creation but we could not find any art creation activities in the message exchange of e-mails and twitters. The big question is that how we could ennoble the communication behaviors of network age people. Our good dream would to realize the way of life described in The Tale of Genji in network with new media.

6 Conclusion

Various types of new entertainment represented by games, mobile phones, etc. are being introduced into our society, our life style rapidly. It seems we are in the era of chaos and it is not certain what kind of life style and society we would have in the future. In such a situation it is important to observe the phenomena happening around us carefully and extract some kind of findings, rules, tendencies, etc.

In this paper by trying to answer the question of what is entertainment the author tried to clarify the basic trends that exist behind various surface phenomena. At first tt was clarified that human behaviors consists of two aspects, logical and emotional aspects and entertainment is closely related to the emotional aspect.

Then the observation by Plato was referred. He tried to separate logical aspect of human behaviors and emotional aspect of human behaviors. His definition originated the long history of Western philosophy. As entertainment was closely connected to the emotional aspect of human behaviors, it has been considered as informal aspect of our behaviors and thus has been considered to be hidden, although entertainment is an essential aspect of our life as well as eating and sleeping.

Also it was pointed out that the recent development of information and media technologies made it possible for us to notice the importance of emotional aspect of our life. In that sense what is happening is the renaissance of entertainment. But on the other hand this means that logos and pathos that have been clearly separated in our long history are going to approach and even merge together.

Also it was pointed out that there are several dangers for such a trend. One is that the cultural differences that made the history of human being very rich might fade away. Another is that the behaviors of human will go back to those of animals.

So far there is no clear solutions to these dangers. However finally it was mentioned that one way to solve this difficult situation is to look back and learn from

our old history. In our history we had several eras where logos and pathos were together like in the era of Illiad and The Tale of Genji. Probably by learning from the behaviors in these famous classics we could imagine the future we are directing.

References

1. http://www.weforum.org/en/index.htm
2. Ptato :The Rebublic. Penguin Classics
3. Tosa, N.: Cultural Computing. NTT Publishing Co. (2009)
4. Nakatsu, R.: Future of Communications Empowered by Technologies. Ohm-sha publishing Co. (2010)
5. Homer: Ilias. Hackett Publishing Company, Inc, Indianapolis (1997)
6. Murasaki, S.: The Tale of Genji. A Divition of Random House, Inc. (1976)

Digital Conservation of Cultural Assets

Yoshihiro Okada[1] and Tetsuo Shoji[2]

[1] Graduate School of Science and Technology, Ryukoku University
1-5 Yokotani, Seta Oe-Cho, Otsu, Shiga, Japan
okada@rins.ryukoku.ac.jp
[2] Faculty of Sociology, Nara University
1500 Misasagi-cho, Nara, Japan
g_rcnl@daibutsu.nara-u.ac.jp

Abstract. Research goal of archiving is not simply digitalization of assets ,but archiving based on clarification of the original structures of the assets. For example, the information on old and deteriorated documents may be difficult to decipher. We aims to make such material widely available after using advanced analytic techniques to revive script and illustrations, thus restoring the document to its original form. First Cases: The Nishi Hongwanji holds a large collection of paintings and artifacts. To preserve the cultural property we are investigating preservation, conservation and conservation documentation methods. Second Cases: we report the virtual reconstruction and conservation of a lost cave shrine. The purpose of this research is to search for the possibility of the digital restoration that maintained the resolution for the realistic representation and the interactive media for digital archives of ruins in the future.

Keywords: Digital Archives, Photo Realistic Representation, Mixed-Reality, Image-Based Rendering (IBR), Model-Based Rendering (MBR).

1 Introduction

Nowadays, large numbers of cultural properties are found that have deteriorated through exposure to sunlight and dust over a long period of time, and that require conservation, repair or restoration. Recent years have seen active use of digital archiving technology to preserve or restore paintings, artifacts and other cultural properties that contain highly valuable historical information. In the study presented here, First Cases: The Nishi Hongwanji holds a large collection from the Momoyama era, paintings and artifacts, as well as the Momoyama style architecture. To preserve the cultural property we are investigating preservation, conservation and conservation documentation methods.

In the Nishi Hongwanji digital archive project, pigments data and measuring data from digital images are gathered and the database is developed.

In this paper, pigments on ‘shouhekiga’, pictures on (room) partition, or ‘ranma’, Japanese traditional transom are analyzed, and 3D computer graphics is constructed using the analysis data of pigments and gold foil.

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 147–160, 2010.

Second Cases: we describe a digital conservation method for Bezeklik Cave No.4 in Turfan. We digitized wall paintings of Cave No.4 scattered all over the world. Then, we integrated these paintings, made up the lack parts of paintings by using similar paintings, and corrected the color tone of paintings. Finally, we generated 3D-model of Cave No.4 with restored wall paintings.

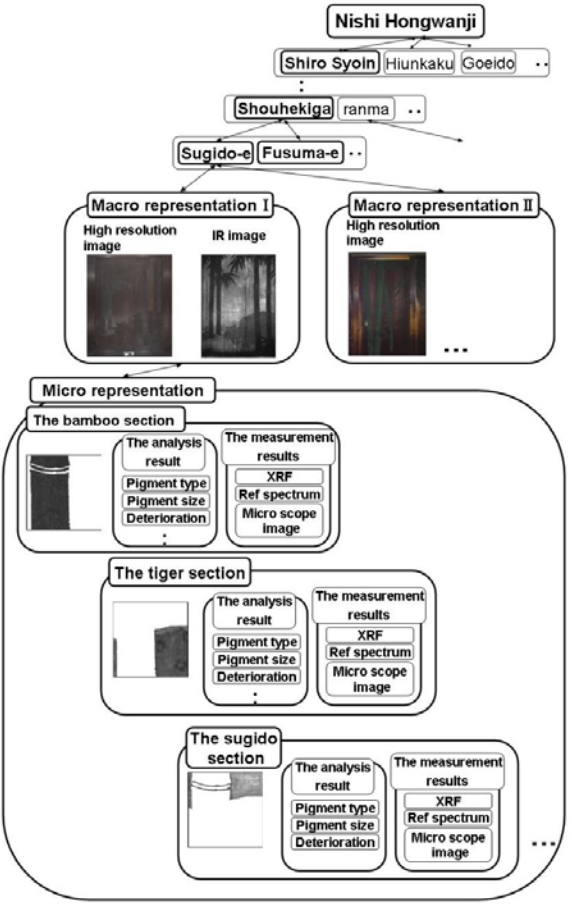


Fig. 1. Nishi Hongwanji Digital Archives structure

2 Nishi Hongwanji Digital Archives

Conventionally, digital archives of paintings store such data as the artist's name, year of production, pigments used and reflection spectra. Nishi Hongwanji Digital Archives store similar data on *shouhekiga*, though a different arrangement of data is required. This is because Nishi Hongwanji's *shouhekiga,* which are still in use as part

of the temple buildings, are in varying degrees of discoloration, deterioration, pigment exfoliation and so on, depending on their location, building structure, exposure to the elements and other factors. For example, all the paintings on the northern side of Tora-no-ma have blackened, whereas those on the southern side of the same room retain their original colors. In consideration of this fact, therefore, instead of archiving each *shouhekiga* simply as one picture, a multi-scale structure—as illustrated Fig.1.—has been developed for Nishi Hongwanji Digital Archives, to take into account the factors affecting the conditions of the respective paintings.

2.1 Analysis of Pigments

Shouhekiga and *ranma* are painted with pigments called *suihi enogu* and *iwa enogu*. *Suihi enogu* is made from mud or earth found mostly in mountains, which is washed with water to remove impurities and then dried in plate form. *Iwa enogu*, granular pigments made of various crushed minerals and semi-precious stones, are used with *nikawa* (glue) as a binder. They come in only a few basic colors, but this single raw material can yield a variety of hues, depending on the granule diameter. This characteristic is used to achieve various nuances and a three-dimensional effect. Granule diameters are indicated by numbers: the higher the number, the finer the granule. The finest granule is called *byaku* (see Fig.2. and Table 1).

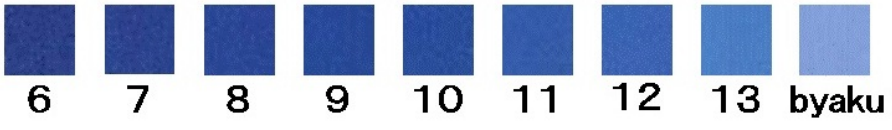


Fig. 2. Iwa enogu numbers and gradations

Table 1. Table 1 *Iwa enogu* numbers and granule diameters

Number	6	7	8	9	10	11	12	13
Av. diam(μm)	120	100	70	50	30	15	10	7

- X-ray fluorescence (XRF) analysis
- Granular diameters
- Reflection spectra

The results of these measurements revealed that the pigment *rokusho* (green) No. 8 is used in the bamboo sections of well-preserved *shouhekiga*. At the same time, the results led to the assumption that the same pigment is used in the bamboo sections of deteriorated paintings and that the pigment *oudo* (yellow ocher) is used in the tiger sections.

2.2 Reflection Models

Using the color-restored *shouhekiga* images, we attempted to create a representation of *shouhekiga* that also expresses the characteristics of the pigments used. As stated above, pigments have different granule diameters, which, as they become larger, render the surface of a painting coarser, due to the diminishing intensity of reflected light. We tried to recreate this effect using a surface scattering model that takes pigment characteristics into account. In the study we used the microfacet distribution function (Beckman distribution function, Formula (1).

$$D(x,k_2,k_1)=\frac{1}{S^2\cos^4 d}\exp\left(-\left(\frac{\tan d}{S}\right)^2\right) \quad (1)$$

$$[d=n\cdot h]$$

here x = denotes a sample point, k_1 = the incident vector

k_2 = the reflecting vector, n = the normal vector

h = the half-angle vector, S = the coarseness parameter

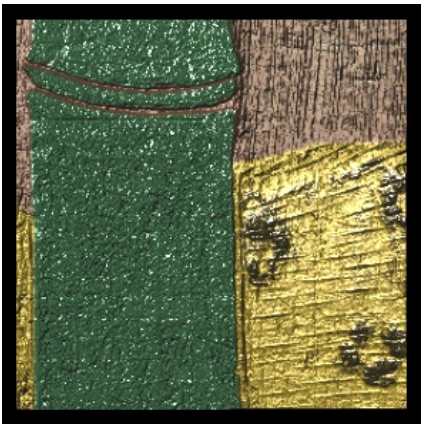


Fig. 3. Representation of *shouhekiga* using pigment reflection model

2.3 Gold Foil Reflection Model

Since Nishi Hongwanji has large numbers of gilded art objects, realistic representation of the surfaces of such objects constitutes an important aspect of the digital archiving. Gold foil, thinly spread gold obtained by striking the metal wrapped in deerskin or a similar material, is cut into squares for application. A gilded surface is minutely uneven and reflects light diffusely. As well, gold foil being a metal, it is

known to cause anisotropic reflection. In this study, therefore, we combined the Ashikhmin[1] and Beckman models to form a gold foil reflection model that reproduces images of gilded complex surfaces. The Ashikhmin model can be obtained from the sum of specular reflection $\rho_s(k_1,k_2)$ and diffuse reflection $\rho_d(k_1,k_2)$, which can be obtained from formulas (2) and (3), respectively:

$$\rho_s(k_1,k_2)=\frac{\sqrt{(n_u+1)(n_v+1)}}{8\pi}\frac{(n\cdot h)^{(n_u(hu)^2+n_v(hv)^2)}}{(h\cdot k)\max((n\cdot k_1),(n\cdot k_2))}F((k\cdot h)) \tag{2}$$

$$\rho_d(k_1,k_2)=\frac{28R_d}{23\pi}(1-R_s)\left(1-\left(1-\frac{(n\cdot k_1)}{2}\right)^5\right)\left(1-\left(1-\frac{(n\cdot k_2)}{2}\right)^5\right) \tag{3}$$

where u , v = orthogonal vectors to the normal direction

R_d = denoting the intensity of diffuse reflection

R_S = denoting the intensity of specular reflection

$F((k\cdot h))$ = Fresnel coefficient

Surface coarseness can be expressed by applying Formula (1) to the highlighted part of Formula (2).

Since the minute surfaces of *shouhekiga* and *ranma* are uneven, the mutual interference of microfacets must also be taken into consideration. This can be expressed by the following formulas:

$$G_{out}(n,k_2,h)=2\frac{(n,h)(n,k_2)}{(k_2,h)} \tag{4}$$

$$G_{in}(n,k_2,h,k_1)=2\frac{(n,h)(n,k_1)}{(k_2,h)} \tag{5}$$

$$G(n,k_2,h,k_1)=\min(G_{in},G_{out},1) \tag{6}$$

Ranma portion representations realized using the gold foil reflection model are also shown further below. Here, only a texture reproduced using the gold foil reflection model is applied. In the future, we intend to reproduce more realistic *ranma* surface images using the gold foil reflection and pigment reflection models.

3 3D Spatial Representation of Nishi Hongwanji

We produced a three-dimensional representation of Nishi Hongwanji's Tora-no-ma using a three-dimensional geometric model based on temple structure data. Model-Based Rendering (MBR), collective term for methods for synthesizing and expressing input data on structures, involves creating a three-dimensional geometric model in a

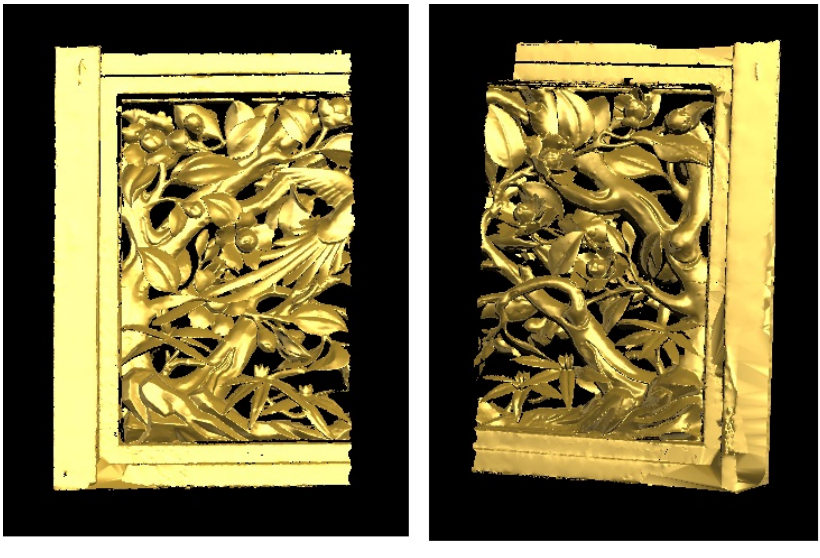
virtual world using input data on structures or a light reflection model or other physical models. Texture mapping, one MBR method, is particularly effective for realistic three-dimensional geometric model production.

3.1 Three-Dimensional Spatial Representation

In the study, we attempted a three-dimensional spatial representation of Nishi Hongwanji's Tora-no-ma in the following manner. First, we fixed the three-dimensional coordinates of the room on the basis of its structural data. We then conducted polygonal approximation with triangular patches, to generate a three-dimensional geometric model of a scene. We then texture-mapped the actual images onto the three-dimensional geometric model. For actual images, we used three types of image: digitalized high-resolution images, infrared photographic images and color-restored images.

3.2 Digital Color Restoration and Three-Dimensional Spatial Representation

We performed digital color restoration of Nishi Hongwanji's cultural properties using data stored in digital archives, such as pigment types, granule diameters, reflection spectra and infrared photographic images. Using color-restored images thus obtained, we generated digital representations of *shouhekiga* in Tora-no-ma using the reflection model, taking into consideration pigment characteristics and gilded surfaces using the gold foil reflection model. Finally, we created a three-dimensional spatial representation of Tora-no-ma, using *shouhekiga* images already in the digital archives and images color-restored in this study.



(a) Part of *ranma* (front) (b) Part of *ranma* (back)

Fig. 4. Representation of *ranma* using gold foil reflection model

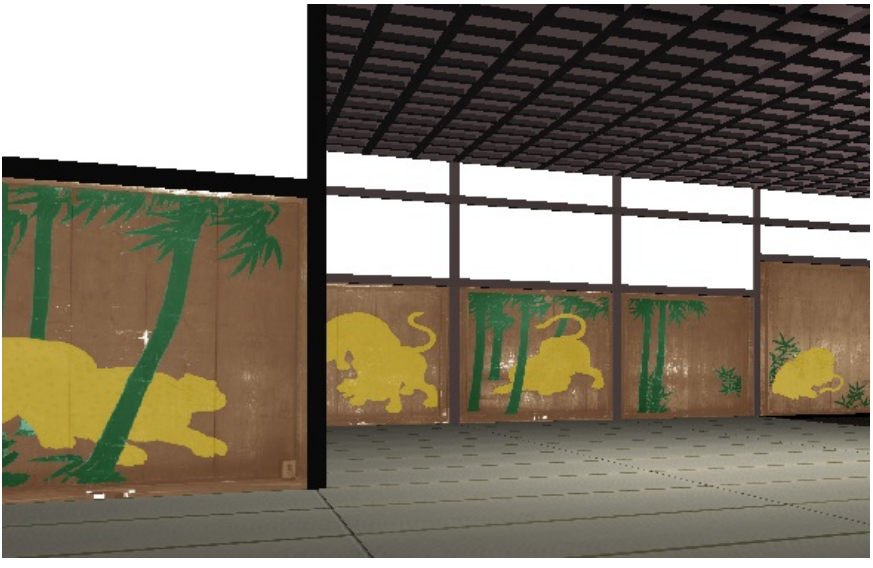


Fig. 5. Three-dimensional spatial representation of Tora-no-ma

Such is the scope of our study conducted thus far. In the future, we intend to go further by applying a reflection model that considers gold foil reflection to three-dimensional spatial representation, so as to include large numbers of Nishi Hongwanji's gilded art objects, such as *ranma* and *shouhekiga* in the archives. For this purpose, we intend to examine techniques for adding pigment and gold foil textures to images of objects having a shape similar to that of *ranma*.

4 Bezeklik Wall-Paintings

Fig. 6. shows Bezeklik temple shown in Turfan area in Dunhuang the northwest. Bezeklik means "Place with the decoration" in Uigur. This cave shrine was built from the 6th century to about the 14th century by Uigur, and there are about 83 cave shrines.

Several wall painting of the cave shrine were taken out by the expedition teams in the foreign countries, such that A. von Le Coq (Germany), Aurel Stein (England), SF. Oldenburg (Russia) and Otani expedition (Japan). Only few wall paintings remain in the original cave shrine. It is possible to see the fragments of the collected wall paintings in Chotscho[2] which published by A. von Le Coq, at National Museum of India (NewDelhi), at The State Hermitage Museum (St Petersburg) and at Tokyo National Museum. Fig.7. shows Bezeklik No.4 Cave. The ceiling of the cave shrine is

Fig. 6. Bezeklik temple

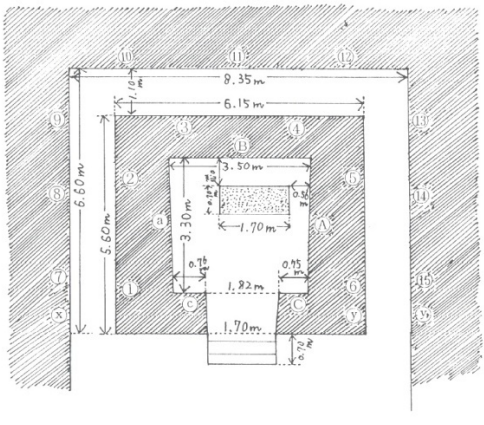


Fig. 7. Measurement of the Bezeklik No.4 Cave

composed of dome-shaped and the barrel type. The width of the corridor is about 1.2m and the height is about 3m.

4.1 Collection of Wall Painting Data and Database Building

Though a restoration object is the wall paintings of the cave No. 4, there are many lack parts. To fill the missed parts, we consider that it refers to the wall painting of other similar cave. Then, we use the wall painting of the cave No. 9 which similar to the No. 4. (A pictorial book for the No. 9 one was published.) The wall painting is classified based on the item shown in Table 2, the image-database is made. (Fig. 8.) For reconstruction of the wall painting, lack parts of wall painting are analogized religiously and statistically using the database.

Table 2. Classification attributes of wall painting

Attribute	Title
Cave number	Number of Cave which painted the wall painting
Title number	Title number for each wall painting
ID	Characters' identification numbers
File name	Image data file name
Source	Source of the wall painting data or publication (drawings, the actual locale, and museum etc.)
Character	Characters'name (bodhisattva, biku etc.)
Item	Item (flower, bread etc.)
Hair	Decoration applied to hair (hat or turban, etc.)
Cloth	Kind of cloth (kesa, armor etc.)
Ear	Shape of ear, ear item (long, earring etc.)
Face direction	Direction of face (left, front etc.)
Nimbus	Color of Hicasom (green, red etc.)

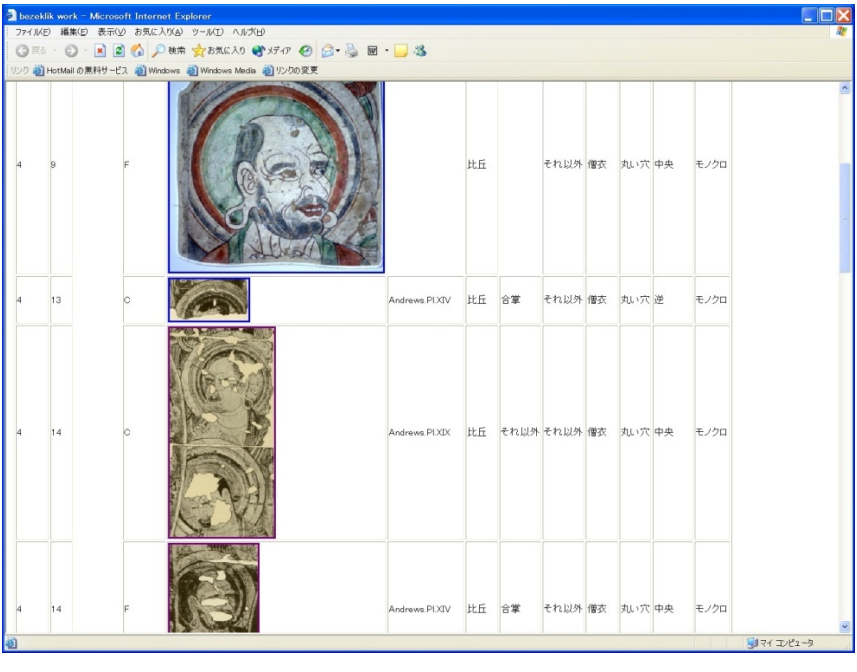


Fig. 8. Wall painting database

4.2 Analysis of The Wall Painting

The groundwork and pigments of the Bezeklik wall painting are given in Table. 3. Using i1 (made by Gretag Macbeth), we measured the spectrum characteristic of the groundwork and pigments for gypsum, lime, vermilion, yellow ocher, azurite, malachite, and the carbon obtained by marketing now. The color tone changes by roughness of pigments. If pigments fine the color brightness grew stronger.

Table 3. The groundwork and pigments of the Bezeklik wall painting [3]

Groundwork	Red	yellow	blue	green	black
gypsum, lime	vermilion, bengara	Yellow ocher	azurite	malachite	carbon

4.3 Reconstruction of The Wall Painting

For reconstruction of the wall painting, we consider a positional matching for fragments of the wall painting (drawing, photographic image in the cave and at museums) which are exist. (Fig.9.) Since the wall painting of the cave No. 4 is very similar to one of the cave No. 9, characters in wall painting are also similar. Although the wall painting of the cave No. 9 do not exist now, the facsimile, i.e. drawing, of it is exist. Using digital high vision broad casting, we consider a zoom of an arbitrary position on the reconstructed wall painting, thus a resolution is 7,000 x 7,000 pixels. A digital reconstruction procedure of the cave No. 4 is as follows.

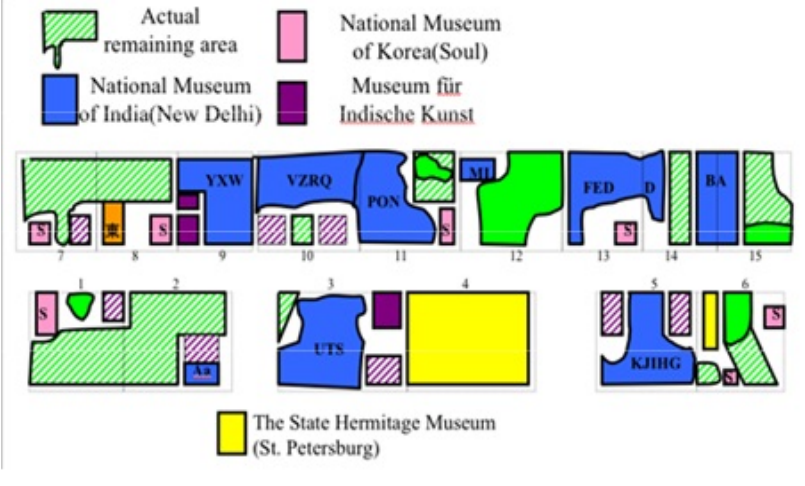


Fig. 9. Wall paintings of No.4 were divided by foreign expeditions

(a) Cave No. 9 (b) Positional match of wall painting fragments (c) Wall painting restoration result

Fig. 10. Reconstruction of wall painting

First, we carry out a positional matching for fragments of the wall painting (Fig.10.(b)) such that a location of characters of drawings for No. 9 cave wall paintings. (Fig. 10.(a)) Each fragment is in National Museum of India (NewDelhi), Museum für Indische Kunst (Berlin-Dahlem). Right and left fragments of Museum für Indische Kunst are not exist, but its drawings are exist.

Next, the lack part of the wall painting is filled using the image database of wall paints. Fragments are chosen religiously and statistically by shape of hands, shape of clothes, expression of face and so on. The fragments are done rotation, reduction, and expansion, and fill the lack space.

Finally, the tone of the entire wall painting is corrected. In this paper the appearance of the wall painting at the time of be drawn is reproduced. To achieve this, the spectrum characteristic of the fragment in existence wall painting is measured, it is necessary to investigate the level of deterioration while comparing it with pigments previously shown. However, it was difficult to measure the spectrum characteristic of the fragment of the wall painting. Then the basic color (Blue, white, green, red, flesh-colored, black, tea, and bitter orange) used by the wall painting based on the analysis result of pigments previously described was made. The tone is corrected based on these. Fig.10.(c) shows the result of correcting the tone.

5 3D- Models' Generation of The Cave No. 4

The purpose to generate three dimension model is to present to the user the position of the wall painting in a buddhist temple intuitively. There is a method of measuring actual cave with the range finder for the method of generating three dimension model of cave. But, by occasion of the above-mentioned, three dimension model is generated based on the data measured in drawing and several places open to the public without doing an accurate measurement here. The cave No. 4 is VR reconstructed by doing the wall painting restored to the generated three dimension model ahead in texture mapping. Fig.11.(a) shows the appearance of a present cave shrine, Fig.11.(b) shows the result of VR restoring of the cave shrine.

(a) Present (b) CG reproduction

Fig. 11. Three dimension reconstruction result

6 Expression of Wall Painting That Uses Virtual Source of Light

Here, we consider the method of expression of the wall painting previously restored. When cave shrine builds the worshipper is thought the wax candle to be light, and tries to reproduce this appearance virtual. That is, the user operates a virtual source of light, mode of expression into which optical how to see wall painting changes virtual.

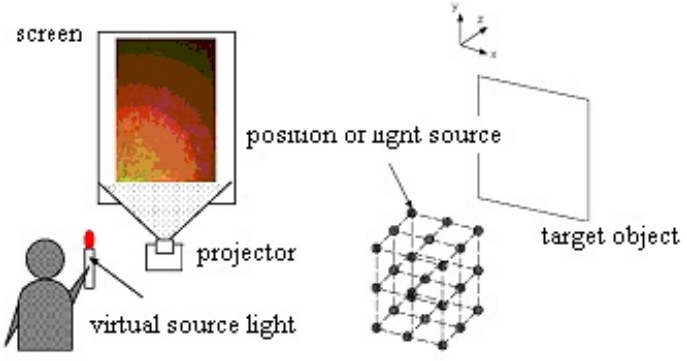
There are some relation researches, technique for projecting optical pattern to screen object, interaction concerning lighting and shadow was achieved by using absorbed screen (CABIN)[3]. Here, it pays attention to the reflection distribution of the source of light, neither a precise reflection model nor shape information in a real environment are used, the reflection distribution filter is generated from the reflection distribution image taken a picture. The reflection distribution filter is the one with the RGBA value (light source color and penetration level). A virtual effect of the lighting can be achieved by the alpha blending of restored wall painting and reflection distribution filter. When an actual object doesn't exist, the effect of the lighting can be achieved virtual by using this reflection distribution filter.

6.1 Generation of Reflection Distribution Filter

The reflection distribution filter is generated as follows.

1. The screen where the wall painting is projected to the position of the object body is set up. (Fig.12.(b)) The source of light uses the wax candle. Moreover, the moving range of the source of light assumes the cube as shown in Fig.12.(b). It takes an image of the screen that is a plane of projection, while changing the source of light position. (Fig.13.(a))
2. It regularizes it (takes to 0 from 1) based on the maximum value of the brightness of the obtained image. The penetration level (α value) is set based on the regularized image, and the reflection distribution filter is generated.

3. Using the generated reflection distribution filter (It has the RGBA value) and using the restored wall painting image (Fig.13.(b)), the image compositing is done by alpha blending. (Fig.13.(c))



(a) Outline (b) Taking a picture method

Fig. 12. Achievement of virtual source of light that uses reflection distribution filter

(a) (b) (c)

Fig. 13. Image compositing that uses reflection distribution filter

6.2 Interaction by The Virtual Source of Light

To change the optical image of the wall painting it is necessary to presume the position of a virtual source of light according to the position of the virtual source of light (i.e. penlight) that the user had. That is, the luminescence part of a virtual source of light is assumed to be a marker. It takes image of the marker with two cameras. The position of a virtual source of light is presumed by using the obtained stereo image. A synthetic image corresponding to the presumed position is displayed. Thus, when the position of taking a picture and the position of a virtual source of light are the same, a synthetic image is displayed as it is. When the position of a virtual source of light is different the weight putting is done to the image in neighborhood at the virtual source of light position and the image is generated. Fig.14. shows the result of doing the interaction by using a virtual source of light.

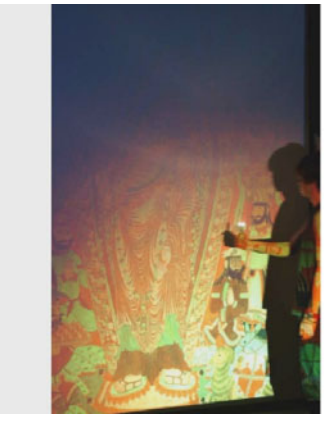


Fig. 14. Expression of wall painting that uses virtual source of light

7 Conclusion

In this paper presented here, we performed digital color restoration of Nishi Hongwanji's cultural properties using data stored in digital archives, such as pigment types, granule diameters, reflection spectra and multi-band photographic images. Using color-restored images thus obtained, we generated digital representations of *shouhekiga* in Tora-no-ma using the reflection model, taking into consideration pigment characteristics and gilded surfaces using the gold foil reflection model. Finally, we created a three-dimensional spatial representation. Second cases, we try to do VR reconstruction. For the restoration of the wall painting, the wall painting that lies scattered in all parts of the world is digitalized, a positional match is done, the lack area was analogized and the tone was corrected. Moreover, three dimension model is generated based on a open to the public drawing and local measurement data, and texture mapping did the wall painting restored to three dimension model. For expression of wall painting in addition, without assuming a complex reflection model, the effect of the lighting can be simply achieved by using the reflection distribution filter. Moreover, expressing the presence became possible through the interaction with the user. One wall painting image is projected in front of the user, and a virtual effect of the lighting has been achieved under the present situation.

References

1. Ashikhmin, M., Premoze, S., Shirley, P.: A Microfacet-based BRDF Generator. In: Proc. ACM SIGGRAPH, pp. 65–74 (2000)
2. von Le Coq, A.: Chotscho,Berlin,Tafel17-29 (1913)
3. Naemura, T., Nitta, T., Mimura, A., Harashima, H.: Virtual Shadows in Mixed Reality Environment Using Flashlight-like Devices. Trans. Virtual Reality Society of Japan 7(2), 227–237 (2002)

Automatic Knowledge Acquisition from Historical Document Archives: Historiographical Perspective

Katsumi Tanaka and Adam Jatowt

Graduate School of Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, 606-8501
Kyoto, Japan
{tanaka,adam}@dl.kuis.kyoto-u.ac.jp

Abstract. Recently many archives containing historical documents have been created and made open for public use. The availability of such large collections of past data provides opportunities for new kinds of knowledge extraction. In this paper we discuss the potential of web and news archives for automatic acquisition of historical knowledge. We also describe some aspects of the data and we draw parallel to historiography – the science of making the history.

Keywords: web archive, news archive, historical information, archive usage, text mining.

1 Introduction

Recently many historical document archives have been created. News and web archives are probably the most well-known ones. Although, web and news archives are sometimes regarded simply as loose collections of web pages or news articles, we limit our focus to their most common form that is the collection of historical documents. We define document archive as a repository of past documents or their copies containing the evidence of the state of the past frozen at particular moments in time. According to this view, a document archive is treated in this paper as the collection of documents which have been collected in the past, remained unchanged from their original form and contain metadata such as document timestamp. Page versions in web archives (or news articles in news archives) are either grouped thematically or according to their other characteristics such as language, location, domain name, etc. However, often the major underlying order is the chronological one.

Many technical and sociological issues are related to selecting archival material and enabling its efficient access and preservation. These topics have been frequently discussed by the web archiving and digital libraries communities [11]. In case of news archives the selection and preservation problems seem to be of lesser importance due to the much smaller data amount as well as the relatively long tradition of collecting and storing news articles by libraries or other institutions.

Nowadays, people leave more and more traces of their activity in a digital form. The digitalization is also commonly available making it possible to convert also traditional print documents to the digital form. In addition, the web encourages people to publish and interact with others, thus, leaving numerous historical traces that could

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 161–172, 2010.

be preserved for future generations. Therefore it has become possible to collect large amount of real-world data and convert it into document archives with unified access to all the individual artifacts. Such collections can be then easily accessed and analyzed using state-of-the-art computer technologies.

However the popularity of archives, especially of web archives, and their user awareness are still relatively low despite the availability of traditional access methods such as browsing or searching. This situation may raise questions of the necessity of archiving as well as may hinder the archival process. In this paper we argue that automatic knowledge acquisition from the archives is what could boost the usefulness of the archives and what could increase their value to the society. We discuss various issues related to the discovery of historical knowledge and possible applications. Our view is partially inspired by the historical methods and the notion of historiography[1] - the methodology of the discipline of history. Automatic temporal knowledge acquisition from historical document archives by using text mining applications can be useful for computational journalism [17], education, entertainment, verifying accuracy of existing historical descriptions and so on.

The remainder of this paper is structured as follows. The next section contains the description of document archive usage with emphasis on the various types of knowledge acquisition and their related issues. Section 3 provides deeper discussion of selected aspects of archived documents and the process of historical studies using document archives. The next section contains the related work. We conclude the paper in the last section.

2 Archive Usage

Despite their great potential historical document archives are still only moderately popular within narrow group of users, and few people seem to be aware of their existence and availability. Based on the online questionnaire study made in 2008 on 1000 users in Japan [8] we have found that less than 2% of web users have recently used any web archive. Partly this may be because of the lack of large online web archives in Japan. Nevertheless, this result implies rather low awareness of users about web archives and of the potential models of their usage. On the other hand, during the course of our studies we found that many people were often quite surprised to learn about the existence of repositories preserving large portions of historical web content. They also often expressed enthusiasm on hearing about the possibility of using such content.

News archives such as Google News Archive Search[2] seem to be relatively known and frequently used; though still, to rather limited extent when compared to the other popular web services. We believe that new usage models should be introduced in order to boost the usefulness and popularity of historical document archives.

2.1 Browsing and Searching

Browsing archive collection is one of the most fundamental ways of access. In the case of a web archive this access may be similar to traditional web browsing with

[1] http://en.wikipedia.org/wiki/historiography
[2] http://news.google.com/archivesearch

added time dimension. Browsing paths can be specified by the archive administrators or can simply reflect the previous link structure that has been recorded at the time of document archiving.

Wayback Machine is the best known application for accessing data in web archives using web-based interface [11]. It is used as the gateway to the Internet Archive's[3] web collection, which is the largest online repository of past page versions. Accessing past content of pages can be done by directly entering specially modified URL containing requested date. Another way is through the directory page listing available past versions. Users can access the content of versions and even follow their links as the links are rewritten to point to the corresponding versions within the archive's collection. The directory page indicates also page versions that contain changes when compared to their consecutive versions by marking them with asterisks.

Search is usually regarded as the process of retrieving particular required information or document, or for finding starting point for browsing activity. In case of temporal archives it often involves determination of the time constraints for the publication dates of documents to be retrieved. Most of online news archives enable temporal search over their collections for returning news articles published within requested time frames. Magazines such as Time[4], Newsweek[5] and plenty of less popular ones provide online searching facilities to their proprietary news articles' collections, which often have been accumulated over long time spans (e.g., 86 years for Time and 29 years for Newsweek). In addition, libraries throughout the world have recently begun digitizing their content and organizing into searchable and browseable digital collections. On the top of that, large web companies like Google, Yahoo![6] or Microsoft[7] started assembling data from multiple distributed news providers.

In the case of web archives, although many collections like the Internet Archive provide only URL-based access, some of them such as Portuguese Web Archive[8] and Pandora[9] already enable textual search.

2.2 Automatic Knowledge Extraction

Browsing and searching offer direct access to stored data. In many cases these are standard and the only way to utilize the archived content. However, manually locating and viewing particular documents in the archives can be tiresome for users. In addition, retrieved "micro-information" such as single web page versions or past news articles may not always be interesting and useful for users. On the other hand, the results of more comprehensive analysis of the larger size archive content could prove attractive. For example, users may not be interested in the details of particular event in the past but may want to know the frequency of similar events, their cause-effect relations, associated trends, differences from the present similar events and so on.

[3] http://www.archive.org
[4] http://www.time.com/time
[5] http://www.newsweek.com
[6] http://www.yahoo.com
[7] http://www.msnbc.msn.com/
[8] http://arquivo-web.fccn.pt/portuguese-web-archive-2?set_language=en
[9] http://pandora.nla.gov.au/

Individual historical facts may not be much meaningful to users, however, assembling or relating them with other evidence should make the returned information more useful and interesting.

Knowledge is basically the information about "relations". Document archives by grouping the artifacts and enabling their automatic access provide a powerful framework for knowledge acquisition in which topical and chronological order allows inferring higher-level relations (e.g., changes in ideas, evolution of ideas or topics, event co-occurrence and periodicity, etc.). For example, even if a given idea is known, its evolution over time and relation to other ideas provides new kind of knowledge.

In general, recent technological advancements in computing offer possibility of semi-automatic or automatic interaction with historical document collections and for producing interesting knowledge. Historical document archives should be then considered more as sources of topically and chronologically arranged data to be mined for useful knowledge. One way to infer knowledge is to use only the data in the collection itself while another is to measure relations (e.g., correlation) to other external data sources.

In reality, often, it is impossible to scan the entire collections for knowledge extraction in document archives, and the only allowable access way is through a search interface. This is usually due to their huge size, proprietary character or access restrictions. Effective mining applications should thus harness the provided searching facility as a means of knowledge acquisition. Already, effective ways for mining search engine indices through their search interfaces have been proposed in web mining area [4,5]. Bollegala et al. [4] measured inter-term similarity by analyzing web search results. Cilibrasi and Vitányi [5] proposed Google Normalized Distance based on web count values in order to use it for such tasks as hierarchical clustering, classification or language translation.

Figure 1 portrays a schematic way to sample historical document collection via its search interface, for example, to obtain longitudinal statistics such as the occurrence of particular feature over time. First, an application issues query to an online news archive. It is there transformed into a series of sub-queries spanning over a predefined time period $T=[t_{beg},t_{end}]$, each with a temporal constrain. The initial time period T is partitioned into R number of continuous and non-overlapping time units, which serve as temporal constraints for the sub-queries. The number of partitions (i.e. granularity)

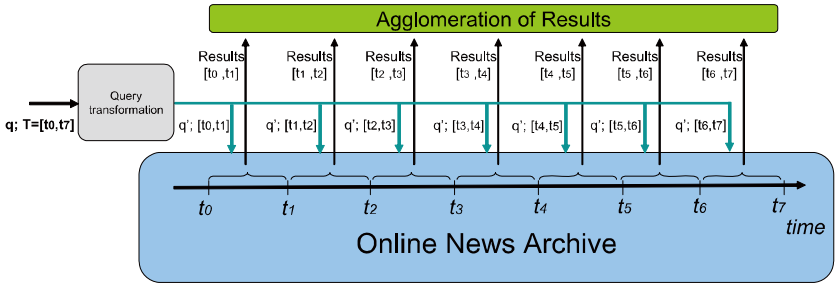


Fig. 1. Simple model of data collection for longitudinal knowledge acquisition from historical archives by querying search engine interfaces

can be set according to the query. This querying approach is made in order to reflect the actual distribution of relevant documents over time within the required time frame *T*. By dividing the query into many sub-queries we decrease the effect of temporal aspect used in the ranking algorithm of the particular collection, and thus, we manage to rely more on the actual document relevance.

In a crude distinction, the knowledge obtained from historical archives can be divided into two broad classes:

- knowledge about a particular source or group of sources and their changes and evolution
- knowledge about the past outlook of the world and the society as well as about the evolution of a particular topic or information over time

Below, we describe the both classes in more detail.

2.2.1 Knowledge on Sources

The first kind of knowledge relates to a given source or information container such as a web page or newspaper. Among the basic information are the frequency and changes in appearance of certain words or topics, the age of document components, the document change frequency or change degree.

For news archive this means characterizing a particular newspaper, magazine etc. through analyzing past editions and contributions. Such information may be useful for measuring the characteristics of news sources, identifying the relevant and high quality ones and so on. In case of web archives this kind of knowledge could add missing information for users browsing a current page version. For example, the users could learn about common topics that were discussed on the page recently or long time in the past. It would be then possible for them to contrast such topics with the ones published on the present page version. This could provide a context for better understanding of the current page version as well as the consistency, periodicity and other temporal characteristics of the page [8,9].

As another kind of knowledge users could receive the information on the age of certain components on pages in order to support the evaluation of their freshness and validity. This information would be obtained by comparing past page versions with the current one. For example, a page component annotated with “new” description may be discovered to be actually quite old as a result of the comparison of the current page version with the old page versions [9].

2.2.2 Knowledge on World and Society

The second kind of knowledge can be helpful for understanding the past as well as for learning about the present – e.g. trends, events, their origins and causes. There are myriads of potential kinds of such knowledge and the ways in which it could be utilized. In the simplest form, it can be extracted using summarization, filtering, association and other text mining technologies on the time series of features.

NY Times API[10] is an example of a programmable interface that offers an effective tool for news collection mining for such kind of knowledge. For a given time period one can find the names of objects mentioned in the news articles such as place or

[10]http://developer.nytimes.com/

person names and their related statistics. This kind of knowledge can have not only educational or entertaining purpose but can be used as an input to higher-level mining or reasoning processes.

In general, one of the prime objectives of the historical knowledge is to understand its connection to the present, find the ways to explain the present and support future prediction, for example, through discovering significant trends or analyzing periodical events. Figure 2 shows an example of successful detection of periodical events in the past from Google News Archive based on finding bursts in hitcount values using the query model depicted in Figure 1. It shows also the forecasting of the next event occurrences on the basis of the calculated periodicity [7].

Overall, by mining web or news content that appeared in the past one could reconstruct the collective image of the past, present and future that authors had in mind at that time. The changes in collective society images are particularly interesting to be studied. For example, one could analyze how the particular future predictions appeared, evolved or disappeared over time.

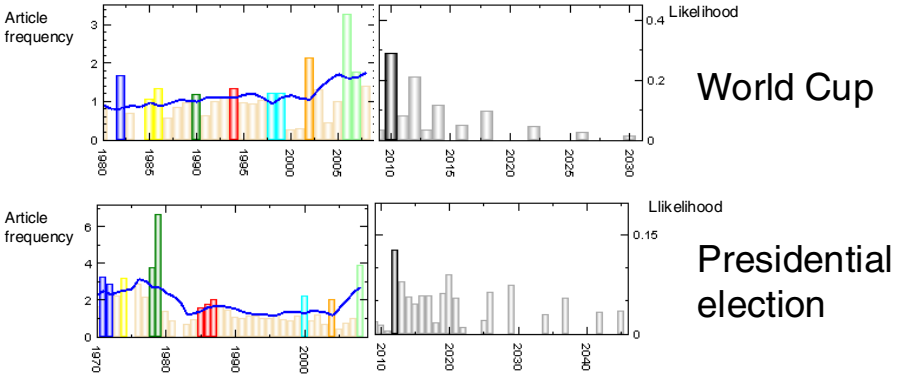


Fig. 2. Detection of periodical events for queries "world cup" and "presidential election" from a news archive and forecasting of their expected occurrences (indicated as the highest bars in the right-hand side charts) [7]

3 Historical Studies using Digital Archives

Digital archives offer possibility for individual users to quickly and easily obtain knowledge without the necessity of manually checking multiple historical resources. Using archives is thus possible not only to support historians in their work but enable history writing for average users.

Historical knowledge can be crudely divided from epistemological viewpoint into knowledge on "what happened" and "how happened". Naturally for the latter we need methods capable of high level reasoning, associating and so on, which may be still difficult to be realized with the state-of-the-art technologies. However, the second knowledge type can be already realized to successful degree by detecting events through automatic clustering or other approaches. By manipulating queries to be executed on historical collections one can also write thematic or contextualized

histories from different viewpoints, for example, the history of women in Japan or history of education in high schools in the last century and so on. These kinds of domain-oriented historical summaries have been usually done before only for selected topics that were of particular interest as their manual creation requires much effort and time.

In addition, historical archives can be used not only for writing historical summaries but also for evaluating the credibility of the already existing historical knowledge. According to the meta-history view, history is not credible and requires constant process of revision[11]. We believe that easy access to archives and the development of text mining and reasoning technologies will offer possibility for automatic verification of history in the future.

3.1 Primary and Secondary Sources

According to the historiography, historical evidences can be divided into three classes: primary, secondary and tertiary[12]. Suppose an event *e* occurred in the past at time *t*. Documents about *e* that were created around *t* are regarded as the primary sources on *e*, while the documents relevant to *e* but produced some time after *t* are considered to be secondary sources. Secondary sources concerning historical events are often created on the basis of primary resources.

The authors of secondary sources usually have more distant view on the events having access to more varying and complete information regarding the event when compared to the authors of primary resources. This is true as certain implications of the event as well as its context can be noticed and understood only some time after the event. On the other hand, there is a risk of missing some important details of the event due to the time passage or even distorting the view of the events, especially, if other secondary sources have been used in the document creation process. Historians generally believe that the closer to the event the more reliable the sources are.

In general, the web is not a self-preserving medium but a self-updating one. Due to constant pressure for new, up-to-date content, the stale fractions of the web are becoming neglected, less densely linked and in consequence less frequently visited, forcing their authors to keep the content up-to-date. This seems to be corresponding to the characteristic of our society according to which the great value is put on freshness and novelty, while the old fades and is cherished in only few cases (e.g., wines, antiques, historical buildings).

Web archive and news archives are examples of collections of primary sources regarding the time period when they were created. On the other hand, the current web can be viewed as a mixture of primary sources regarding the current time and secondary sources on the past, especially the distant past[13]. Figure 3 shows the conceptual view of primary-secondary distinction of web and web/news archives.

Historians usually face an issue of incomplete data when dealing with primary sources. Similarly, in the case of digital document archives, there is often lack of complete information on document evolution over time or some documents are

[11] http://en.wikipedia.org/wiki/Historical_revisionism

[12] In this paper, we treat tertiary sources simply as secondary sources.

[13] Naturally, online web and news archives or other online primary sources physically also belong to the web; however, for the sake of clarity we treat them separately here.

missing from the collection, etc. Therefore, exhaustive data accumulation and preparation steps are critical to the effectiveness of knowledge discovery processes. In case of web archives, two types of uncertainties can be distinguished here for a given series of past page versions. The first type, called *content uncertainty*, is caused by the lack of information about the transient content that appeared in the page within the time periods constrained by the time stamps of the consecutive past page versions. Consider two versions of a page, v_{left} and v_{right}, captured at time points t_{left} and t_{right} ($t_{left} < t_{right}$). The probability, $P(v_i)$, that there is some v_i satisfying $t_{left} < t_i < t_{right}$ and containing content different from that in v_{left} and v_{right} depends on many factors such as the length of the period $[t_{left}, t_{right}]$, the type of the page, the content difference between page versions v_{left} and v_{right}, etc. Basically, the longer the gaps between the page versions, the greater is the probability of transient content occurring in the page.

The second one, called *time uncertainty*, relates to estimating dates of detected content changes. In the above example, the exact timing of the content changes estimated from the comparison of v_{left} and v_{right} is unknown and can only be crudely approximated. The time uncertainty, like the content one, depends also on the number of acquired past page snapshots and their distribution in page history.

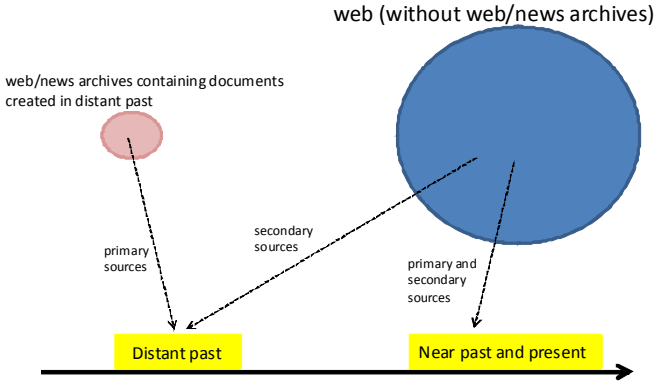


Fig. 3. Concept of primary and secondary sources in web and web/news archives

Detecting the differences between primary and secondary sources can also provide interesting insight. Given a popular object (e.g., company, person, place, etc.) one could compare the amount of attention in both primary and secondary resources about this object as well as the way in which the authors referred to it. For example, a scientist could publish a paper that was unrecognized by his peers at the time of the publication's appearance. Yet, sometime later the paper would be considered as highly influential. Table 1 portrays this concept. Also, sentimental attitude to given events may change over time. The differences in sentimental attitude could be detected by comparing the sentiment expressions used in primary and secondary sources (Table 2).

Table 1. Concept of event characterization according to its view in primary and secondary sources

	Event exists in primary data	**Event does not exist in primary data**
Event exists in secondary data	Established, well-remembered event	Event discovered later
Event does not exist in secondary data	Forgotten event	-

Table 2. Concept of event characterization according to its sentiment view in primary and secondary sources

	Event was recognized positively in the past	**Event was recognized negatively in the past**
Event is recognized positively now	Constant positive recognition of the event	Change: event become positively recognized
Event is recognized negatively now	Change: event become negatively recognized	Constant negative recognition of the event

3.2 Data Normalization

For the case of primary sources regarding the distant past (e.g., news articles from the 19th Century) there is a problem of the change in wider context such as language, culture or society rules. Generally, the further we move into the past, the harder is to understand the historical sources due to the gap brought by sociological and technological changes. For example, certain words may be no longer used or their meaning can differ from the one used currently. Hence, there is a need for a kind of "data normalization" or its "translation" so that the information extracted from distant primary sources could be understood and directly used for knowledge acquisition in combination with information obtained from documents created in more recent time periods. Although, the web has still relatively short history as compared to the history of print, nevertheless, is has existed in the times of rapid technological and cultural change. Therefore the same problems can be found in web archives here, yet, naturally, to lesser extent.

Somewhat similar data normalization is also often needed when we compare the results of statistics taken from different time points. For example, it is commonly known that more news articles appeared recently than in the distant past due to the rapid increase in the rate of journalistic activity. Therefore, the counts of documents created in unit time periods of near and distant past should be comparable only after their normalization with respect to the whole collection size in the both periods. A simple way of approximating the rate of article growth over time could be done measuring hitcount values obtained for series of stop words within the different time periods [7], provided such data is available.

3.3 Information Credibility

Both for the primary and secondary sources the credibility is of paramount importance. As historians have to effectively deal with forged or corrupted

documents, in the same way, users of online historical documents should be warned against incredible or incomplete sources. We distinguish here three types of credibility aspects from the viewpoint of automatic knowledge acquisition in historical document archives:

- credibility of document metadata
- credibility of document content
- credibility of collection archive

Credibility of metadata is about the trustworthiness of the document description. One common metadata problem of historical artifacts is the inability of determining the correct authorship. Many times the creator of an artifact is not known or only his pseudonym is revealed. Another common metadata issue is related to dating artifacts. Here the question to be asked is: "was the source actually created or published at the given time?" Documents which cannot be accurately positioned in time or whose timestamp is inaccurate have little value or can even harm the knowledge acquisition process. Time uncertainty of past page versions discussed above is related to the credibility of metadata of web page components.

The credibility of document relates to the question whether a given document is original and has not been altered in any form or whether all the previous alterations are explicitly known. In the previous section we briefly explained the related concept of content uncertainty in the history of a web page.

A very simple solution for automatically evaluating the credibility of documents and their metadata is to employ machine learning methods for outlier detection. For example, if a given document contains terms very different from the ones appearing in other documents created at the same time, then its credibility or at least the credibility of its creation time is questionable and should be manually examined.

The last credibility type relates to the possible bias in the collection construction. In the process of longitudinal knowledge extraction from historical archives, one often implicitly assumes that the collection reflects the popularity of information or frequency of published documents as it was actually in the past. However, if a given archive has been constructed in a way in which certain information or sources are over- or under-represented; then, using such archive may result in biased or inaccurate knowledge. The archive could be useful for the knowledge creation process only if one knew the scope of the bias introduced during the collection creation.

4 Related Work

Web archiving community has been recently actively involved in the issues of content selection, preservation and management. The overview of the current state-of-the-art as well as future directions in this area can be found in [11].

Particular cases in which web archives should be useful for users such as the ones in legal trials or topic-focused report writing were listed by the International Internet Preservation Consortium [6]. *Visual Knowledge Builder* [14] was an early proposal of an application for history navigation in private hypertexts. The authors' objective was to enable users to playback the history of hypertexts much like in VCR players. Users

could then witness the authoring styles of hypertexts and understand their various historical contexts.

From a social viewpoint, Wexelblat and Maes [16] demonstrated the *Footprints* system that adds social context to browsed document structures by utilizing historical data on user visits. In result, new users could be guided to useful and popular resources.

Ohshima et al. [12] proposed an approach for showing the changes in rivals or peers of user-defined objects over time based on data obtained from querying online news archives. In general, mining text streams has been studied relatively well (for example, see [15,10,1]).

Overall, until now there was relatively little research that explicitly aimed at mining content stored in web archives despite the fact that it presents a great potential for knowledge discovery. Apart from a few exceptions, most approaches neglected temporal dimension of page content. Aschenbrenner and Rauber [3] surveyed work that had been done towards mining large portions of web content with consideration of its temporal aspect. The authors provided also a general outlook on the potential of mining archived data. Rauber et al. [13] discussed the possibility of mining past web data for identifying and portraying changes in web-related technologies, particularly in such characteristics of pages as file format, language, size, etc. Arms et al. [2] have reported on building a research library for scientists to study the evolution of content and the structure of the web.

5 Conclusions

In this position paper we have discussed several issues related to the process of knowledge acquisition from document archives containing historical documents. We have compared the documents in archives to primary sources common in historical studies and described their characteristics from the viewpoint of automatic knowledge acquisition. We believe that historical document archives could be more useful for society and should have more value after wide range of applications had been developed for effective mining of historical knowledge.

Acknowledgement. This work has been partially supported by National Institute of Information and Communications Technology, Japan and by MSR IJARC CORE6 project entitled “Mining and Searching Web for Future-related Information”

References

1. Allan, J. (ed.): Topic detection and tracking: event-based information organization. Kluwer Academic Publishers, Norwell (2002)
2. Arms, W.Y., Aya, S., Dmitriev, P., Kot, B.J., Mitchell, R., Walle, L.: Building a research library for the history of the web. In: Proceedings of the Joint Conference on Digital Libraries, pp. 95–102 (2006)
3. Aschenbrenner, A., Rauber, A.: Mining web collections. In: Masanes, J. (ed.) Web Archiving, Springer, Heidelberg (2006)

4. Bollegala, D., Matsuo, Y., Ishizuka, M.: Measuring semantic similarity between words using web search engines. In: WWW 2007, pp. 757–766. ACM Press, New York (2007)
5. Cilibrasi, R., Vitányi, P.M.B.: The Google Similarity Distance. IEEE Trans. Knowl. Data Eng. 19(3), 370–383 (2007)
6. IIPC's Access Working Group. Use cases for Access to Internet Archives (2006), http://netpreserve.org/publications/iipc-r-003.pdf
7. Jatowt, A., Kanazawa, K., Oyama, S., Tanaka, K.: Supporting analysis of future-related information in news archives and the web. In: Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2009), pp. 115–124. ACM Press, Austin (2009)
8. Jatowt, A., Kawai, Y., Ohshima, H., Tanaka, K.: What can history tell us? towards different models of interaction with document histories. In: Proceedings of the 19th ACM Conference on Hypertext and Hypermedia (HT 2008), pp. 5–14. ACM Press, Pittsburgh (2008)
9. Jatowt, A., Kawai, Y., Tanaka, K.: Detecting age of page content. In: Proceedings of the 9th ACM International Workshop on Web Information and Data Management (WIDM 2007), pp. 137–144. ACM Press, Lisbon (2007)
10. Kleinberg, J.M.: Bursty and hierarchical structure in streams. Data Mining Knowledge Discovery 7(4), 373–397 (2003)
11. Masanes, J. (ed.): Web Archiving. Springer, Heidelberg (2006)
12. Ohshima, H., Jatowt, A., Oyama, S., Tanaka, K.: Seeing past rivals: visualizing evolution of coordinate terms over time. In: Vossen, G., Long, D.D.E., Yu, J.X. (eds.) WISE 2009. LNCS, vol. 5802, pp. 167–180. Springer, Heidelberg (2009)
13. Rauber, A., Aschenbrenner, A., Witvoet, O.: Austrian Online Archive processing: analyzing archives of the World Wide Web. In: Proceedings of the 6th European Conference on Digital Libraries, pp. 16–31 (2002)
14. Shipman, F.M., Hsieh, H.: Navigable history: a reader's view of writer's time. New review of hypermedia and multimedia 6, 147–167 (2000)
15. Swan, R., Allan, J.: Automatic generation of overview timelines. In: Proceedings of the 23rd Conference on Research and Development in Information Retrieval, pp. 49–56 (2000)
16. Wexelblat, A., Maes, P.: Footprints: history-rich tools for information foraging. In: Proceedings of Conference on Human Factors in Computing Systems, pp. 270–277 (1999)
17. Villano, M.: Can computer nerds save journalism? Time Magazine (2009)

Virtual Kyoto Project: Digital Diorama of the Past, Present, and Future of the Historical City of Kyoto

Tomoki Nakaya[1], Keiji Yano[1], Yuzuru Isoda[2], Tatsunori Kawasumi[1], Yutaka Takase[3], Takashi Kirimura[3], Akihiro Tsukamoto[3], Ayako Matsumoto[3], Toshikazu Seto[4], and Takafusa Iizuka[4]

[1] College of Letters, Ritsumeikan University,
56-1 Tojiin-kita-machi, Kita-ku, Kyoto, Japan
{nakaya,yano,kawasumi}@lt.ritsumei.ac.jp
[2] College of Asia Pacific Studies, Ritsumeikan Asia Pacific University,
Beppu City, Oita, Japan
yuzuruisoda@gmail.com
[3] Kinugasa Research Organisation, Ritsumeikan University,
56-1 Tojiin-kita-machi, Kita-ku, Kyoto, Japan
takaseyutaka@yahoo.co.jp, {tkiri,atv28073}@fc.ritsumei.ac.jp,
ayakom@fj9.so-net.ne.jp
[4] Graduate School of Letters, Ritsumeikan University,
56-1 Tojiin-kita-machi, Kita-ku, Kyoto, Japan
{gr0008sf,gr0051xe}@ed.ritsumei.ac.jp

Abstract. Using the recently developed 3D GIS (Geographic Information System) and related visualisation technologies, we have created a digital diorama of an entire historical city, which can be used to virtually travel through different realistic landscapes at different times in the history. The digital diorama called Virtual Kyoto is the virtual geographic environment of the past, present, and future of the historical urban spaces in Kyoto City by constructing geotemporal-referenced 3D models of cityscape elements at different eras. In order to promote digital humanities studies on the arts and culture of traditional Kyoto, Virtual Kyoto is used as a digital platform for constructing a web-based digital museum interface with geographic data-linkages to numerous historical and cultural digital contents. We also explore the possibility of using Virtual Kyoto as an information environment to discuss the future of the historical city of Kyoto with the effects of city planning activities such as landscape policies or the possible damage due to disasters on historical landscapes.

Keywords: 3D GIS, virtual reality, digital humanities, landscape planning, 3D urban model.

1 Introduction

The original diorama, invented by Louis Daguerre in 1822, exhibited large landscape paintings inside a theatre-sized facility for experiencing virtual trips to various

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 173–187, 2010.

monumental landscapes [1]. By controlling light effects and turning physical stage sets, the device gave striking panoramas that swiftly switched between day and night and between different time periods. Its smaller variant has been widely adopted as a museum displaying tool for collected objects in an artificial scene. The recently developed 3D GIS and virtual reality methods provide an opportunity to create a digital variant of the diorama to contextualize spatial objects in a virtual geographic environment (VGE) with flexible controls to alter the viewer's position and the contents of the environment [2].

Digital diorama has several advantages over the original diorama: (1) a large physical space to store the contents is unnecessary, even for modelling a large geographic environment; (2) the contents can be easily edited or replaced for various purposes; (3) the system is open to other databases so that any set of digital contents can be linked to geographic positions; and (4) the virtual environment can be shared through networks, particularly the Internet.

In this study, we aim to explore the possibility of using a set of digital dioramas of the Japanese historical city Kyoto, called Virtual Kyoto [3], as a digital platform to enhance information sharing of the arts and culture of Kyoto as well as encouraging the discussion on the future of this historical city and its rich cultural heritage. Virtual Kyoto was created as a GIS-based VGE of the past and present historical urban spaces in the city by constructing a large amount of geotemporal-referenced 3D models of cityscape elements from different ages. So far, our published studies [4-6] have mainly focused on the restoration and modelling of historical landscapes using various modern and historical sources of geographic information. Such efforts of restoring the past can be regarded as a virtual journey through memories of this great historical city. We here take up the ideas and methods of the digital diorama Virtual Kyoto, to retrieve and assemble information on the arts and culture of Kyoto and apply it to planning considerations of the city's future landscape.

This Virtual Kyoto project has been a part of the 'Kyoto Art Entertainment Innovation Research (2002-2006)' undertaken at Ritsumeikan University under the auspices of the 21st Century Center of Excellence (COE) program funded by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). The program is now managed by the Digital Humanities Center for Japanese Cultures and Arts under the Global COE program (2007–2011). In accordance with the recent trend to use new-generation digital recording techniques for chronicling cultural heritages [7], a large amount of digitally archived materials from the arts and culture sector have been accumulated through these programs, including collections of digitalized fine arts, scanned old maps, laser-scanned floats, CAD models of existing historical buildings, and motion captured traditional artistic performance, e.g. 'Noh' and 'Kabuki'. Virtual Kyoto was originally intended to be a platform for integrating such various digitalised contents for contextualizing them in geographic locations (Fig. 1).

The rest of this paper is organized as follows. We provide a brief explanation of the construction of Virtual Kyoto in section 2. Then, section 3, we describe a method to use Virtual Kyoto for constructing a new online digital museum interface with geographic data-linkages to the numerous historical and cultural digital contents including information on culturally valued landscapes. Then, we explore possibilities of using landscape simulations and GIS-data overlays to discuss the future of the historical city of Kyoto and the effect of city planning activities such as landscape policies (section 4) or the possible damage due to disasters on the historical landscapes (section 5).

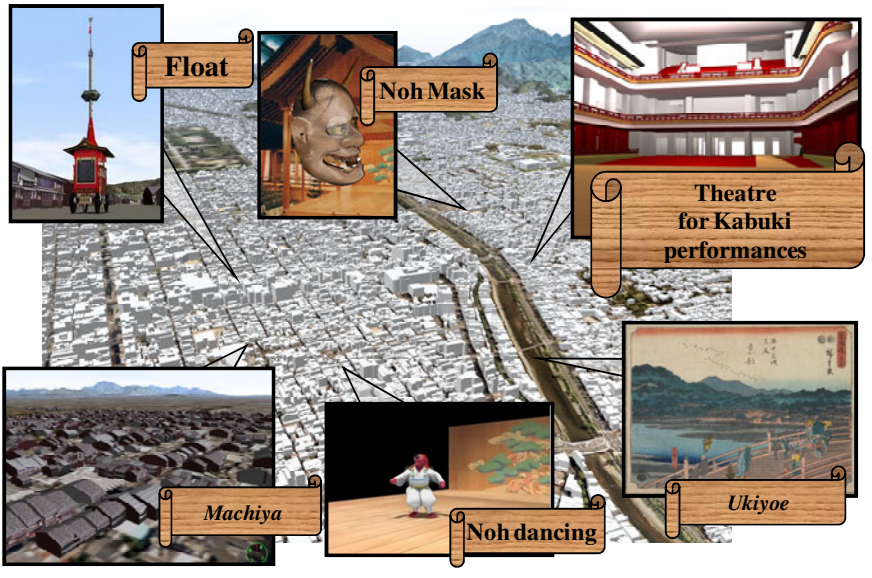


Fig. 1. Concept of Virtual Kyoto integrating digitally archived contents of Kyoto arts and culture

2 Virtual Kyoto

VGEs have emerged along with the rapid development of geotechnologies to enhance 'realism' of geographic representation [2]. A popular application of VGE is the so-called 'virtual city' that is a digital representation of the geographical and geometrical reality of a city. One of the most advanced virtual cities is being created in the Virtual London project by CASA (Centre for Advanced Spatial Analysis), University College London. Virtual London is a large scale 3D GIS with CAD models of Greater London using state of the art computer technology. It is designed for professional use by architects and planners but is also targeted at the wider public through various processes of public participation across the web [8,9].

By highlighting the city's function as an area for collective interactions of various people, the concept of the digital city was also coined by researchers in informatics. In Kyoto, the project of 'Digital City Kyoto' ran from 1999 to 2001 supporting and facilitating social interactions for urban everyday life by developing a web-based social infrastructure. Although the digital city does not necessary entail a VGE of a city, Digital City Kyoto included an interface layer of VGE for the main street, Shijo Street, and a world heritage site, Nijo Castle, in which users can virtually walk through the environments [10,11].

Our Virtual Kyoto is similar and deeply related to conventional virtual city projects such as Virtual London, and digital city projects with VGE interface such as Digital City Kyoto. However, Virtual Kyoto has a unique feature that distinguishes it from its counterparts. In particular, it is a 4D GIS of almost the entire city including a time dimension added to the conventional three dimensional GIS/CAD models. It can store

various geographical information from different historical eras. This unique feature has been crucial from the outset as the goals of the Virtual Kyoto project were focused on depicting the history of the city, as well as reconstructing the city's past urban landscape. Compared to Digital City Kyoto, Virtual Kyoto does not currently provide a general social interaction platform but it is expected to facilitate public-participation for collecting and sharing the information on arts and culture of Kyoto, and eventually to contribute to create an environment for better everyday living in Kyoto.

In Virtual Kyoto we have adopted MAPCUBE® data from CAD Center, Increment P and Pasco, Ltd., for creating the present landscape modelling. The data consists of prismatic 3D building block models based on the building footprint and height for the whole city of Kyoto (Fig. 1). The 3D models are created by extruding the building footprint to the height obtained by an air-borne laser-profiler which recorded height values at an interval of 2.5 m with an error within 15 cm (http://www.mapcube.jp/en/product.html). We replaced the simple extruded building model, 'white model', with detailed CAD models with photo realistic texture images for buildings regarded as landmarks in the historical Kyoto landscape, such as world heritage temples and shrines.

An important aspect of Kyoto rises from the fact that the city, founded in 794 AD, was the historical capital of Japan and still retains a large number of historical architecture such as temples, shrines, machiya, and modern western-style buildings of the pre-war period. This is a rare case in Japan; the city managed to escape damages during World War II. Our strategy in restoring historical Kyoto involves the modelling of present day Kyoto as a virtual city, then replacing the newer buildings with traditional buildings as we go back in time. For this purpose, various sources of past landscapes including aerial photos or old maps, landscape drawings and paintings were collected and georeferenced in a GIS environment. For restoring the early landscape of ancient Kyoto called Heian-kyo (ca. 8^{th}–12^{th} centuries), landscape modelling was independently conducted by following blue-prints of physical miniatures of the Heian-kyo diorama that were made to mark the 1200^{th} anniversary of the founding of the city in 1994. Fig. 2 demonstrates 3D sceneries portraying many geographical objects at three different eras in Virtual Kyoto (see [3,4,5,6] for details on the modelling procedure of these geographic environments). For the landscape visualisation, we employed UrbanViewer™ (CAD Center, Ltd.), a virtual reality desktop application for real-time viewing of detailed urban scenes using a 3D urban model of MAPCUBE®.

We have thus created the digital diorama of Kyoto as a 4D GIS environment (3D space plus time dimension) of Kyoto landscape elements. This supports a number of new areas of geographic inquiry, especially with its ability to represent the implications of historical changes in landscape on the city's current and future policies.

Virtual Kyoto comprises the following four functions: (a) archiving georeferenced materials such as current digital maps, old topographic maps, cadastral maps, aerial photos, picture maps, street photos, landscape paintings, archaeological sites data, and historical documents; (b) creating a database of all existing buildings including machiya, early modern buildings, and shrines and temples including those of historical and cultural significance; (c) creating 3D models of the buildings mentioned above; and (d) estimating and simulating land use and landscape changes over the study periods using the aforementioned materials.

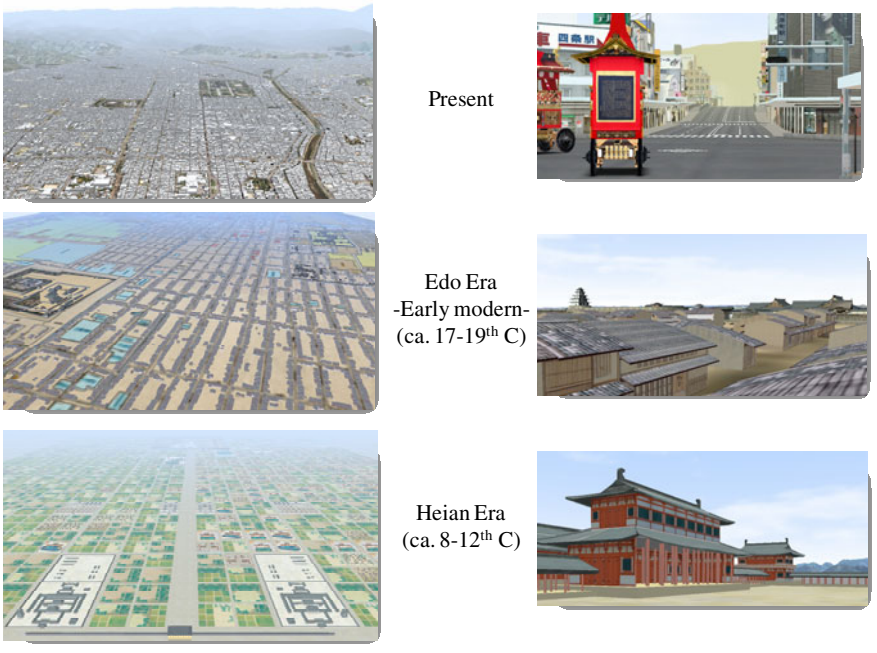


Fig. 2. Digital dioramas of Virtual Kyoto at present and in two historical eras

3 Digital Museum of Kyoto Arts and Culture on Web-Based Digital Diorama

Following the original intention of the Virtual Kyoto project, we currently maintain a web-site using the 3D web-GIS engine, UrbanViewer™ for Web (CAD Center, Ltd.), to disseminate the digital diorama contents of Kyoto with links to other digital archives of Kyoto arts and culture. The URL of the site is as follows:

http://www.geo.lt.ritsumei.ac.jp/uv4w/frame_e.jsp

Like GoogleEarth (Google Inc.) and Virtual Earth (Microsoft Inc.), the engine enables us to transmit and receive a great amount of 3D landscape information including buildings, landforms and textured images with interactive manipulation of view settings on ordinary DSL-level-speed internet infrastructure by LOD (level of detail), data compression and streaming techniques of smoothed images. LOD is a technique that is widely used for virtual reality to control in real time the quality of 3D models' display. It automatically changes the display/non-display of 3D objects and also displays the object's geometry and texture images with the appropriate data amount and resolution according to the distance between the client and the viewed objects and the client's viewing angles.

To reduce the amount of data to be transmitted, a simple way of data reduction is used on extruded building models that often occupy a large space in a view. This

reduction method stores the geometry data as the polygon of its base shape and height instead of the polygons of all the surfaces. In order to achieve smooth real-time display of the texture images, different resolutions of the images as well as digital terrain models are stored in the database server. According to the distances between the viewing positions and viewed objects, and the client's viewing angles, the web server orders the database server to extract the image data with the most appropriate resolution for the client's view and the extracted image data are sent to the client's PC. For browsing through a website constructed by the engine, a client is required to use Microsoft Internet Explorer with the UrbanViewer^{TM} for Web plug-in installed.

In the web-based digital diorama of Virtual Kyoto, users can freely fly over or walk through landscapes of Kyoto at different times (present, early 20th century, and 12th century). Fig. 3 shows an example of such an instant virtual time-travel as viewed on the web. A pair of views using the same viewing position and angle but showing scenes at different time periods vividly demonstrate how one could feel the sense of the ancient landscape around Heian-kyu, the rectangular ancient imperial palace that was built at the beginning of Heian-kyo in late 8th century, and which was largely transformed into the current built-up areas with irregular skylines of buildings. We

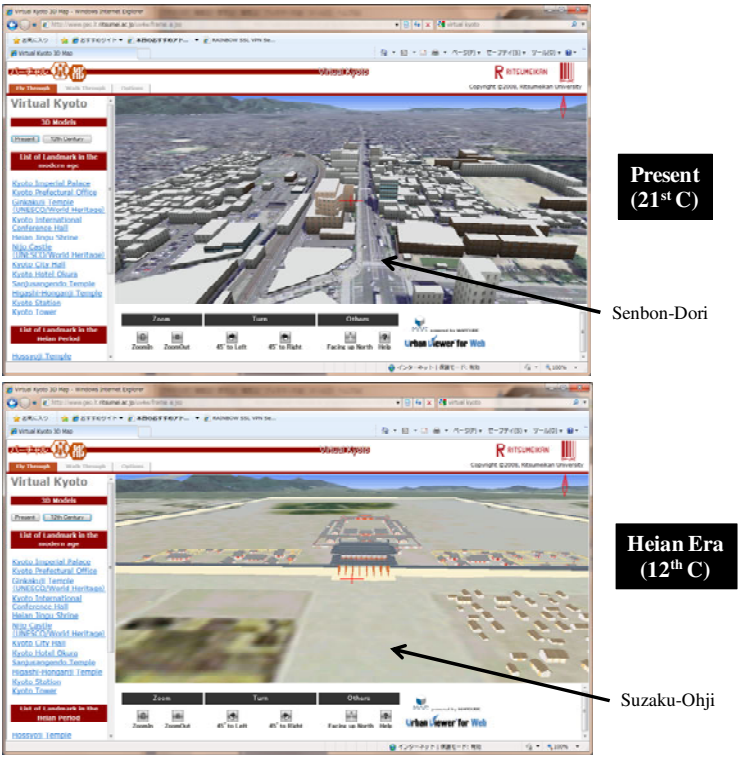


Fig. 3. Virtual time-travel of Kyoto landscape on web

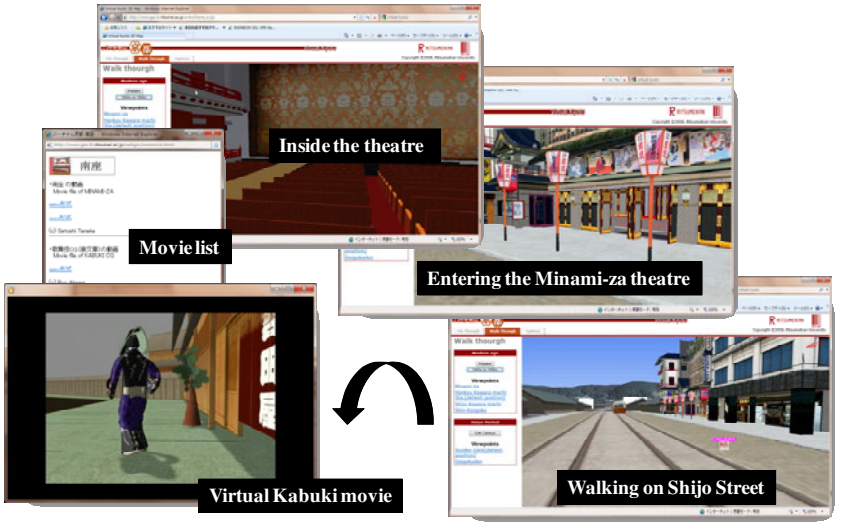


Fig. 4. Virtual tour of virtual Kabuki performance in Virtual Kyoto

see that a current major street, Senbon-Dori still runs along the ancient route of the main Heian street, Suzaku-Ohji, connected to the former palace main gate, but that the width of the current street is much narrower compared to that of the ancient one.

Glancing across the digital diorama, we can browse existing historical arts and cultural materials by clicking geolocated markers with labels of the digital contents. We herein make use of the digital diorama as a gallery or museum platform of digitised fine art objects related to specific locations in Kyoto.

Fig. 4 demonstrate how the system enables a virtual tour that includes experiencing a virtual Kabuki performance, created by the technology of motion capture of Kabuki actors, at the historical theatre, Minami-za, on Shijo Street. The tour begins with a walk down the street that may represent the early Showa era (ca. 1920–1930) then a change in the skyline of buildings and vistas of the mountains conveys the feeling of the rapid urban development after World-War II. We can seamlessly walk into the theatre from the road to see a virtual Kabuki performance; the inside of the theatre is also modelled and incorporated the online virtual environment. By clicking the link marker in the theatre, a popup window showing the list of streaming movies appears. Although there are currently only two movies of digitised performances in our system, in the future it would be possible to experience a series of virtual experimental performances. In earlier townscape settings, we hope to present a variety of older style performances in old theatre stages by combining motion-captured movements with old costumes reflecting those in historical records.

As for historical landscapes, it is difficult to evaluate how physical landscapes are culturally assessed and valued. A clue may be obtained from popular landscape paintings that are often read by architects or geographers for debating humanistic values of historical landscapes [12]. The web-based Virtual Kyoto has a set of links to

Fig. 5. Browsing Ukiyoe landscapes in Virtual Kyoto

digitalized art contents, scanned images of traditional Japanese Ukiyoe woodblock prints (Fig. 5). The sets of Ukiyoe landscapes that date back to the Edo Era (ca. 17th–19th centuries) provides a way to see lost landscapes. The views from the Ukiyoe paintings are geolocated in the diorama and are also linked to the Ukiyoe Database System maintained by the Ritsumeikan Art Research Center, which provides detailed information of each items and enable a search for other Ukiyoe contents by using keywords such as theme, painter and year.

As seen in Fig. 5, we often find mountains in the background of Ukiyoe prints, contrasting with the subject of the urban scene. Since many of the well-known places in Kyoto are situated near mountains, the main theme of the Ukiyoe stands out when it is set against those backgrounds which are used as a 'borrowing landscape' – a concept of traditional landscape design applied in many Japanese gardens. It exemplifies that the vistas on surrounding mountains from specific viewing points in the city have been appreciated as landscapes with the cultural values of Kyoto.

4 Landscape Simulator Using Editable Diorama of Kyoto

As seen in the previous section, historical Kyoto vistas of mountains often have a cultural meaning. We are particularly conscious of this when bonfires are lit to send off the dead during the All Souls (Obon) festival on August 16th, which marks the peak of Kyoto's hot late summer. Publicized to tourists under the name of 'Daimonji Gozan Okuribi', (the Five Mountain Great Character Sending Off of the Dead), a bonfire formed as a great character is lit for each of the five mountains at night. This broadly popular annual event was well established in the Muromachi (ca. 14th–16th centuries) and Edo (ca. 17th–19th centuries) eras.

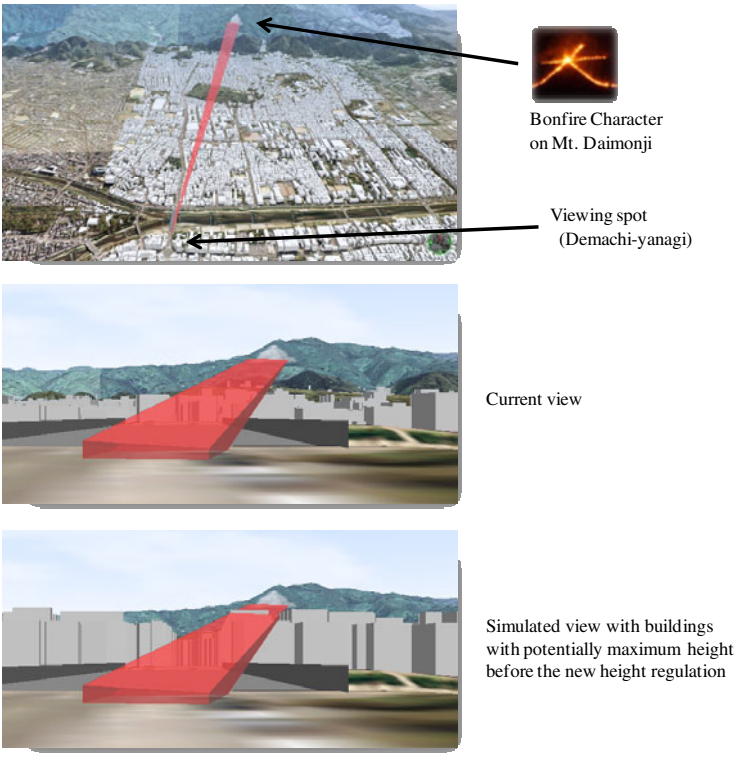


Fig. 6. Evaluating potential risk of obstructed view from possible development

Demachi-yanagi, the meeting point of the Kamo and Takano rivers, is a popular viewing spot to see the first bonfire lit on the hillside of Mt. Daimonji (Nyoigatake). Virtual Kyoto was applied to evaluate the potential risk of obstructing this view due to possible future development of buildings with the maximum height allowed under the previous building and development regulations. In Fig. 6, a red half-transparent board represents the vista line from the viewing spot to the fire bed of the bonfire character.

While no building height currently exceeds the line, the clear view could potentially be obstructed by future development and buildings that comply with the building height restrictions (see the bottom image in Fig. 6). To avoid such potential destruction of historical views in Kyoto, the Kyoto City Government initiated a new landscape policy in 2007 (the Kyoto City Landscape Policy)[13], to protect the beautiful historical urban landscape of the city. The policy puts direct control on the appearance and height of buildings from four aspects: height restriction in the city centre; conservation of scenic views and 'borrowed landscapes' from certain viewing spots; design restrictions on buildings; and regulation against the display of outdoor advertisements. Following this new scheme, 38 magnificent vistaed views and 'borrowed landscapes' were designated as heritage views to be conserved and new height regulations based on altitude were introduced, with the help of landscape simulations, so as not to hamper those significant vista.

Current street view

Simulated street view without exterior signs and electric poles

Web-GIS interface for simulating the street view with radio buttons to display or hide exterior elements

Fig. 7. Simulating street view on Shinmachi Street

The changes in building exteriors, the impact of removal of business signs and billboards, burying electrical cables, and other related changes from proposed regulations can be also envisioned with landscape simulations in Virtual Kyoto. We developed a simulator of detailed landscape elements on Shinmachi Street in the central district of Kyoto (Fig. 7). The local town along the street is one of historical community called 'Hoko-machi', where a large float called 'Yamahoko (or Yamaboko)' that parades in the Gion Festival is maintained.

Walking through such a simulated streetscape, users including local residents and local governmental officers can get a feeling for the possible results of new street exterior regulations or voluntary actions to improve streetscapes. Considering the fact that some landscape modification such as removing electric poles are quite costly, and more importantly that the new regulations are quite controversial among different stakeholder groups such as developers, local residents and newcomers living in high-rise apartment houses, it is vital to publicly share the possible results associated with the landscape regulations for meaningful debate on the future of the city. As part of web-based Virtual Kyoto, we also developed web-based simulators with buttons to display or hide each landscape elements in Shinmachi Street. As an environment of PPGIS (public participating GIS) [14], the web-based system is expected to encourage public participation in discussions on landscape policy and promote involvement and commitment from a wide range of public sectors.

5 3D Hazard Map of Kyoto Using Digital Diorama

As an extension of the previous landscape simulation, we can also see the impact of disasters like floods or earthquakes by overlaying thematic layers of hazard risks over the digital diorama of the city. We have created and distributed another version of the web-based digital diorama of Virtual Kyoto, 'the Safety and Security 3D Map of the Historical City Kyoto' combined with the hazard map and crime statistics of the city. This can be found at http://www3.rits-coe.jp/ritsumei_kyoto/.

Hazard maps are maps that show information regarding evacuation facilities/spaces and the degrees of hazard risks, to help prepare for future disasters. The Kyoto Fire Department created an award-wining hazard map (a city-wide map was prepared in 2004; maps of each ward were published in booklet form in 2005) showing disaster-related facilities such as hospitals, and various hazard risks from landslides, floods and earthquakes. It combines a voluminous amount of geographical data in a GIS. We use this information for our 3D hazard map system.

As in the case of the previously shown web-based Virtual Kyoto, the 3D map is an interactive map displaying realistic landscapes in which we may freely fly through the virtual scenery. Most hazard maps have two-dimensional views of broad areas but a 3D map permits movement within the map, allows exploration of the distribution of various risks at different scales and angles so that a user can explore and understand the hazard risks related to their own spheres of activity. Using the relational database engine with spatial index, the system has a function to search locations of facilities such as police stations, hospitals, and evacuation shelters with their attributes (such as capacity of shelters) as well as cultural heritages and national treasure architectures around the current view point.

Kyoto hosts a large number of historical buildings such as temples, shrines, machiya (traditional wooden townhouses), and western-style buildings of the pre-war period. Particularly, a great number of world heritage and national treasure architectures are accumulated in Kyoto. These architectural heritages significantly contribute to Kyoto's historical urban landscape. It is estimated that if a serious disaster hit Kyoto, a large number of heritage sites would be lost and their restoration would be extremely expensive. Thus, establishing a policy to mitigate possible damages of historical sites by future disasters is important to sustain the historical landscapes in Kyoto [15,16].

Currently our system includes only cultural heritages and national treasure architectures but in the future it would include other types of historical buildings, such as machiya. Machiya in Kyoto is specifically called *kyo-machiya* for its distinctive character and there remain about 48,000 kyo-machiya in various conditions across the historical parts of the city. We have constructed a GIS-based monitoring system for kyo-machiya with the aim of supporting the government's step towards preserving this traditional architectural style [17,18]. The system is based on the 'Kyo-machiya Community-building Survey', on which the Kyoto City, Ritsumeikan University, volunteer architects and citizens have been working together since October 2008. In this field survey, we have used the Mobile GIS (ArcPad) in PDA to extract the exact geospatial information on houses and buildings in addition to their photographs. The

A bird's eye view on flooding risk along Kamo river

View of building collapse risks caused by a possible earthquake (Hanaore fault) around Higashi-Honganji temple

Fig. 8. 3D hazard map generated using Virtual Kyoto

information is maintained by ArcGIS and PhotoField [19], a photo-album application with geographic information (Fig. 9). The machiya survey entries include: its type, its condition, design elements of the façade, and whether or not it is vacant. Such information of overall machiya condition with accurate geographical coordinates may be useful for assessing hazard risks of neighbourhoods filled with machiya (e.g. for fire propagation or road blocking) as well as for realising fragmented historical cityscapes at a risk of being lost by a possible disaster.

Combining such field survey results of historical buildings with the 3D hazard map using Virtual Kyoto would efficiently empower policy makers as well as local residents to consider the necessity of formal and informal rescue organizations to protect or prepare endangered cultural landscapes for future disasters.

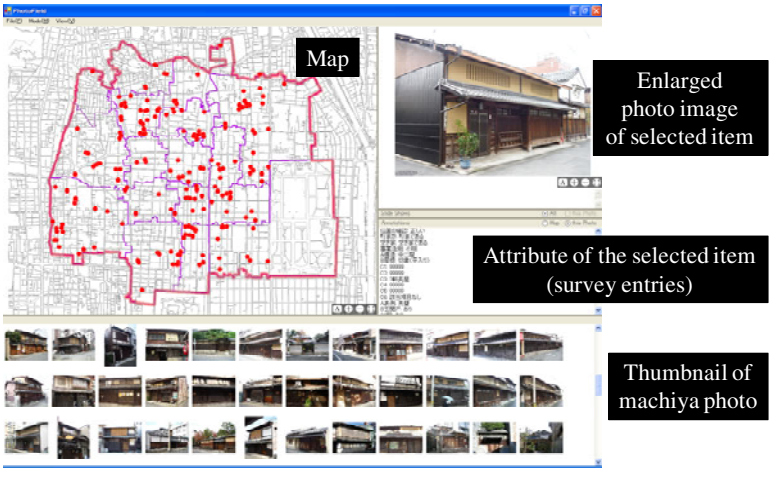


Fig. 9. Sample screenshot of kyo-machiya GIS database using PhotoField: geographic locations, attributes, and photographs of well-preserved kyo-machiya identified by field survey in northern central part of downtown area of Kyoto

6 Concluding Remarks

Since the passing of laws regulating the preservation of old Kyoto in 1966, the city has introduced many development restrictions and ordinances to preserve its historic character. The largest number of streetscapes and buildings that need to be preserved as cultural assets in Japan exist in Kyoto. Post-war Japan has seen rapid economic growth, and Kyoto's streets have already undergone extremely fast changes and innumerable historical monuments and places have been destroyed by the impact of development. The remaining heritage sites and landscapes are still endangered by future developments and disasters.

In this paper, we have described how to apply the digital diorama, Virtual Kyoto, to share valuable information regarding the historical city, from historical vistas shown in arts to simulated sceneries with possible future scenarios. In particular, by using the web-based diorama to distribute 3D panoramic scenes to clients, it will be possible for many people to use the perspectives to learn from and appreciate the city's historical environment combined with arts and cultural contents of Kyoto. In the process, the city's historical memory will become the common domain of many people and a base for discussing the future of the historical city. Digital humanities [20], a newly emerging field for studying humanities using digital information, may encourage debates on city culture between humanity and policy-planning researchers by sharing digital information in a shared information environment like the digital diorama.

One challenge is to extend the use of the digital diorama, Virtual Kyoto, and enrich it through possible public input involving a wider range of people for collective thinking of the past, present, and future of the historical city, Kyoto. Further research is needed to evaluate how the nature of digital diorama that enables shared realistic viewing experience of landscapes might change public involvement in of archiving historical materials and decision making in city planning.

Acknowledgments. This research project was financially supported by the Global COE Program (Digital Humanities Center for Japanese Cultures and Arts), Academic Frontier Project (Disaster Mitigation of Urban Cultural Heritage), and MEXT KAKENHI (Grant-in-Aid for Scientific Research (B) 21320161). We would like to thank our collaborators at Ritsumeikan University for permitting us to use the digital archives of arts and cultures of Kyoto: Prof. R. Akama (Virtual Kabuki in Fig. 4) , Prof. K. Hachimura (motion captured Noh dancing in Fig. 1), Prof. H. Tanaka (3D model of Noh mask in Fig. 1), and Prof. S. Tanaka (CAD model of Minamiza theatre in Figs. 1 and 4). We obtained the images of Ukiyoe shown in Figs. 1 and 5 from the Ukiyoe Database System maintained by the Ritsumeikan Art Research Center.

References

1. Gernsheim, H., Gernsheim, A., Daguerre, L.J.M.: The History of the Diorama and the Daguerreotype. Dover Publications, New York (1968)
2. Lin, H., Batty, M.: Virtual Geographic Environments. Science Press, Beijing (2009)
3. Yano, K., Nakaya, T., Isoda, Y. (eds.): Virtual Kyoto, Nakanishiya, Kyoto (2007)
4. Yano, K., Isoda, Y., Nakaya, T., Kawasumi, T., Matsuoka, K., Takase, Y., Kawahara, D., Kawahara, N., Inoue, M., Tsukamoto, A., Kirimura, T.: Creating virtual time-space of Kyoto. E-journal GEO 1(0), 12–21 (2006)
5. Yano, K., Nakaya, T., Isoda, Y., Takase, Y., Kawasumi, T., Matsuoka, K., Seto, T., Kawahara, D., Tsukamoto, A., Inoue, M., Kirimura, T.: Virtual Kyoto: 4D-GIS comprising spatial and temporal dimensions. Journal of Geography (Chigaku Zasshi) 117(2), 464–478 (2008)
6. Isoda, Y., Tsukamoto, A., Kosaka, Y., Okumura, T., Sawai, M., Yano, K., Nakata, S., Tanaka, S.: Reconstruction of Kyoto of the Edo era based on arts and historical documents: 3D urban model based on Historical GIS data. International Journal for Humanities and Arts Computing 3(1-2), 21–38 (2009)
7. Baltsavias, E., Gruen, A., Gool, L.V., Pateraki, M.: Recording, Modeling and Visualization of Cultural Heritage. Taylor & Francis, London (2006)
8. Hudson-Smith, A., Evans, S., Batty, M.: Building the Virtual City: Public Participation through e-Democracy. Knowledge, Technology & Policy 18(1), 62–85 (2005)
9. Hudson-Smith, A., Milton, R., Dearden, J., Batty, M.: Virtual Cities: Digital Mirrors into a Recursive World. In: CASA Working Paper 125 (2007)
10. Ishida, T.: Digital City Kyoto. Communications of the ACM 45(7), 76–81 (2002)
11. Ishida, T.: Activities and Technologies in Digital City Kyoto. In: van den Besselaar, P., Koizumi, S. (eds.) Digital Cities 2003. LNCS, vol. 3081, pp. 166–187. Springer, Heidelberg (2005)
12. Higuchi, T.: The Visual and Spatial Structure of Landscapes. MIT Press, Cambridge (1988)
13. Kyoto City: Kyoto City Landscape Policy: Forming Timeless and Radiant Kyoto Landscapes (September 2007), http://www.city.kyoto.lg.jp/tokei/cmsfiles/contents/0000062/62129/HP-English.pdf
14. Ramasubramanian, L.: Geographic Information Science and Public Participation. Springer, Heidelberg (2010)

15. Muranaka, A., Nakaya, T.: Economic valuation of the historic landscape restoration from disasters: a case study using the CVM in Kyoto, Japan. Rekisitosi-bosai-ronbunsyu (Papers on Disaster Mitigation of Historical Cities) 3, 245–252 (2009)
16. Muranaka, A., Nakaya, T.: Investigation into the Willingness-to-Pay for the historic landscape restoration from disasters by structural equation modeling. Papers on Environmental Information Science 22, 73–78 (2008)
17. Seto, T., Matsumoto, A., Iizuka, T., Yano, K.: Public Participation GIS of Historical Landscape: A Case Study of Kyot-Machiya Community-building Survey in Kyoto City. In: Proceedings in the 22nd CIPA Symposium, p. 5 (2009)
18. Iizuka, T., Matsumoto, A., Seto, T., Yano, K.: GIS-based Monitoring Systems for Kyo-Machiya in Kyoto City. In: Proceedings in the 22nd CIPA Symposium, p. 5 (2009)
19. PhotoField – Spatial Photo Album, http://www.s-it.org/photofield/
20. Schreibman, S., Siemens, R., Unsworth, J. (eds.): A Companion to Digital Humanities. Blackwell, Oxford (2004)

A Platform for Mining and Visualizing Regional Collective Culture

Shin Ohno, Shinya Saito, and Mitsuyuki Inaba

Digital Humanities Center for Japanese Arts and Cultures,
Ritsumeikan University
56-1 Tojiin-Kitamachi, Kita-ku, Kyoto 603-8577, Japan
{shin,ken,inabam}@arc.ritsumei.ac.jp

Abstract. This paper proposes computational methods for mining and visualizing collective culture among the community members of a region. This paper first outlines a procedure to extract significant narratives with text mining technique and spatiotemporal analysis on the textual data transcribed from oral-history interviews with the regional community members. It also introduces the KACHINA-CUBE system that imports the narratives as contextualized fragments of sentences based on spatiotemporal information, visualizes them onto a virtual 3D space, and assist researchers to discover commonalities and diversities among them based on the trajectory equifinality model (TEM), which is a theoretical framework to clarify both the similarities and differences among the trajectories of individual life courses. At the end of this paper, we illustrate a test case on collective culture regarding the once-flourishing film industry in Kyoto.

Keywords: collective culture, personal culture, text mining, narrative analysis, qualitative GIS.

1 Introduction

A new interdisciplinary approach, Cultural Computing [9] focuses on applying computer technology for preserving, sharing, and analyzing tangible and intangible culture. Since the concept of culture is broad, complex, and ambiguous, it is important for researchers to recognize their perspective or scope for approaching to culture before they apply computing technology to cultural phenomena or cultural artifacts.

In this paper, we set our viewpoint on culture by adopting the concept of culture proposed by J. Valsiner [10]. A socio-cultural psychologist, he argues that the culture is a dynamically emerging phenomenon that is co-constructed by the interplay between "personal culture" and "collective culture." The former consists of personal sense, memory and knowledge, while the latter is considered as collection of the former. There exists dynamic interplay between personal and collective cultures, and they go through the process of "internalization" and "externalization." Through the process of communication with others, personal

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 188–199, 2010.

culture is externalized. In other words, shared within the community, the personal culture affects the collective counterpart. On the other hand, collective culture is internalized to the individual's sense and personal culture.

Since both personal and collective cultures have intangible, complex, and ambiguous characteristics, capturing them in socio-cultural contexts is one of challenging tasks for researchers. To understand this contextual feature of culture, we specifically focus on spatiotemporal information as a key to capture personal and collective cultures. When people recount their personal experience, they frequently give utterances directly or indirectly related to spatiotemporal information. We also found different people sometimes refer to the same location or time in their stories. Intersection of such stories could be considered as a clue to understand collective culture. Although personal culture may not be externalized directly, we could have a glimpse of the collective culture from collection of independent stories as a result of indirect externalization of personal culture.

In the following, we propose a method to extract the information for personal culture from transcribed oral histories. The extraction procedure is followed by the introduction of the KACHINA CUBE (KC) system, a web-based platform for visualizing and mining regional collective culture. We, then, test KC with a case of collective culture from three research participants, regarding the once-flourishing film industry in Kyoto. Finally, we evaluate our methods and the platform by comparing to the related works.

2 Mining Personal and Collective Cultures

This section elucidates the procedure of extracting information of personal and collective culture from the narrative data. In reality, the extraction process follows the research that involves collecting narratives from research participants. It should be also mentioned that the linguistic processing in this section is partially assuming the Japanese textual data transcribed from oral-history interviews with the regional community members in a certain area of Kyoto, although the main part of the following procedure would be applicable to other languages.

2.1 Extracting Personal Culture

We consider the spatiotemporal information appeared in narrative data would be a guide to split them into chunks related to personal culture. Fig.1 explains procedure to extract personal culture from the narrative data. The procedure contains morphological analysis, syntactical analysis, and spatiotemporal analysis. After obtaining the narrative data, researchers first perform morphological analysis on them, splitting them into morpheme units. Using both the basic and user dictionaries, we can detect morpheme units from the original transcripts. To explain this in more detail, the user dictionary covers miner or specific terms that are not registered in the basic dictionary. For example, "Uzumasa-Eigamura," or Uzumasa Movie Village, indicates one specific location, which the basic dictionary might not register as a proper noun. Such names of locations, events, persons, and buildings are considered as entries for the user dictionary.

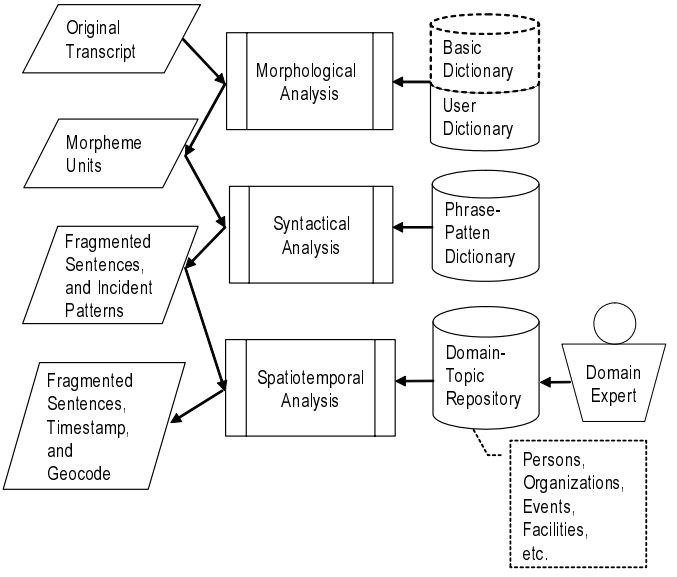


Fig. 1. The flow of extracting personal culture from the original transcripts. The whole process entails three phases of analyses, morphological, syntactical, and spatiotemporal.

The next step is syntactical analysis that examines the transcripts based on the phrase pattern dictionary that has frame sets of incidents. The analysis outputs fragmented sentences with incident patterns. For example, the sentence, "My uncle started Makino Cinema Theater at Kyoto Senbon area". would be detected by a syntactical analysis. In this case, the phrase pattern dictionary has one entry with temporal information, which is "starting Makino Cinema Theater," and another with spatial information of "Kyoto Senbon area".

Final process is spatiotemporal analysis that maps between spatiotemporal information and incident patterns. Meta information such as timestamp and geocode is added to the fragmented sentences. For this process, we propose the domain-topic repository, which is developed based on the information from domain experts in order to define the semantic associations between persons, organizations, events, facilities and so on. The linkages from the information to the spatiotemporal information are also stored in the repository. With the repository, it becomes possible for the researchers to traverse the terms or phrases in the narratives and add meta information to the fragmented sentences.

In short, our way of analyzing personal culture makes use of three different databases: the user dictionary for morphological analysis, the phrase-pattern dictionary for syntactical analysis, and the domain-topic repository for spatiotemporal analysis.

2.2 Bridging Personal Culture to Collective Culture

The databases we mentioned in the previous section are reusable for researching the same periods, the same areas, or both. The same spatiotemporal information,

such as names of buildings and events are more likely to turn up if referred locations are close enough. Since the dictionaries and domain-topic repository are reusable, the more pieces of information are stored to them, the easier the process gets to extract spatiotemporal information from other sources.

The collection of fragmented sentences based on the narrative data from the regional community members can be considered as a clue to understand collective culture in the region. As we discussed, even though the community members did not meet each other those days, and did not go through the externalization process as Valsiner proposes, it is highly likely that they share the similar experience, which means there exists the possibility of indirect externalization. For example, a research participant has some story about his/her experience in the elementary school, and another research participant went to the same elementary school in the similar time period. They might or might not have known each other then, but it is likely that they had some friends, teachers, and school events in common. Through such intermediates, each personal culture gets externalized and shared by the community. In the end, the collection of personal stories based on spatiotemporal information can be considered as collective culture like this case.

3 Visualizing Collective Culture

The following sections introduce the KC system as a platform for visualizing regional collective culture. It imports the contextualized fragments of sentences generated by the procedure above, visualizes them onto a virtual 3D space based on spatiotemporal information, and assists researchers to understand collective culture shared by the community members in a region.

3.1 User Interface

The KC system provides a web-based virtual 3D space to its user (see Fig. 2). The space includes a two dimensional map with timescale, and is used to store and plot the contextualized fragments of sentences extracted from the narrative data based in spatiotemporal information.

Fig. 3 shows a 3D space from two different angles. It consists of a map with timescale. The user can rotate the 3D space to browse the fragmented sentences or narratives from suitable angles.

3.2 Data Structure

Expecting KC to be sharable, we apply RDF (Resource Description Framework)/ OWL (Web Ontology Language) [12] to define the classes for storing narrative data. RDF and OWL are Semantic Web standards that provide a framework for sharing and reuse of data on the Web. Fig. 4 shows a part of the class definition in KC. Under the **Document** class, the following classes are defined.

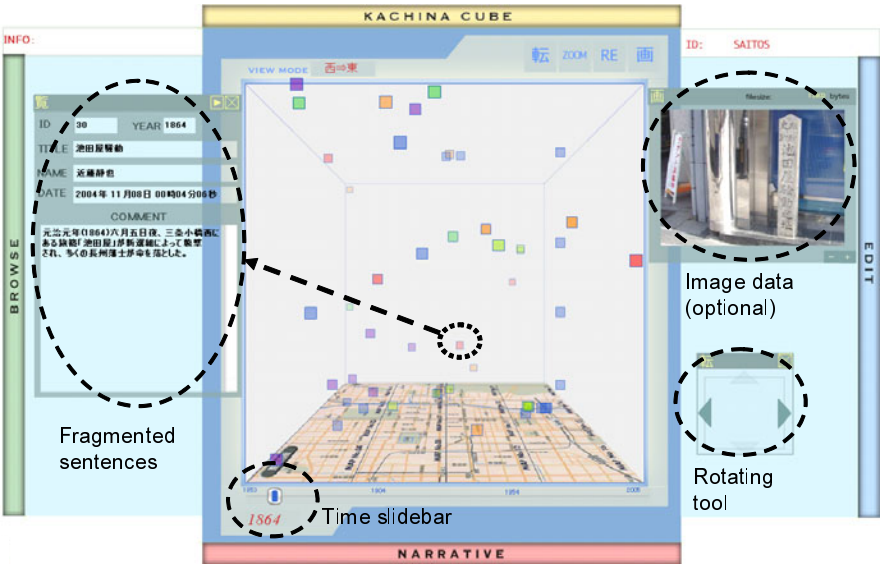


Fig. 2. User interface of the KC system. The central window shows a 3D space with a map and markers to fragmented sentences. The user can define links to images or video clips to the markers. The user also can turn around the 3D space with the rotating tool.

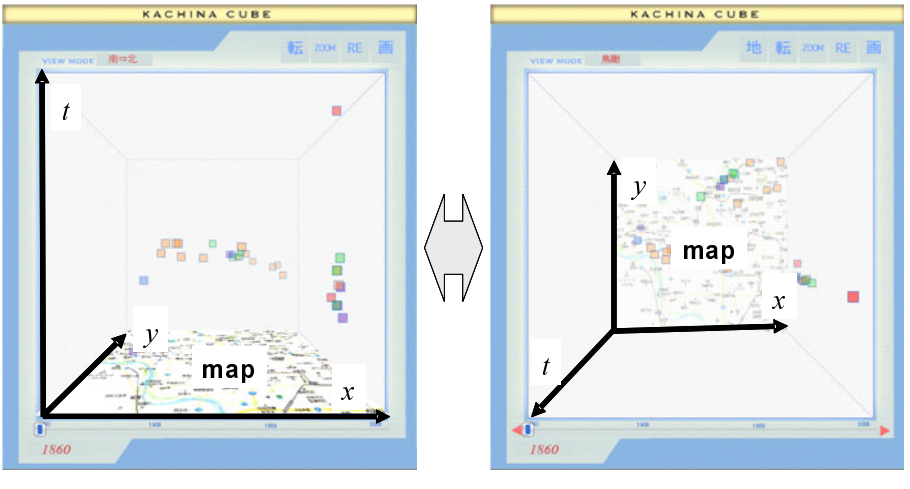


Fig. 3. A virtual 3D space from two different angles. It consists of a map (dimension x and y) with timescale (dimension t).

Fragment class:
Generic class for fragmented sentences
StoryFragment class:
Fragmented sentences from narrative data
HistoryFragment class:
Topics that are considered as historical fact
Storyline class:
Contextualized data consists of story/historical fragments

The **Attribute** class is used to represent the spatiotemporal information of the narrative data in KC. It has the followings two subclasses.

Geography class:
Geographical information of story and historical fragments
Temporal class:
Time related to the incidents in story or historical fragments

Since each fragment has both spatial and temporal information, the standard format is suitable for KC and motivates other researchers to access the data. It should be noted that this data format allows us to define semasiological relationship and class structure among fragmented sentences.

3.3 OPP Detector

Valsiner and Sato [11] proposed the concept of Trajectory Equifinality Model (TEM) as a framework to analyze personal experience with diverse and possible trajectories. One of the components to support TEM, OPP (Obligatory Passage Point) is the point which most of the people have to go through because of their socio-cultural contexts, institutions, and customs.

Implemented in KC, OPP detector can find fragmented sentences related to similar space or time in multiple narratives provided by research participants. Fig. 5 illustrates how KC finds OPP among the narratives. Narratives A to D in the figure go through similar space and time. OPP detector of KC recognizes the intersection of the narratives as OPP. In other words, OPP detector assists researchers' understanding on regional collective culture based on the analysis on multiple personal life courses.

4 Application

4.1 A Test Case of Implementation

To advocate the method, we implemented tools and databases to extract the information related to personal culture based on the method we discussed above, and imported the extracted data into the KC system.

As a test case, we used the narrative data of movie culture in Kyoto's Rakusai area, a.k.a., Japan's Hollywood. We utilized oral history data collected by [6], [7],

```
<owl:Class rdf:ID="Document"> </owl:Class>

<owl:Class rdf:ID="Fragment">
  <rdfs:subClassOf rdf:resource="#Document" />
</owl:Class>

<owl:Class rdf:ID="StoryFragment">
  <rdfs:subClassOf rdf:resource="#Fragment" />
</owl:Class>

<owl:Class rdf:ID="HistoryFragment">
  <rdfs:subClassOf rdf:resource="#Fragment" />
</owl:Class>

<owl:Class rdf:ID="StoryLine">
  <rdfs:subClassOf rdf:resource="#Document" />
</owl:Class>

<owl:ObjectProperty rdf:ID="consistOf">
  <owl:domain rdf:resource="#StoryLine" />
  <owl:range rdf:resource="#Fragment" />
</owl:ObjectProperty>

<owl:Class rdf:ID="Attribute"></owl:Class>

<owl:Class rdf:ID="Geography">
  <rdfs:subClassOf rdf:resource="#Attribute" />
</owl:Class>

<owl:Class rdf:ID="Temporal">
  <rdfs:subClassOf rdf:resource="#Attribute" />
</owl:Class>
```

Fig. 4. RDF/OWL representation of class definition in KC

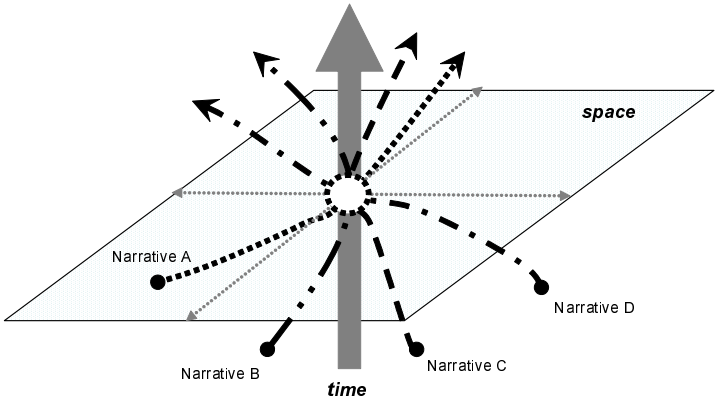


Fig. 5. Image of OPP as an intersection of the trajectories of multiple narratives provided by research participants

and [8]. Those oral history data were collected from three research participants who had involved in the movie industry in the area from the 1910s to the 1930s. Each research participant did not refer to each other in the data, but their experiences are considered as connected in some cases. Therefore, extracting personal cultures from all three research participants can give us some ideas about collective culture of that area. The step-by-step procedures to use the method are as follows:

1. Read the oral histories and collect direct spatiotemporal information from them.
2. Enter the direct spatiotemporal information to the user dictionary.
3. Enter the indirect spatiotemporal information such as names of persons, buildings, and events to the phrase-pattern dictionary.
4. Extract fragmented sentences from both morphological and syntactical analyses by using the dictionaries.
5. Add the domain-topic repository to map the spatiotemporal information to computer-readable data.
6. Add timestamp and geocode to the fragmented sentences.
7. Store the data to KC, and see how it works.

4.2 Analyzing Personal Culture

For morphological analysis, we add the user dictionary to the MeCab program [4]. Furthermore, to make the syntactical analysis easier, we also add the phrase-patterns to the program. After doing so, we prepare a tool for these two analyses to split text into lines which can go through the line-by-line analysis. As an experiment, we set the threshold parameter as three lines, which are treated as a unit of fragmented sentences. With the domain-topic repository, we add timestamp and geocode data to the fragmented sentences, extracted from the first two analyses.

4.3 Visualizing and Exploring Collective Culture

After the procedures above, we make instances of the StoryFragment class from the fragmented sentences with spatiotemporal information. The instances of Storyline class are generated based on the contextual information of the fragmented sentences. Fig. 6 shows a result of plotting those instances on KC.

We could get the narrative data from three research participants in this test case. Since some stories in their narratives were referring to a similar time and place, KC's OPP detector could automatically find the intersection point of their storylines in the virtual 3D space (see Fig. 7). With the assist of the OPP detector, we could know a focal point to start exploring the sharable experiences in the community, and get a clue to understand the indirect collective culture regarding movie industries in Kyoto.

In this test case, KC found one OPP where three storylines from three different research participants went through a similar time and place in the virtual 3D

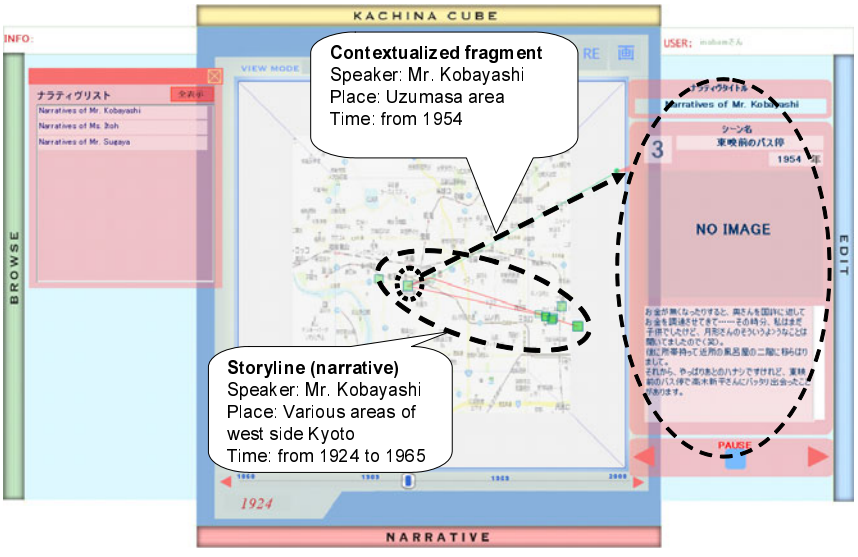


Fig. 6. Fragmented sentences and storylines plotted in the virtual 3D space of KC

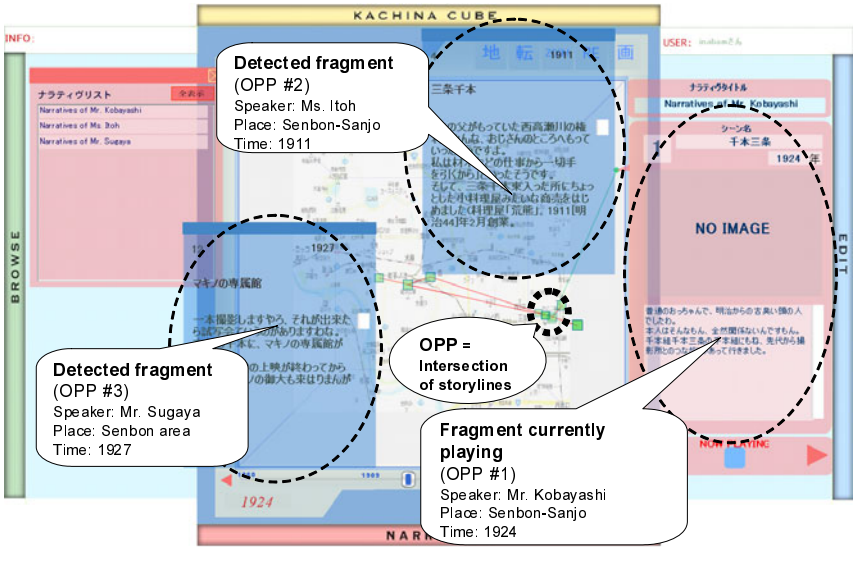


Fig. 7. OPP detector finds an intersection among storylines from three research participants

space. The point indicates the Senbon-Sanjo area in the central Kyoto from the 1911s to the 1927s. According to the storyFragments provided by research participants, there used to be some movie theaters and an enterprise related with movie industry in that area. Since there are no movie theaters or movie-related firms in the area at the present day, the place and time can be considered as a crucial point to explore or restore collective culture regarding the once-flourishing film industry in Kyoto.

5 Related Works

Our approach is closely related to that of GIS-based narrative analysis. For example, Kwan et al. [5] develop the method called "geo-narrative," which extends current GIS capabilities to analyze and interpret narrative materials such as oral histories, life histories, and biographies. In this method, researchers are supposed to go over narratives line by line, interpret sentences, and then plot them onto GIS space. Although it provides a precise way to examine narrative materials, its process of data plotting cannot exclude the researchers' subjective interpretation.

Other researchers are exploring the possibilities of automatic geocoding from text data. For example, Yao et al. (see [13], [14], and [15]) have been working on the fundamental research on "qualitative location (QL)" which refers to a spatial location using linguistic terms such as qualitative spatial descriptions. Their research proposes a way to process multiplicity and uncertainty in QL described in natural language. Angel et al. [1] also advocate a methodology for the semiautomatic geocoding of persistent Web pages in the form of collaborative human intervention to improve on automatic geocoding results. While their techniques are universally applicable, their approaches cannot handle various proper nouns.

CulturalSampo [3] is an integrated virtual museum with the combination of semantic web technology and web-based geographical system. It allows its user to browse cultural artifacts in a specific region in a certain time. The target domain of this system, however, is the public history or cultural artifacts which are regularly exhibited in real or virtual museums.

Different from the researches above, our approach uniquely combines text mining and semantic web technology to extract spatiotemporal information from the narrative data, and helps the researcher understand personal and collective cultures with the visualization of narratives in the virtual 3D space.

6 Conclusion

Taking Valsiner's socio-cultural definition of culture, this paper focused on spatiotemporal information to capture personal and collective cultures, since such information regularly appears in people's life stories. To analyze the culture, we propose a method to extract personal culture from their stories.

We analyzed oral histories, narrative data from three research participants in Kyoto's Rakusai area, and stored extracted spatiotemporal information to

KC in order to examine our method. As a result, KC demonstrates significant potentials for study of regional culture, based on spatiotemporal information. Reviewing similar researches in Cultural Computing, we evaluate our original method to tackle extracted culture.

As the next step of our research, we have to work on three major challenges. First, we have to enhance our user dictionary, pattern-phrase dictionary, and domain-topic repository to handle the wide range of stories. Since these databases are reusable for researchers interested in similar areas and times, collaborative development and sharing of such databases could contribute to enhance a variety of socio-cultural research activities. Secondly, we still have trouble dealing with ambiguous and fuzzy spatiotemporal information. As GIS researchers are experiencing similar problems, we should collaborate with each other to develop techniques or corpus to handle such ill-defined spatiotemporal information. Finally, we need to apply our method and platform to more cases for enhancing generalizability of our way on mining and visualizing collective culture.

Acknowledgments. The authors would like to thank Prof. Mika Tomita and her colleagues who provided the narrative data for our research. This research has been partially supported by Digital Humanities Center for Japanese Arts and Cultures, Ritsumeikan University.

References

1. Angel, A., Lontou, C., Pfoser, D., Efentakis, A.: Qualitative geocoding of persistent web pages. In: Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems, Irvine, California (2008)
2. Arpinar, I.B., Sheth, A., Ramakrishnan, C., Usery, E.L., Azami, M., Kwan, M.: Geospatial ontology development and semantic analytics. Transactions in GIS 10, 551–575 (2005)
3. Hyvönen, E., Mäkelä, E., Kauppinen, T., Alm, O., Kurki, J., Ruotsalo, T., Seppälä, K., Takala, J., Puputti, K., Kuittinen, H., Viljanen, K., Tuominen, J., Palonen, T., Frosterus, M., Sinkkilä, R., Paakkarinen, P., Laitio, J., Nyberg, K.: CultureSampo – Finnish Culture on the Semantic Web 2.0: Thematic Perspectives for the End-user. In: Museums and the Web 2009, Indianapolis, USA (2009)
4. Kudo, T.: MeCab: Yet Another Part-of-Speech and Morphological Analyzer, http://mecab.sourceforge.net/
5. Kwan, M.-P., Ding, G.: Geo-Narrative: Extending Geographic Information Systems for Narrative Analysis in Qualitative and Mixed-Method Research. The Professional Geographer 60, 443–465 (2008)
6. Tomita, M., Itakura, F.: Voices from Kyoto: Interview with ITO Asako – An aspect of Japanese film history. Art Research 1, 127–138 (1999)
7. Tomita, M., Kamiya, M., Gondo, C.: Voices from Kyoto: Interview with KOBAYASHI Masanori – An aspect of Japanese film history. Art Research 2, 115–136 (2000)
8. Tomita, M., Kamiya, M., Oya, A.: Voices from Kyoto: Interview with SUGAIE Koyo - An aspect of Japanese film history. Art Research 3, 241–270 (2001)

9. Tosa, N., Matsuoka, S., Ellis, B., Ueda, H., Nakatsu, R.: Cultural Computing with Context-Aware Application: ZENetic Computer. In: Kishino, F., Kitamura, Y., Kato, H., Nagata, N. (eds.) ICEC 2005. LNCS, vol. 3711, pp. 13–23. Springer, Heidelberg (2005)
10. Valsiner, J.: Personal Culture And Conduct Of Value. Journal of Social, Evolutionary, and Cultural Psychology 1, 59–65 (2007)
11. Valsiner, J., Sato, T. (eds.): Historically Structured Sampling (HSS): How can psychology's methodology become tuned in to the reality of the historical nature of cultural psychology?. Transcript Verlag, Bielefeld (2006)
12. The World Wide Web Consortium: OWL Web Ontology Language Overview, http://www.w3.org/TR/owl2-overview/
13. Yao, X., Jiang, B.: Visualization of qualitative locations in geographic information systems. Cartographic and Geographic Information Science 32, 219–229 (2005)
14. Yao, X., Thill, J.-C.: Spatial queries with qualitative locations in spatial information systems. Computers, Environment and Urban Systems 30, 485–502 (2006)
15. Yao, X., Thill, J.-C.: Neuro-fuzzy modeling of contextcontingent proximity relations. Geographical Analysis 39, 169–194 (2007)

A Web Strategy for Cultural Inheritance Centered on Agriculture Case Study Approach: The Olive Project in Shodoshima Japan

Takashi Togami[1], Yoshitaka Motonaga[2], Ryoei Ito[1], Atsushi Hashimoto[1], Takaharu Kameoka[1], and Tsuyoshi Nakamoto[3]

[1] Graduate School of Bioresources, Mie University,
1577 Kurimamachiya-cho, Tsu City, Mie Prefecture
514-8507, Japan
kameoka@mie-u.ac.jp
[2] Department of Production and Environment Science, Faculty of Agriculture,
Niigata University
[3] AEON RETAIL Co.,Ltd, Food & Delicatessen Merchandising Div, Food Artisan Group

Abstract. The olive culture in Shodoshima has the 100 years history, yet faces to the crisis when looking at the century ahead. In modern society, it is assumed that there has been a strong link between food culture and two kinds of aspects which consist quality necessary for cultural succession. One is the promotion of target products and the consideration of consumer behavior and quality requirements from consumers. The other is the creation of cultivation recipe which clarifies the method and component necessary for the production of products fulfilling the required quality by consumers. Therefore, we attempted to construct the strategic website for consumer driven food culture extension and to make cultivation recipe by the installation of Field Server in agricultural field and its data utilization. This paper shows the case that adopted the ICT in both consumer-led promotion and agricultural production for passing food culture down the generations.

Keywords: Consumer-led promotion, cultivation recipe, Field Server, olive, food-culture, Shodoshima.

1 Introduction

In Meiji era, Japanese government introduced cultivation of foreign vegetables to Japanese agriculture in order to cope with the increase in food demands resulting from the population growth. In other words, they began assimilating foreign food culture and merging it with Japanese food culture. Olive was first introduced at the time. Foreign vegetables such as cabbage, tomato and onion were well adapted. However, olive was not since olive cultivation purpose was unclear. It was just one part of government policies that nebulously introduce fruits, vegetables and grain from leading countries and adopt crops, which may be helpful for Japanese agriculture, among the options. Hence, olive cultivation could not be focused.

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 200–213, 2010.

After Russo-Japanese War in 1904 to 1905, Japan conquered broad fishing grounds at northern region and huge amount of fish haul of fish and seafood became available. As for the avenue available for storing and transporting the fish and seafood, marinating was adopted. And so, full-scale olive cultivation was implemented under the government in order to fulfill the oil demands. The cultivation was attempted in Mie-Pref., Kagawa-Pref. and Kagoshima-Pref., but only olives grown in Shodoshima Kagawa-Pref. remained due to the environmental conditions and so on.

Thus, the olive cultivation in Shodoshima is the symbolic representation that brings down the changes in Japanese food culture, and also the memorial token that represents a fusion of traditional Japanese and foreign food cultures. Therefore, the inheritance of unique culture in Shodoshima, which is the olive cultivation going on for 100 years is very meaningful and absolutely important.

However, there are tasks that should be implemented for olive culture succession when looking at the century ahead. Figure 1 shows the conceptual diagram for inheritance of food culture. At first, information on Shodoshima and Shodoshima olive should be spread out over a lot of people. For food culture extension, consumer and their demands are important parameters to be taken into account since consumer generated media (CGM) such as blogs and BBS has positive influence on other consumer's behavior. In addition, products should meet the strong quality requirements from consumers in order to popularize the product even when the time changes. Along with that, production of primary products with the required quality becomes necessary, yet there are issues in agriculture. For instance, the persons engaged in agriculture decreased rapidly from 14.54 million to 2.98 million in past 50 years [1], and there were over 60 % of the persons aged over 65 in 2008 [1]. Moreover, there are a lot of quality requirements and regulations that should be satisfied shown in Figure 2, and primary food production has become difficult due to unstable and varying climate. Therefore, cultivation recipe that clarifies methods and components for primary products with required quality level is significant.

However, the question remains: how to satisfy these requirements? The development of modern technology in recent years enables to answer the question, and that is the introduction of information and communication technology (ICT) to each field. The construction of a website assists the extension since the number of internet users has been increasing [2], and the introduction of an IT device in cultivation enables the collecting essential data for the production. As just described, the introduction of ICT makes different aspects overlap each other and creates common gateway leading to the solution. In fact, ICT has been introduced into a lot of different fields such as culture and tourism [3] [4], education [5] and agriculture.

In 2008, Shodoshima marked its 100th anniversary for olive grass establishment and the "Olive Food Industry Board" was established for the purpose of the regional improvement of Shodoshima centered on olives. Mie University Japan helped establish the board with AEON RETAIL Co., Ltd. and they have implemented the supportive approach for the inheritance of regional food-culture and the cultivation by the use of IT device.

Therefore, the aims of this study were the construction of website enabling consumer driven food culture extension and making olive cultivation recipe for cultural succession necessary for relaying olive culture to future generations.

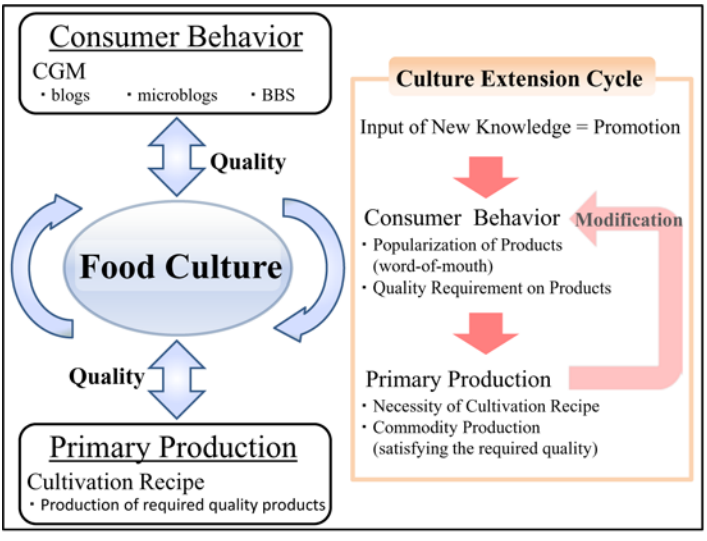


Fig. 1. Conceptual Diagram for Food-Culture Inheritance

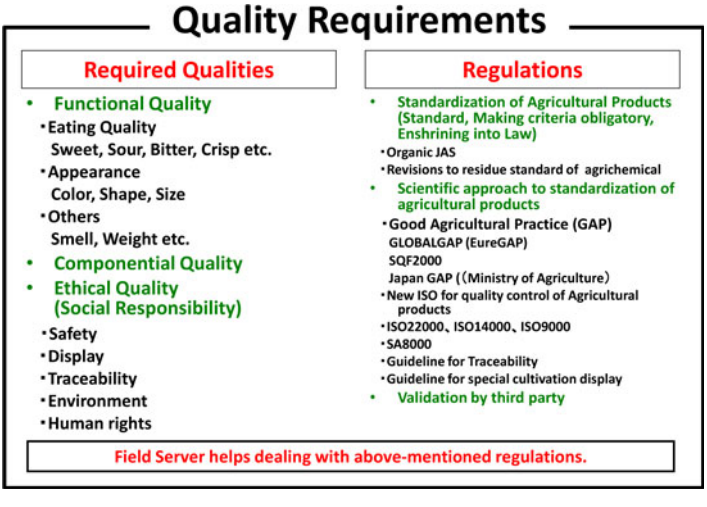


Fig. 2. Quality requirements for food production in modern society

2 ICT Application for Olive Culture Promotion

There is a fact that inheritance of olive culture was carried out since olive has been cultivated for 100 years in Shodoshima. However, in the world changing from day to day, cultural succession cannot be expected when looking ahead to the next 100 years.

For instance, olive growing area and the amount of crop decreased rapidly in past 50 years [6] [7]. Therefore, It is absolutely necessary to consider "quality" consisting of two aspects for inheritance of food culture. One is the quality of products required by consumers. The consumption behavior of modern customers can be expressed as AIDEES, which is the abbreviation for Attention Interest Desire Experience Enthusiasm Share, and the adoption of Consumer Generated Media (CGM) such as blogs and BBS when considering popularization and edification of products as a part of conserving culture is indispensable for the system that cognizes the customer demanding quality varying across the ages. Moreover, the structure that stimulates their shopping behavior by correct understandings and knowledge providing through education and experience will also be requisite.

The other is the satisfaction of the strong quality requirement on primary products. Today, when issues of food mislabeling and residual pesticide are often on mass media, there have been increasing in the number of integrants making up the quality that fulfill the requirements by modern society on primary products. Moreover, even the cultivation of crop itself is becoming difficult due to the unstable climatic change. Therefore, the informatization of the integrants and method for primary products satisfying the required quality level is essential. Furthermore, olive and olive cultivation in Shodoshima will not disappear if passing the integrants and methods down the generations.

Just like CGM mentioned earlier as the example, the informatization, information conservation and management of quality consisted of above-mentioned two aspects are becoming impossible without using ICT in an increasingly complex world. Therefore, as the tactics for relaying olive culture to next 100 years, we attempted to construct the structures fulfilling the quality required by each aspect by website and ICT devices.

2.1 Consumer Driven Food Culture Extension

Regarding the popularization of the olive and its food culture, the consumer behavior and the perception of food quality are major parameters to be taken into account [8]. On top of this, the consideration of a marketing strategy is imperative. With the advent of the internet era, internet communication tools such as blogs, microblogs such as Twitter, social networking service (SNS) and bulletin board system (BBS) have appeared and AIDEES that is the marketing model to adequately apprehend consumption behavior of consumers in modern society was advocated in 2006. This model refers to the process of the behavior that customers detect products, have interests, conceive buying motivation, purchase and experience, become wildly enthusiastic fans, and encourage others through their experiences. And the recent scientific research on the social relationship between word-of-mouth (WOM) senders and receivers showed that the information transmitted by customers has positive influence on consumption behavior of other customers [9]. Therefore, providing CGM, which is the website becoming the new media by compiling a database of transmitted information, makes a huge contribution to acquire consumer needs and prevalence of food culture. Additionally, nowadays, when everyone is able to provide information on the web, provision of precise and reliable information and knowledge on food and food culture, and the education to younger generations are critical prerequisites.

Therefore, the webpage development to construct a strategic website fulfilling the requirements as previously stated was implemented. In this regard, following four points were considered. The points are; (1) website design for olive extension, (2) the selection of target website audience and human interface, (3) the approaches to Web2.0 and (4) the linkage between the website and the database.

2.2 Making Olive Cultivation Recipe for Cultural Succession

As for passing olives down the generations in Shodoshima, what becomes important and should be considered is “quality” of products. Nowadays, there are a lot of integrants that makes product’s quality as is shown in Figure 2. Each integrants should be fulfilled. Otherwise, products are replaced by foreign products of the same kind and therefore the culture would be lost. Thus, the satisfaction of demanded quality in the modern world is the key. However, questions are rising when considering the demanded quality. How to do it? What can be the tool? This is when ICT comes on stage. The powers of ICT are informatization, information conservation and management, and so on. Therefore, the informatization of the components and methods for producing the products that satisfy the demanded quality and passing on the components and methods to the next generation by the beneficial use of ICT are the possible answer. To put it plainly, it means the development of "cultivation recipe", which is similar to recipes in cooking. Thus, the development of cultivation recipes and passing it down from generation to generation will be the essence for cultural succession, and the acquisition and the improvement of required quality level of primary products lead to the improvement of processed foods and to the contribution of inheritance of food culture. In addition, under increasingly complex conditions and in the world, achievement of the goal cannot be expected without using ICT.

Approach to Cultivation Recipe. As for developing a cultivation recipe, it should deal with variety of requisites as is mentioned, and they are the stable production under unstable and varying climate such as global warming, floods and droughts, sustainability which means it has low impacts on environment and energy consumption, cultivation of high quality and functionality products which contain high nutrition and good taste, and food safety and reliability. This means that the paradigm shift from maximization implemented in 20^{th} century, which is the non-sustainable agriculture based on chemistry and engineering, to the optimization is indispensable for the cultivation recipe and cultural succession.

With regard to the optimization, ICT helps the optimization in many aspects. First of all, it enables cost reduction and competitive agriculture. It means optimal farm planning and efficient management of large number of fields are possible. Secondly, robust and stable farm production under extreme weather and global warming can be implemented in next 100 years. Next, sustainable agriculture by optimal agrochemical application is also possible. Fourthly, food safety and reliability by monitoring cultivation and tracing farming records can be gained. Lastly, visualization of quality by image analysis and data analysis will make it possible to produce high quality products. However, in order to achieve the goal mentioned above, four approaches should be carried out. One is that data collection to know what is happening in each

field quantitatively is imperative. Second one is the efficient knowledge transfer. Most of often, long term experience and intuition are required for cultivation. Therefore, quantifying invisible empirical knowledge and transferring tacit knowledge to explicit knowledge are absolutely necessary. And also the case base reasoning is significant. Third one is the optimization and risk management to support making decision during cultivation based on acquired data and knowledge. Last one is the framework to support decision making. Thus, in agricultural field, the construction of ICT-aided cultivation support system that enables collecting local data becomes important.

Field Server for Environmental Information Monitoring. In relation to the construction, there is a strategic device called Field Server (FS). Figure 3 shows a Field Server developed for a long-term monitoring in a field [10] [11] [12] [13]. As it can be seen in the figure, various kinds of data such as ambient temperature, relative humidity, amount of insolation and soil moisture can be acquired as well as growth situation imageries by its onboard network camera. In addition, other environmental data such as carbon dioxide can also be acquired if the sensor is attached to the Field Server. Those data can be collected from remote locations by the use of wireless LAN.

In this study, we have considered the utilization of data acquiring at two locations taking modern-day farm conditions into account. One is the internet-enabled site, which is online site. The other is where internet connection is not available, which is offline site. The types of data we have acquired includes ambient temperature, relative humidity, amount of solar radiation, soil temperature and soil moisture as growth situation data as well as images focusing on olive tree, fruits and trunk near ground. Figure 4 shows the schematic diagram of field monitoring system introduced in Shodoshima. As is mentioned above, data can be collected via internet at online site. However at offline site, data is stored at the data storage of the Field Server. And then, it is manually collected and transported regularly to the data storage at online site. After that, it is collected via internet to the data server at the computer center for

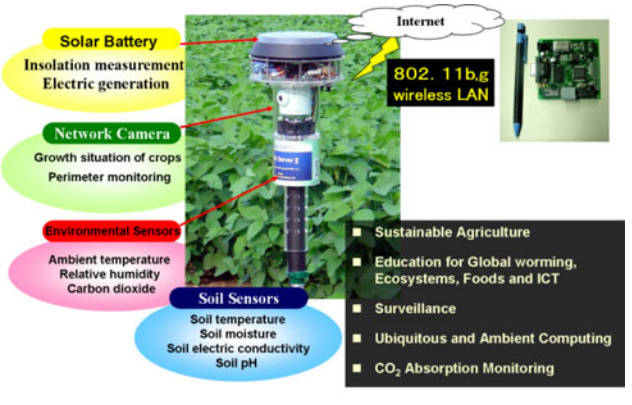


Fig. 3. Field Server and the acquirable data

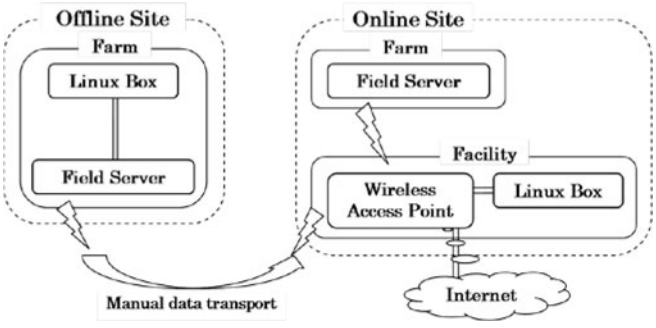


Fig. 4. The schematic diagram of field monitoring system introduced in Shodoshima (Motonaga et al. 2008)

Agriculture, Forestry and Fisheries Research. Finally, it is stored and managed by the data server at the computer center. Motonaga et al. [14] constructed the above-described storage and management system. Therefore, the strategies of Field Server for this study consist of two parts as the part of developing the cultivation recipe. One is "the measurement and control by Field Server software on the web", and the other is "Image utilization and application on Field Server images". In addition, in the case when immediacy is necessary, the data at the online site was used. If it is not necessary, the data at the computer center was utilized.

3 Results and Discussion

3.1 Consumer Driven Food Culture Extension

The construction of webpage for consumer driven food culture extension was implemented with the consideration of four points as is mentioned. In this section, results are presented and current problems and tasks will be discussed.

Website design for olive extension. In order to spread out information and provide precise and reliable knowledge on Shodoshima and their olive culture, the enhancement in both amount and quality of information should be considered to gain more of the user attention. Therefore, the project website was constructed with six contents that are project summary, Agriculture and Environment, food and culture, tourism, Education and English, and most of the sources of information were cited as proof. In food and culture section, we have prepared cultural contents of Shodoshima olives such as history, tradition and centenary celebration being held, and contents relating to olives as cuisine culture such as olive oil, health effect of olive oil, and a broacher of unique cooking using olive oil that consumers are able to meet only at Shodoshima.

The selection of target website audience & human interface. It is significant to consider the selection of intended website audience and human interface in order to convey information easier, make more users to visit the website and preserving the

traditions and cultures of Shodoshima olives. Therefore, farmers, children and general users were selected as the target according to our project objective.

With regards to human interface, reconsideration of information expression was implemented. For instance, the display of numerical data from sensors for farmers was redesigned. The redesignation includes the translation of language from English to Japanese since a lot of farmers cannot understand English, and the elimination of data which does not mean anything to farmers in terms of avoiding any confusion. In addition, various kinds of information expression were adopted such as movies and a quiz game created by macromedia flash. Figure 5 shows the example of information expression. Left hand side of the figure is the flash content that enables give information visually to users and it presents the footstep of Shodoshima olive in the last 100 years. Right hand side of the figure shows one of the educational contents created for children. It is necessary to consider the way of gaining parents and children's attentions and the way of providing enjoyable contents. Therefore, we have created the quiz game, flash contents presenting the process from olives to olive oil and growing stage of olive.

Fig. 5. The example of information expressions. (a) The flash content presenting the footstep of Shodoshima olive in the past 100 years, (b) The quiz game for children as a part of education.

The approaches to Web2.0. Web2.0 [15] refers to the new generation of web based services and communities characterised by participation, collaboration and sharing information among users online such as YouTube, Amazon, Twitter, blogs and BBS, and social networking services[16]. It can be said that the benefit of the adoption of Web2.0 is not only for upvaluing the website, but also for providing a place to internet users for intelligence sharing and information exchange. Therefore, the applications such as blogs and BBS representing web2.0 are absolutely needed as the web and marketing strategies leading to acquiring consumer's needs for the required quality, the popularization of products, and to improve and assist creating a cultivation recipe, and so to the succession of olive culture. However, in this study, only two approaches were carried out. One is mobile phone contents for farmers that they can send a text and an image to the web and alert system if there is a problem at working site was adopted. The other is the map contents. For general users and

children, some map contents made by the use of Google map were constructed in order to give users rich experience of information handling, to spread out the location of Shodoshima, and to educate children. Thus, these approaches showed advancement of project as the first step, yet more are necessary.

The linkage between the website and the database. With regard to the informatization, it is obvious that the construction of data base, the information storage and management system, and the collaboration between stored information and the latest information on the web will make it possible to create and provide new media to customers and will be very effective from the standpoint of consumer. From the perspective on cultivation on the other hand, databases and the collaboration will contribute to pinpoint the components and the method necessary for the production of products with required quality. Therefore in this study, the precious contents created for the 100th anniversary as the footmark for next 100 years were stored at a server. With regard to the cultivation, data acquired from Field Servers were accumulated and stored at a data server. And also, we have arranged the structure that enables to store the texts and images transmitted by on-site farmers by mobile phones.

Towards the application to CGM & WOM. With regards to the popularization of olive culture, the website was introduced by some mass media [17] and it has been spread via links and blogs on the outside of our website. However, it has become clear that there are urgent and crucial concerns. That is the lack of user generated contents (UGC) such as blogs and BBS. The installation of blogs or BBS enables information transmission, exchange, sharing, and most importantly correction by their feedback function. In other words, it enables information control and management. Moreover, it is also possible to acquire consumer demands. However, on our current version of website, there is no way to correct or stop information even if the information that has bad influence on the extension of food was put on other websites. Therefore, more of the approaches for providing CGM are urgent needs and must. The approaches include the adoption of UGC in order to gain more of the user access, to transmit correct information, and to provide a virtual place for better information exchange and intelligence sharing to users regarding to the acquisition of consumer demands, the satisfaction of required quality and the improvement of products. In addition, a database is necessary for CGM. Therefore, the construction of database should be carried out in parallel with the CGM installation.

3.2 Making Olive Cultivation Recipe for Cultural Succession

The development of the web application enabling the measurement and control as well as utilization and application of images acquired by Field Servers were carried out to support fulfilling the strong quality requirements of primary products. In this section, the results will be presented with the related figures and discussed various issues and tasks for creating cultivation recipe.

The measurement and control by Field Server software on the web. Figure 6(a) shows the webpage displaying the latest images and data acquired from sensors. This page was designed to monitor an agricultural field and an olive tree. Therefore, the access to the data storage at online site was carried out due to the necessity of the

immediacy. As it can be seen in the figure, the thumbnail of images configured on the network camera is displayed. When clicking the buttons just under the small images, larger image can be displayed at the field next to the thumbnail. In addition, when clicking a small image by itself, much larger image is displayed on a new window of web browser.

The graph of data acquired from sensors is shown in Figure 6(b). As it can be seen in the figure, the ComboBox that enables users select year and month, and day, was adopted to refer past data. In addition, for farmers to compare their selecting data with Shodoshima's past mean temperature, the mean temperature of selected month can be automatically displayed. For the mean temperature, the data from 1979 to 2000 were used.

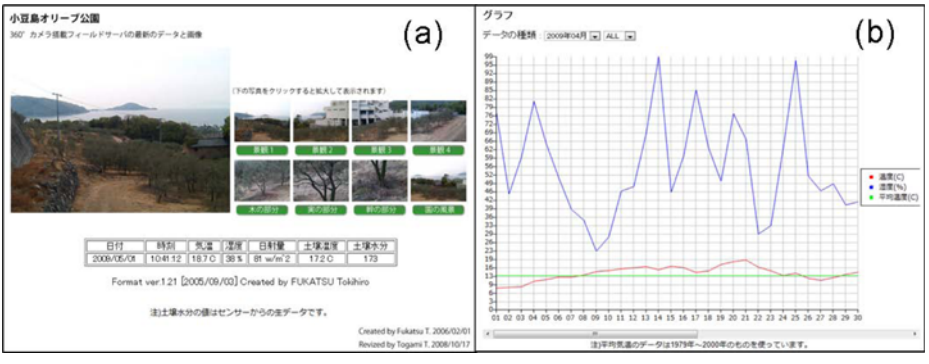


Fig. 6. (a) The web page displaying the latest images and data acquired from sensors. (b) The graphs of data acquired from sensors.

We have arranged the past numerical data for farmers to draw upon the data for decision making in cultivation management shown in Figure 7(a). From the information utilization point of view, we have designed the page that allows users download their selecting data in CSV format so that users can use the data on other application software and analyze. The reason for the adoption of the data format is because it has broad utility, which means a lot of software supports the format.

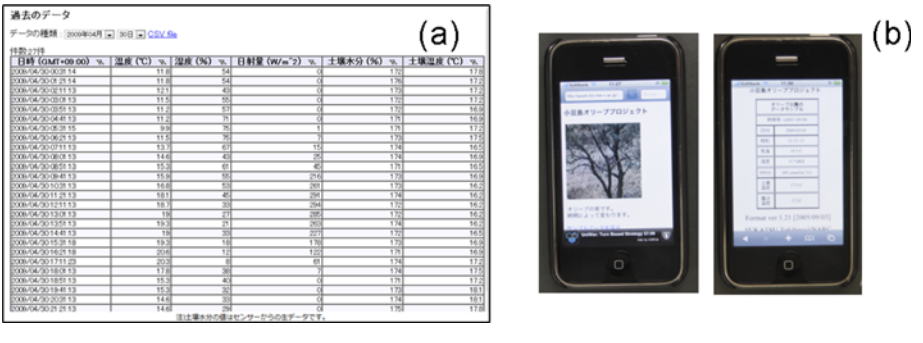


Fig. 7. (a) The past data reference and download in CSV format. (b) The latest image and sensed data displayed on a mobile phone.

The latest image and sensed data could successfully be displayed on a mobile phone shown in Figure 7(b). The posting and mailing system were also adopted so that users can send a text and an image to the web while working at a field and they can send an alert to all the registered users if there is a critical situation. Therefore, it can be said that the construction of mobile contents, posting and mailing system have created more flexibilities in monitoring and cultivation management for farmers.

Image utilization and application on Field Server images. Figure 8 shows the conceptual diagram for the utilization of Field Server images. As is shown, by the figure, the images can be utilized by three ways that are the image combination, archives, and real-time images. In this study, the image utilization by the image combination and archives has been attempted.

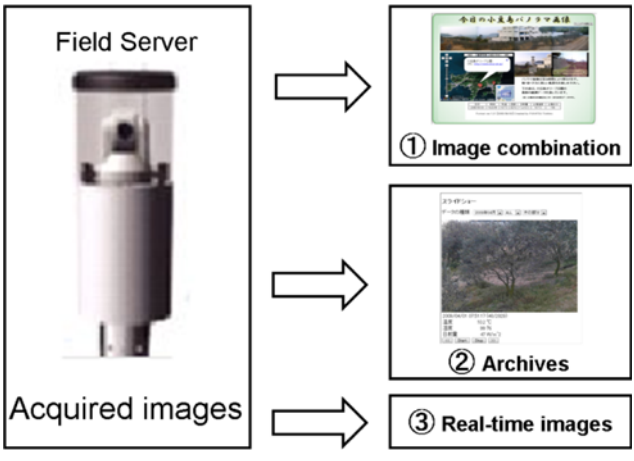


Fig. 8. The conceptual diagram: The utilization of FS images

The example of the image utilization by an image combination is shown in Figure 9(a). The panoramic image on the top of the figure was created by the combination of acquired images from the support system. In this study, interesting approach to urge user's understandings of significance of scientific cultivation technology was implemented. The images have been used not only for the purpose of monitoring an agricultural field and olives, but also for providing an opportunity to users to meet the data acquired by the system. And so the map and sensed data were adopted on the page.

Figure 9(b) and 9(c) are the examples of image utilization by archives. The images were used for the reference of past growth situation of an olive tree shown in the Figure 9(b). As for the reference and cultivation support, the field information such as ambient temperature, relative humidity and amount of solar radiation were displayed with the image. On the other hand, stored images were put in chronological order and used for the creation of the educational content that presents the growing stages of an olive.

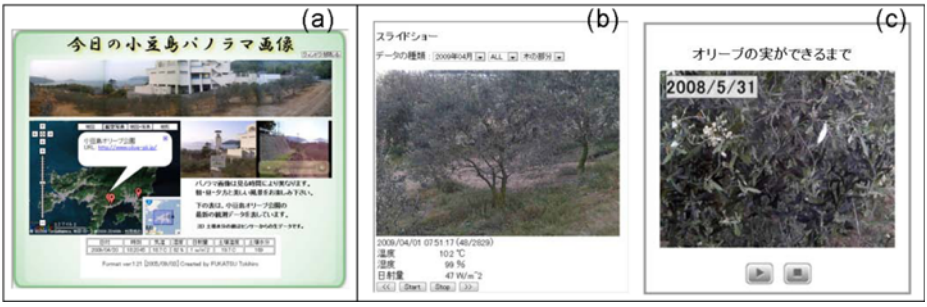


Fig. 9. (a) The example of an image combination (Panoramic image). (b) The example of image utilization by archives: the reference of past growth situation. (c) The image utilization by archives: the educational content.

For the creation of cultivation recipe. It became clear that there had been data precision problems on Field Server when developing the web application. Hence, only data acquired by sensors were put on the Web. However the data acquisition and utilization of information related to agricultural crops and growth environment information from a cultivation perspective are very important to keep a stable cultivation and obtain the quality of products, and also to improve the productivity. For that reason, the data preparation for data analysis on cultivation is absolutely necessary and the data includes accumulated temperature, accumulated amount of solar radiation, amount of rainfall and mean temperature. In addition to the data, the stabilization of data precision by the parallel use of our cultivation support system and Automated Meteorological Data Acquisition System will also be necessary. Moreover, web application that equip the mathematical model for the estimation of growing stages such as the time of bloom based on cultivation index are very much needed at cultivation sites, and it should be user-friendly especially for farmers to handle data for operation.

With regards to image utilization and application on Field Server images, it has been affirmed that images are valuable as information. However, it has become clear that the current system has difficulties in data handling of imageries such as processing for secondary contents and sorting out images. In addition**,** website design that allows users easier operation should be reconsidered since it has still some difficulties for farmers to use the website for decision making during cultivation.

Thus, web application that possesses cultivation recipe and function of guidance visually for better cultivation will be necessary. This means “Guideware” is necessary. In addition, Cultivation guidance and registration that specify cultivation period, cultivation method, picking season, nutrition and health effect will be necessary for consumers.

4 Conclusion

In this study, we showed the case study approach that adopts ICT for the promotion of food culture extension, and the creation of cultivation recipe in agriculture. We

attempted to construct the strategic and integrative website along with the adoption of CGM, but the approaches to CGM is the action assignment and urgent. Furthermore, the construction of databases is imperative in regards to the provision of CGM. In relation to CGM and databases, they should be adopted in agricultural sector to satisfy and improve the quality required on primary products, and to bring the cultivation recipe mentioned in earlier section to the next generation. From making the recipe point of view, the utilization of data acquired by Field Servers was implemented as the first step. In the future, the preparation of the data acquisition system with the cultivation perspective and data which will become major factors of the recipe such as accumulated temperature and accumulated amount of solar radiation are must. In addition, more of the efforts on quantifying invisible empirical knowledge and transferring tacit knowledge to explicit knowledge are very much needed since the aging of the persons engaged in agriculture has been becoming even more serious issue. At that time, the development of a faked object that can be criterial by touching the object will be necessary since meaningful data acquisition by ICT in everything is almost impossible. In other words, apprehending the human as a sensor and implementation of the informatization will be significant.

As the next step, the website with blogs will definitely need to be merged with cultivation recipe and made them into one and small packet as the guideware mentioned in earlier section in order to satisfy strong quality requirements and better cultivation.

Finally, we focused on a virtual experience in this case study, yet an actual experience for consumers such as farming experience for primary productions and food processing for secondary products will be necessary. It means putting consumers through farm works such as pruning and ingathering throughout a period of time, and through food processing such as olive extraction. This will surely generate a synergistic effect between virtual and actual world and more of the outcome of food culture extension can be expected.

The URL for the official website for Shodoshima olive project is shown below.
http://www.quark.bio.mie-u.ac.jp/olive/en/index_en.html

Acknowledgement. We would like to express our sincere gratitude to Dr. Masayuki HIRAFUJI, Dr. Tokihiro FUKATSU and Mr. Takuji KIURA who have contributed to this project.

References

1. Ministry of Agriculture, Forestry and Fisheries: The Statistics for Farm Labor Force, http://www.maff.go.jp/j/tokei/sihyo/data/08/html (accessed February 3, 2010) (in Japanese)
2. Ministry of Internal Affairs and Communications: Information & Communications Statistics Database, http://www.soumu.go.jp/johotsusintokei/new/ (accessed February 9, 2010) (in Japanese)
3. Fotakis, T., Economides, A.A.: Art, Science/Technology and History Museums on the Web. Int. J. Digital Culture and Electronic Tourism 1, 37–63 (2008)
4. Noel, L., Carloni, O., Moreau, N., Weiser, S.: Designing a Knowledge-based Tourism Information System. Int. J. Digital Culture and Electronic Tourism 1, 1–17 (2008)

5. Tondeur, J., Braak, J.V., Valcke, M.: Curricula and the Use of IT in Education: Two Worlds Apart? British Journal of Educational Technology 38, 962–976 (2007)
6. Shodoshima Town Office: About Shodoshima Town and Olive, http://www.town.shodoshima.lg.jp/olive_station/olive/index.html (accessed Februray 14, 2010) (in Japanese)
7. Shodoshima Town Kagawa Pref.: The official text for Shodoshima Olive Certificate, Shodoshima Kagawa Pref., Kagawa Pref. (2008) (in Japenese)
8. Cayot, N.: Sensory Quality of Traditional Foods. Food Chemistry 101, 154–162 (2007)
9. Yamamoto, H., Matsumura, N.: The Power of Grassroots Influentials: The Optimal Heterophily between Sender and Receiver. In: 2009 International Conference on Computational Science and Engineering, pp. 456–462. IEEE Press, Vancouver (2009)
10. Fukatsu, T., Hirafuji, M.: Development of Field Servers for a Field Monitoring System. Agricultural Information Research 12, 1–12 (2003) (in Japanese)
11. Fukatsu, T., Hirafuji, M.: Field Monitoring Using Sensor-Nodes with a Web Server. Journal of Robotics and Mechatronics 17, 164–172 (2005)
12. Fukatsu, T., Hirafuji, M., Kiura, T.: An Agent System for Operating Web-based Sensor Nodes via the Internet. Journal of Robotics and Mechatronics 18, 186–194 (2006)
13. Fukatsu, T., Hirafuji, M., Kiura, T.: A Distributed Agent System for Managing a Web-based Sensor Network with Field Servers. In: 4th World Congress on Computers in Agriculture (2006)
14. Motonaga, Y., Ito, R., Fukatsu, T., Hashimoto, A., Ninomiya, S., Kameoka, T., Nakamoto, T.: Monitoring the Olive Cultivation in Shodoshima using Field Server Systems. In: World Conference on Agricultural Information and IT (2008)
15. O'Reilly, T.: What is Web2.0 Design Pattern and Business Models for the Next Generation of Software, http://oreilly.com/web2/archive/what-is-web-20.html (accessed February 8, 2010)
16. IT Words and Terms Dictionary: Web2.0, http://www.sophia-it.com/content/Web+2.0 (accessed February 8, 2010) (in Japanese)
17. The Sanyo Shinbun: Regional NEWS, http://www.sanyo.oni.co.jp/sanyonews/2009/03/18/200903181140 0393019.html (accessed July 18, 2009) (in Japanese)

Area Informatics
– Concept and Status –

Shoichiro Hara

Center for Integrated Area Studies, Kyoto University
46 Shimoadachi-cho Yoshida, Sakyo-ku, Kyoto 606-8501 Japan
shara@cias.kyoto-u.ac.jp

Abstract. Area studies are interdisciplinary science of humanities, natural and technological researches. Area informatics is a new information paradigm in area studies to integrate and analyze data from variety disciplines quantitatively and objectively. Spatiotemporal data gives a few quantitative attributes that are relatively easy to be derived from area studies' data. The Humanities GIS research group (H-GIS) is a leading joint research group in Japan to apply spatiotemporal informatics to area studies. This paper sums up current situations on area informatics and introduces the H-GIS outcomes about applying spatiotemporal informatics to area studies.

Keywords: area informatics, area studies, spatiotemporal tools, digital gazetteers, HuMap, HuTime, H-GIS.

1 Introduction

Area studies are interdisciplinary science to investigate and understand particular areas comprehensively. Natural and technological sciences such as meteorology, biology, ecology, forestry, hydrology, pedology, medicine, and disaster prevention are the major disciplines of area studies, and furthermore, humanities such as anthropology, ethnology, folklore, history, politics, linguistics, and sociology also play essential rolls in area studies. Unfortunately, the concept of comprehensive investigation and understanding of area studies is hardly realized.

Humanities generally use qualitative data and process them interpretively and subjectively, but on the other hand, natural and technological studies commonly use quantitative data and process them numerically and objectively. However, a paucity of appropriate frameworks to use and process qualitative data together with quantitative data makes it difficult to organize a huge volume of humanities' resources into databases and to utilize them for visualizing, comparing, calculating, and analyzing objectively and logically. This is a reason of area studies failing in constructing comprehensive images of particular areas.

Informatics is science to analyze and understand everything quantitatively and objectively by introducing distinctive norms and methods, which develops many algorithms and means to quantify and classify qualitative data. Area informatics is a new information paradigm in area studies to make a breakthrough in constructing

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 214–228, 2010.

comprehensive images of particular areas. Area informatics rather focuses on humanities' side of area studies: by applying informatics methods, area informatics tries to build frameworks of organizing resources, quantifying qualitative data, integrating them with quantitative data from natural and technological sides of area studies, analyzing whole data sets, and finally constructing comprehensive, objective and reproducible images of particular areas.

However, qualitative data sets of area studies are too diverse to treat everything from the beginning, and this is the reason why area informatics begun focusing on time and space attributes. Spatiotemporal perspectives form a common basis for humanities: time and place are a few quantitative attributes that are familiar to humanities and relatively easy to be derived from humanities' resources, and accordingly a huge amount of knowledge about how to use spatiotemporal attributes explicitly and implicitly has been accumulated in the fields. In the past, quantitative approaches to using spatiotemporal attributes were not popular in humanities, because appropriate methods and tools to describe and process these attributes were not matured. Recently, circumstances have gradually changed due to the dissemination and development of excellent and free software such as GIS (Geographic Information Systems/Sciences) tools, digital maps, digital gazetteers, and metadata that treat spatiotemporal attributes appropriately.

The Humanities GIS Research Group (H-GIS) [1] is a leading interdisciplinary research group in Japan to develop and apply spatiotemporal informatics to humanities. The H-GIS was originally established voluntarily by researchers from a variety of fields including information, engineering, history, literature, geography, health science, and ecology. Later, the H-GIS was reestablished as a research group under the Joint Research Project organized by the Center for Integrated Area Studies (CIAS) [2] Kyoto University, and has focused its efforts on area studies. Dialoguing between humanities and information researchers to make mutual consensus, and studying information engineering technologies such as GIS, GPS (Global Positioning System), RS (Remote Sensing), database, metadata, and ontology and so on, the H-GIS has developed many databases, Resource Sharing Systems to integrate databases on the Internet, spatiotemporal tools called HuMap and HuTime, and some ontology dictionaries about places and dates.

In this paper, current situations of spatiotemporal informatics in area studies will be summed up, and then H-GIS outcomes will be introduced as an example of applying spatiotemporal informatics to area studies. Chapter 2 introduces two information models proposed by the H-GIS and explains the concept and structure of area informatics, chapter 3 explains Resource Sharing Systems, chapter 4 overviews some spatiotemporal tools, and chapter 5 briefly introduces ontology databases. Finally, status of area studies is summarized and future directions especially related to ontology are considered in chapter 6.

2 Information Models for Area Informatics

Following chapters describe some outcomes accomplished by the H-GIS. First, this chapter introduces two information models. One is the schematic model that organizes an overall image of area informatics from data, information and knowledge

flows, and the second is the data model that gives the basic concepts of the H-GIS research approaches and tools.

2.1 Research Model of Area Studies

Area studies can be modeled from the viewpoint of data, information and knowledge flow as shown in Figure 1. This picture is developed to organize an overall image of the relationships between research phases of area studies, data/information/knowledge flows, and necessary information technologies and tools [3, 4].

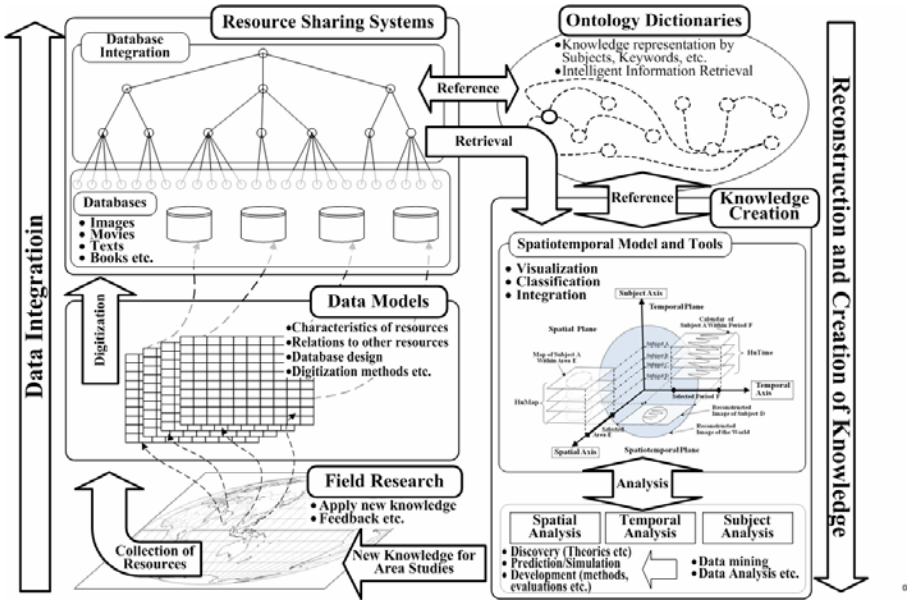


Fig. 1. Schematic Structure Model of Area Studies

Area studies are carried out in "research places" such as research fields, libraries, archives, laboratories and so on. Research resources are collected by observing, measuring, interviewing, reading, searching, and discovering activities. Authoring tools, on-site data entry tools, wireless data terminals are the typical necessary information tools in this phase [5, 6].

Collected resources are organized according to appropriate "data models". That is, resources are digitized and encoded depending on their disciplines (e.g., history, economy, agriculture, ecology, anthropology, and medicine), media (e.g., texts, images, movies, and sounds), data types (e.g., numbers, strings, and symbols), storage media, and collection types (relations with other resources). Digitizing techniques, media technologies, data models, metadata, and encoding methods are the main information topics in this phase. The H-GIS is preparing for several guidelines to digitize and organize resources.

Organized resources are compiled into "databases". Database management systems, information retrieval methods, and user interfaces are the main information technologies in this phase. Databases are the basis of information processing, but in Japan, numbers and sizes of databases especially for humanities are smaller than other countries. The H-GIS has supported database creation to enrich humanities' databases.

Most of the digital resources are distributed over databases. Researchers must know locations of databases for a start, then access databases one by one, and retrieve them by different methods. This is time consuming and tedious jobs. A "Resource Sharing System (RSS)" is a new information system to integrate databases on the Internet and to provide users with a uniform interface to retrieve databases seamlessly in one operation [7, 8]. RSSs are the key technology to integrate digital resources. Structures of RSSs will be explained in Chapter 3.

Each database encodes data by specific language, vocabularies, and terminologies according to the discipline that the database depends on. Since a given word does not always have the same meaning in different disciplines, RSSs alone are not enough to retrieve data across disciplines. Dictionaries that organize words and their meanings in a formalized structure are necessary to solve this problem. Using these dictionaries, computers can automatically relate words according to their meanings (synonym, antonym), hierarchical conceptions (taxonomy) and so on. Ontology is thought to be the key technology in this phase, and RSSs together with "ontology dictionaries" are expected to provide advanced data accessing methods. Some ontology tools developed by the H-GIS will be mentioned in Chapter 5.

"Knowledge Creation" is the final phase of the model. All data are visualized, integrated, classified, compared, and analyzed to discover and/or create new knowledge about particular areas. Spatiotemporal attributes are crucial in this phase, and spatiotemporal information tools, geo-statistics, and data mining techniques are key technologies to process these attributes. New knowledge created in this phase is fed back to the fields, and a new research process will begin. Spatiotemporal tools developed by the H-GIS will be explained in Chapter 4.

All information concepts and technologies indicated here are necessary for area studies, but the following chapters will focus on RSSs, spatiotemporal informatics, and ontology.

2.2 Data Model of Area Studies

When resources are applied to computer processing, their data must be comparable, that is, they must be quantitative values, ordered values or at least symbols clearly distinguishable from others. Typical resources related to area studies are books, magazines, papers, documents, maps etc. These resources have bibliographic attributes such as subjects, creators, titles, contributors that can be used to identify and find resources. Unfortunately, most of the bibliographic attributes are ambiguous and difficult to quantify, but time and place are a few quantitative attributes that are relatively easy to be identified from bibliographic attributes. Moreover, there are also many other resources whose bibliographic attributes are difficult to identify. Nontext materials such as archaeological remains, excavations, and natural phenomena are typical examples. For these resources, times and places (when and where a resource

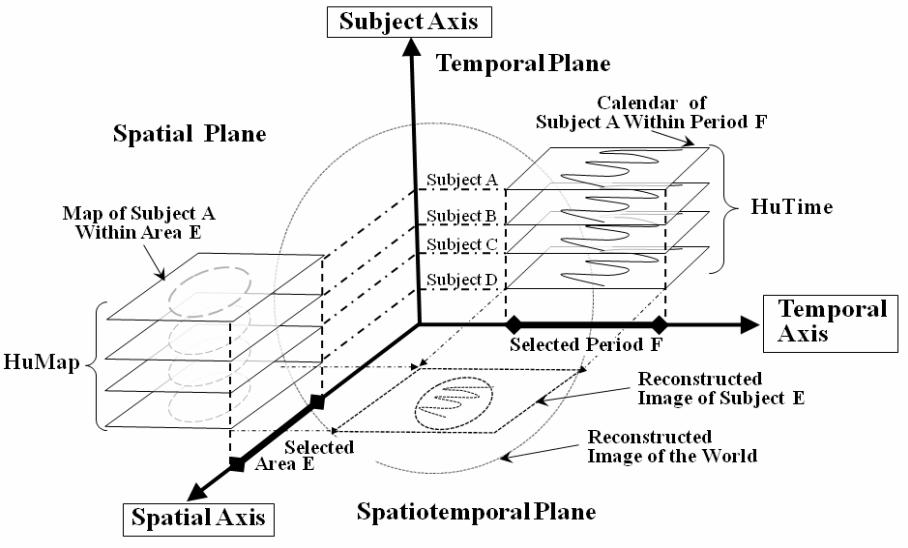


Fig. 2. Data Model of Area Studies

created, used, discovered etc.) might be the only quantitative attributes which can be used to identify and retrieve resources. These are backgrounds why spatiotemporal attributes together with traditional bibliographic attributes are essential in area informatics especially in "Knowledge Creation" phase in the Figure 1.

Figure 2 shows the data model proposed by the H-GIS [9], which is developed to give an overall view of relationships between data types and tools necessary for area studies. In this model, all objects (materials, events, phenomena etc.) are expressed as dots distributed in a 3D space spanned by three axes.

The subject axis groups objects by subjects (i.e., vocabularies/meanings), which corresponds to bibliographic databases. An RSS is the typical database system developed by the H-GIS, which will be explained in Chapter 3. A retrieval operation extracts objects that are distributed within a particular section on the axis (i.e., a particular subject). Zoom-in operation narrows down subjects using narrower vocabularies listed in thesauruses. Zoom-out is the opposite operation. Data mining is to discover some relationships among objects by vocabularies.

The spatial axis arranges objects by place, which corresponds to maps. A mapping operation extracts objects that are distributed within a particular section on the axis (i.e., a particular area). A 2D plane spanned by the subject axis and the spatial axis includes maps about particular subjects and areas (subject maps). These maps correspond to "layers" of GIS tools. HuMap is a GIS tool developed by the H-GIS, which will be explained in Chapter 4. Zoom-in operation enlarges a particular area of a map in great detail. Zoom-out is the opposite operation. Data mining is to find patterns from distributions of objects on a map such as randomness, coherency, periodicity and decrement from the center to periphery, and to find patterns between maps such as coexistence and transition.

The temporal axis arranges objects by time, which corresponds to calendars. A calendaring operation extracts objects that are distributed within a particular section on the axis (i.e., a particular period). A 2D plane spanned by the subject axis and the temporal axis includes calendars about particular subjects and periods (subject calendars). These calendars correspond to "layers" of temporal information tools. HuTime is a temporal information tool developed by the H-GIS, which will be explained in Chapter 4. Zoom-in operation enlarges a particular period of a calendar in great detail. Zoom-out is the opposite operation, showing a broad sweep of history. Data mining is to find patterns from distribution of objects on a calendar such as randomness, coherency, periodicity and decrement from one time point, and to find patterns between calendars such as coexistence and transition.

A 2D plane spanned by the spatial axis and the temporal axis includes spatiotemporal layers; there are no traditional tools (e.g., maps, calendars, GIS tools) corresponding to this plane. Tools that can treat these layers are innovative, which will be discussed in Chapter 6.

3 Resource Sharing Systems and Metadata

Variety types of databases have been created as the result of research activities. These databases must be expected to be the basis of area studies. However, one database cannot always hold all related resources, for example, users must visit many databases to make up the whole map of South East Asia. These users must be disappointed. Since each database is designed and built based on conventions of each institution and/or discipline, even the same materials are organized differently by databases. That is, locations, metadata, languages, vocabularies, terminologies, and retrieval procedures (i.e., specifications) are different from databases, and users have to learn specifications of all databases they want to retrieve data. A "Resource Sharing System (RSS)" is developed to solve these problems by integrating databases on the Internet.

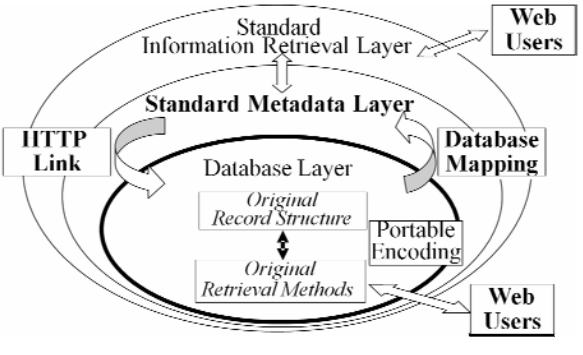


Fig. 3. Resource Sharing Data Model

By the way, a nationwide OPAC system can also integrate all library databases in a country. This integration is achieved by the technique that all databases in an OPAC system must follow the same specification, that is, an OPAC system consists of homogeneous (or same structure) databases. The remarkable innovation of an RSS, different from OPAC systems, is that an RSS can integrate heterogeneous (or different structure) databases on the Internet. This is crucial to area studies, because researchers use and need heterogeneous databases (e.g., libraries, archives, museums). For example, the NIHU RSS [10] seamlessly integrates more than 100 heterogeneous databases that five institutes hold. The CIAS also opens its RSS for area studies [11].

An RSS is modeled as 3-layered structure shown in Figure 3. The first layer is "Database Layer" where original databases are allocated. Databases in this layer are required to encode data by XML. An RRS needs frequent data conversion and exchange processes between databases. Since data structure is different from each database and data definitions are often modified, data conversion process must be easy to follow these differences and modifications uniformly and independently from databases. In addition, since binary data encoding and structures are different from database systems, "data portability" is necessary while exchanging data between database systems. XML has effective data conversion methods (XSLT) and data portability function between heterogeneous information systems.

The second layer is "Metadata Layer". Standard metadata is essential to "hub-data" that gives an independent record structure from databases. In an RSS, records of each database are converted into (or mapped onto) the hub-data records, by which all records of databases integrated into an RSS have the same structure. Figuratively speaking, Metadata Layer is an envelope that wraps databases, and users can see all databases in this envelop as one database (i.e., one record structure). Selection or definition of metadata decides the performance of RSSs.

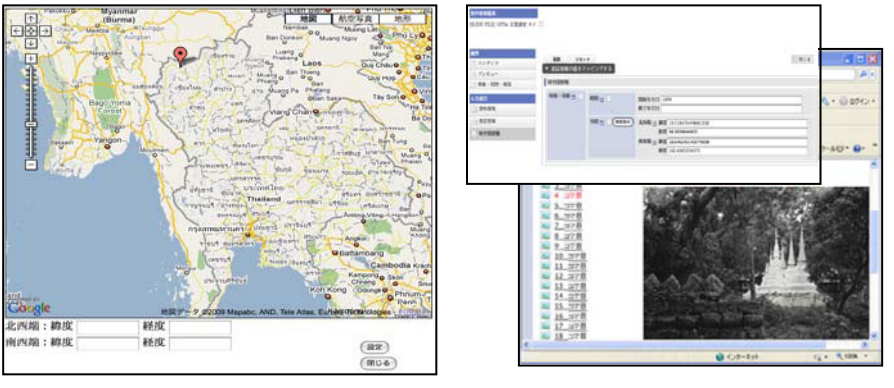
The third layer is "Information Retrieval Layer". Metadata specifications define only record structures. Even if database systems integrated into an RSS adopt the same metadata, retrieval procedures may be different from database systems. Database systems in this layer are required to cope with a standard retrieval procedure. Figuratively speaking, Information Retrieval Layer is also an envelope that wraps database systems, and users can see all database systems within this envelop as one database system (i.e., one record structure and one retrieval procedure).

Spatiotemporal attributes are often difficult to organize as metadata: there are many kinds of spatiotemporal attributes in the world, but most metadata have few elements for them. To solve this problem, some RSSs expand original metadata by adding their own elements. The NIHU RSS is an example. Since the purpose of the NIHU RSS was to integrate heterogeneous databases, a minimum element set was thought to be the best metadata, and NIHU therefore adopted the Dublin Core Metadata Element Set (DCMES) [12] as its base metadata. DCMES is simple and widely used on the Web, but it is too small to describe precise spatiotemporal attributes. To describe a variety of spatiotemporal attributes of humanities' resources, the NIHU RSS adds spatiotemporal elements to expand the original DCMES [13].

On the other hand, the CIAS RSS is designed to organize wide range of resources collected by researchers of area studies. Since these resources are collections of field notes, photographs, maps, and so on, minimum elements' sets are not enough for the purpose. The CIAS RSS needs larger metadata sets [13], and introduces combinations

of EAD (Encoded Archival Description) [14], METS (Metadata Encoding & Transmission Standard) [15] and MODS (Metadata Object Description Schema) [16]. MODS is bibliographic metadata to describe information about each material. However, bibliographic attributes are not enough for collections: descriptions about the origins of resources, histories of ownership and the relationships between other resources are important information for collections. EAD is appropriate archival metadata to describe such information. Materials themselves, their bibliographic data, and their collection data must be linked together. METS is introduced to package these metadata, functioning as an envelope to organize related metadata.

A retrieval example of the CIAS RSS is shown in Figure 4. In this example, locations of the contents are indicated by marks on the map. If a mark is selected, its information will be displayed.



(a) (b)

Fig. 4. Retrieval Example of the CIAS Resource Sharing System. (a) is a map view to retrieve records by locations, and (b) is an example of the retrieved record.

4 Spatial and Temporal Tools

Spatial tools and temporal tools are essential to area informatics to integrate, visualize, and analyze data quantitatively and objectively. This chapter explains a HuMap (a spatial tool) and HuTime (a temporal tool) that the H-GIS has developed.

4.1 A Spatial Information Tool (HuMap: Humanities Map)

HuMap is a newly developed free GIS tool that can handle objects on layers spanned by the subject axis and the spatial axis (i.e., subject maps) in Figure 2. HuMap is derived from TimeMap [17]. The innovation of TimeMap is that it can carry out temporal operations, but weakness is that it is rather a viewer than an analysis tool. The latest HuMap is completely different software from TimeMap: it can carry out visualization operation, spatial operations and temporal operations. Functions of HuMap are summarized as follows.

Basic Viewer Functions

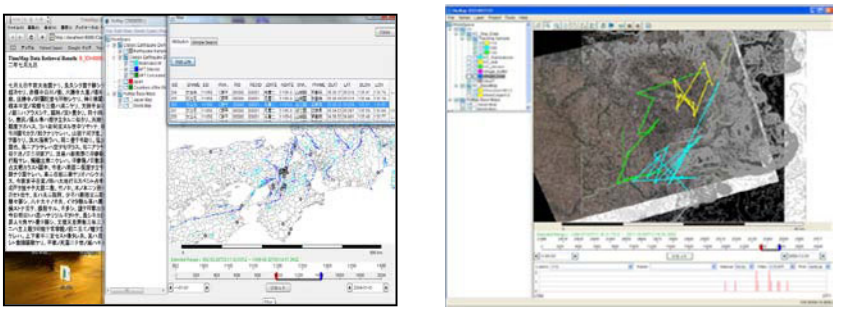
- Arrange/display various data by place and time
- Use maps whose coordinate systems are different simultaneously
- Multi-format: ESRI shape file, CSV, XML metadata, JPEG, JPEG2000 etc.
- Layer selection, change layer order, create new layers, delete layers etc.
- Change symbol/color/size/α-value of an object (feature)
- Zoom-in/out by place and time
- Import/export layer data
- Web-link
- Put and retrieve annotations on layers

Functions as a Spatial Tool

- Link with the data clearing house
- Retrieve objects (features) by place, time, and subject
- Choropleth Map
- Animations/tracking
- Logical operations between layers (Intersection, Union, Merge etc.)
- GIS functions (Dissolve, Buffering, Clipping, Tracking etc.)
- JAVA and R plug-ins for advanced analysis (under construction)

Data visualization is the basic use of HuMap. A typical view of HuMap is shown in Figure 5 (a). Main layers of the view are political-boundaries (polygon features), seismic active-faults (line features), and earthquake-events (point features) that are reconstructed from historical resources about earthquakes. We can see the spatial relationships between active faults and earthquake-events. Spatiotemporal queries are important function of HuMap that is used to select objects within the particular area and period. The upper-right table in the figure shows the result of the query to choose earthquake-events during 1000 AD to 1500 AD in Kyoto area. HuMap also has a Web-link function to link an object and its related information on the Web using URLs. The left side text box in the figure shows the linked document of an earthquake-event in the full-text database.

HuMap can carry out basic logical operations between layers and some spatial operations such as dissolve, buffering, clipping, tracking, etc. Figure 5 (b) is the example of tracking to analyze the movement of objects according to time.



(a) (b)

Fig. 5. Example Displays of HuMap: (a) is a typical view of HuMap and example of spatiotemporal query, and (b) is an example of tracking analysis

4.2 A Temporal Information Tool (HuTime: Humanities Time)

HuTime is an innovative temporal information tool that can handle events on layers spanned by the subject axis and the temporal axis (i.e., subject calendars) in Figure 2. In the same way as HuMap overlays and visualizes maps and images by referencing positions, HuTime arranges calendars, documents, graphs, and images by referencing time lines. Users can easily grasp temporal relations and/or patterns of events from different calendars by time. Although there are some temporal tools such as SIMILE Timeline [18], these tools primarily focus on data visualization. The strength of HuTime is that it can carry out basic temporal operations. Functions of HuTime are summarized as follows.

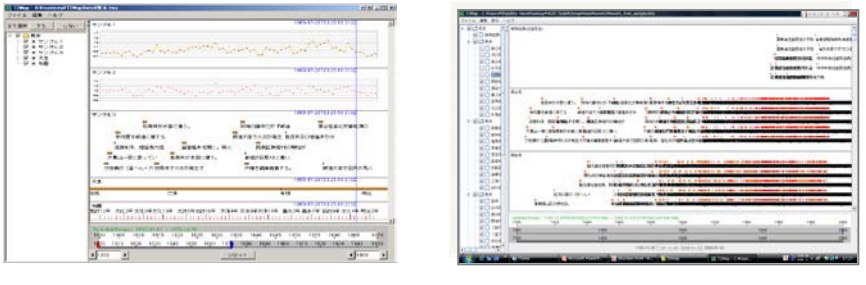
Basic Viewer Functions

- Arrange/display various data by time
- Use calendars whose calendar systems are different simultaneously
- Multi-format: CSV, XML metadata, JPEG, GIF etc.
- Layer selection, change layer order, create new layers, delete layers etc.
- Change symbol/color/size/α-value of an event
- Zoom-in/out by time
- Import/export layer data
- Web-link

Functions as a Temporal Tool

- Link with data clearing house (under construction)
- Retrieve data by time and subjects
- Search/filter functions to select specific events on a layer
- Logical operations between layers
- Analyze periodicity, causal relation etc (under consideration)

Figure 6 shows example displays of HuTime. A typical view of HuTime is shown in (a). This view consists of a wind velocities layer (dots graph), a precipitation layer (dots graph), articles about natural disasters layer (document) and a Japanese calendar layer. We can see the relationships between winds, rainfalls and disasters. Temporal queries are the basic function of HuTime to extract objects occurred within a designated period and/or matched with specific conditions (b).



(a) (b)

Fig. 6. Example Displays of HuTime: (a) is a typical view of HuTime, (b) is an example view of a temporal query

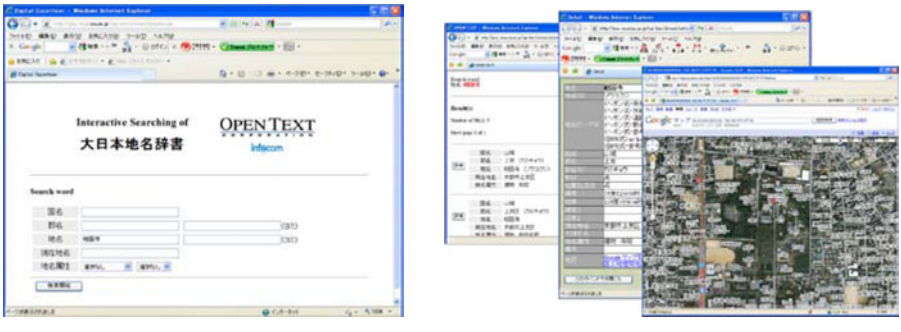
5 Ontology Databases

Two databases, the "Japanese Historical Gazetteer" and the "Calendar Database", have been developed by the H-GIS to support spatiotemporal data organization for databases and to add advanced data retrieval functions to RSSs. Strictly speaking, they are, at present, "thesauri" rather than ontology dictionaries.

5.1 Japanese Historical Gazetteer

The "Japanese Historical Gazetteer" [19] is a simple thesaurus. Entry words of the thesaurus are historical place names in Japan collected from "Dainihon Chimei Jisho (Dictionary of Place Names in Greater Japan compiled by Togo Yoshida)", "Engishi Jinmyocho (Procedures of the Engi Era)", "Nihon Jiin Sokan (Directory of Japanese Temples)", "Jinsoku-zu (Quick Mastery Maps: Kanto Region)", and "Kasei-zu (Temporary Maps: Kinki Region)". At present, about 150,000 place names are registered in the Gazetteer. This is the largest free academic gazetteer database in Japan. Each entry word includes its present place name, attributes (pronunciation of the place name, types of the place, broader and narrower place names and so on), and geographic coordinates (longitude and latitude).

This gazetteer will be used to identify correct place names, to convert a place name into a pair of coordinates, to visualize spatial relationships between places, to analyze spatial patterns of particular place names and so on. Figure 7 shows example displays of the Gazetteer (searching for the Shokoku-ji Temple).



(a) (b)

Fig. 7. Example Displays of Japanese Historical Gazetteer: (a) is a search window, and (b) shows its retrieval displays

5.2 Calendar Database

The "Calendar Database" is a simple table to organize all dates (Japanese dates, Chinese dates, Gregorian dates and so on) according to Julian dates. It is used to convert a date from one calendar to another.

6 Discussions and Considerations

The motivation of the H-GIS was to apply spatiotemporal informatics to humanities (especially of area studies) for facilitation of data usage and data integration with natural and technological studies. The H-GIS has made progresses in collaboration with some international projects such as ECAI (The Electronic Cultural Atlas Initiative) and PNC (The Pacific Neighborhood Consortium). ECAI, based at University of California, is a global consortium of people who share the vision of creating a distributed virtual library of cultural information with a time and place interface [20]. The H-GIS and ECAI keep in exchanging technologies and knowledge about spatiotemporal informatics. PNC, mainly supported by Academia Sinica, explores issues of interdisciplinary collaboration, and the development of the cultural knowledge contents [21]. The focus of PNC is not exactly on spatiotemporal informatics, but many members engage in digitization of materials, programming spatiotemporal tools and so on for humanities' contents. The H-GIS also keeps in communicating with PNC about spatiotemporal issues. Spatiotemporal informatics in the humanities is gradually becoming recognized as an important paradigm in Japan. Some academic societies such as "IPSJ SIG Computers and the Humanities" and "Branch of Historical Geography, The Human Geographical Society of Japan", for example, are interested in spatiotemporal informatics. The H-GIS keeps relationships with these societies.

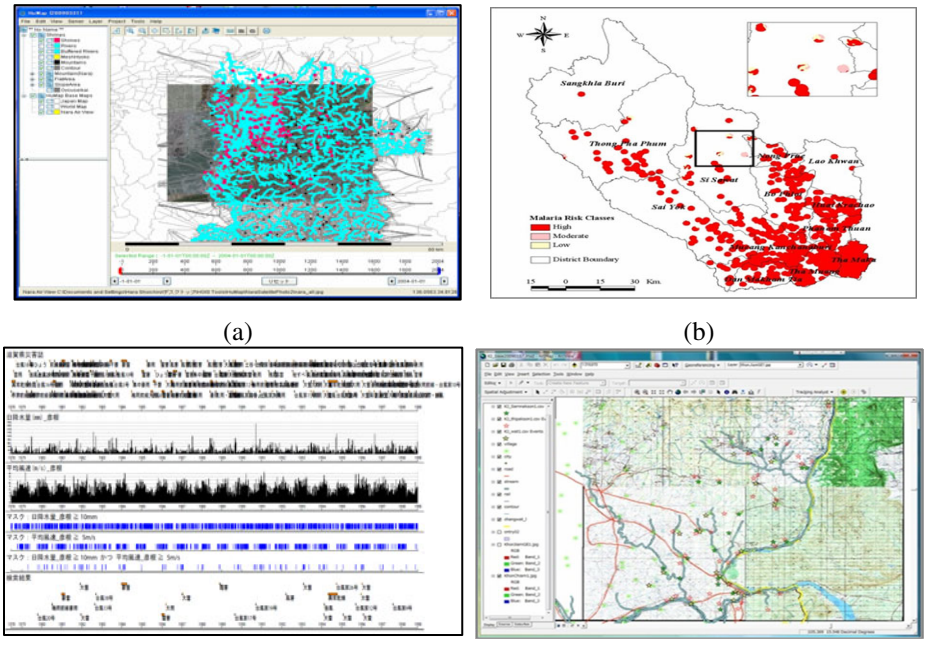


Fig. 8. Research Examples using HuMap and HuTime: (a) is a landscape analysis, (b) is a Malaria risk analysis, (c) is a temporal analysis about weather and disasters, (d) is a spatiotemporal analysis about Thai monks' movement

The H-GIS has at last developed HuMap, HuTime, the Japanese Historical Gazetteer Database and the Calendar Database. These tools are in a trial phase, and they are used by researches. Figure 8 shows some research examples using HuMap and HuTime. The example (a) is a landscape analysis of sacred places around ancient Shinto-shrines. The example (b) is a Malaria risk analysis in Thailand, which is a typical GIS application to health and medical care. The example (c) is a temporal analysis using weather data (Figure 6 (a)) to find relationship between heavy rainfall, strong wind, and disastrous events. The example (d) is a spatiotemporal analysis to find backgrounds that explain pilgrimages of monks in Thailand. These research trials gradually attract interesting of area studies' researchers. On the other hand, the H-GIS gets feedbacks from researchers to modify and expand functions of the tools.

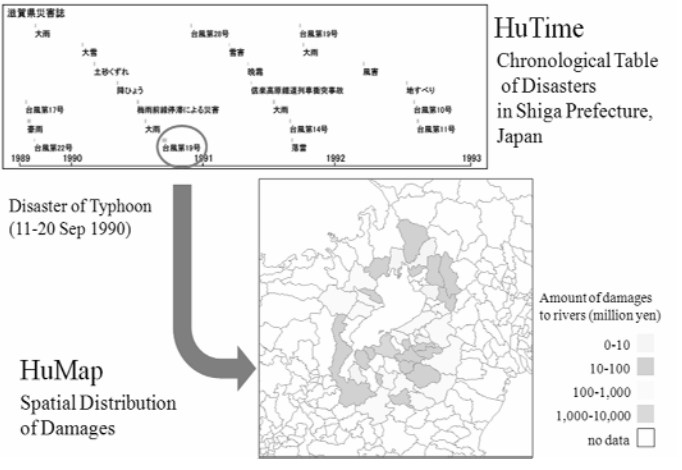


Fig. 9. A Simulation of Spatiotemporal Analysis by Switching HuMap and HuTime

HuMap has rich spatial functions but limited temporal functions, while HuTime has no spatial functions but rich temporal functions. The H-GIS is trying to combine these two tools to create a revolutionary tool that can handle objects on the planes spanned by the spatial axis and the temporal axis in Figure 2. A map is a "piece" of the world we see from a narrow window of particular subject and area, and a calendar is also a "piece" of the world viewed from a narrow window of particular subject and period. Area studies (and sciences) can be said the processes of reconstructing the world image from these pieces. A layer in spatiotemporal plane is an image reconstructed from a pair of map and calendar. Analogically speaking, just as a 3D human body image can be reconstructed from many pieces of CT images, the world image can be reconstructed from variety pieces of maps and calendars.

If such a tool is realized, users will be able to analyze objects from spatial and temporal perspectives simultaneously. However, since maps and calendars have different layout structures and four dimensions are at least necessary to display them simultaneously, it is difficult to design such viewers. One solution may be to switch HuMap and HuTime properly according to users' perspectives. Figure 9 shows a simulation that visualizes relationships between temporal climates patters and spatial disasters patterns by switching HuMap and HuTime. Figure 10 shows the schematic

specification of a spatiotemporal tool. Internal data structures of HuMap and HuTime are not identical due to different development backgrounds, which make data exchange between two tools difficult. To solve this problem, a mutual metadata is defined. Just like a pipeline, data are exchanged via this metadata. This switching function will be realized within this year. These flexible developments are possible because all tools are programmed and owned by the H-GIS. This is the reason why the H-GIS has developed all systems and tool by itself.

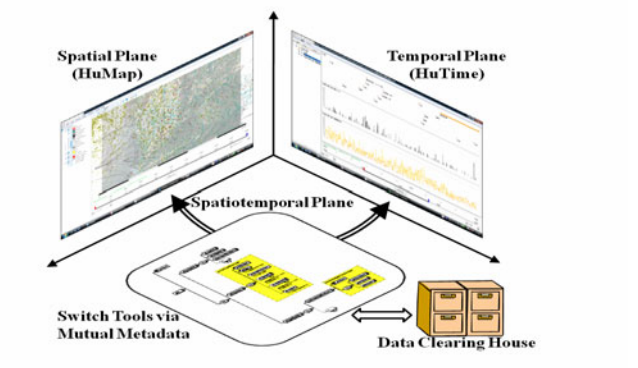


Fig. 10. Schematic Specification of a Spatiotemporal Tool

Accumulations of data whose semantics and spatiotemporal attributes are well organized are essential in order to fully use spatiotemporal tools. An RSS is a solution to this problem, and several systems are in operation. Two challenges now remain to be faced. One is to introduce ontology technology. Since area studies are collection of many disciplines, a given word in a discipline not always has the same meaning in the other discipline. Vocabularies should be organized by meaning so that computers can process them, allowing information retrieval from a wide range of databases.

The second challenge is to expand and refine the spatiotemporal attributes of metadata. The NIHU RSS uses DCMES. DCMES is simple but does not allow precise description of spatiotemporal information. The NIHU RSS expands DCMES to describe spatiotemporal information precisely. However, this makes the metadata different from that of standard DCMES and difficult to link with other RSSs. The CIAS RSS uses a combination of MODS, EAD and METS. These metadata are good enough to describe precise information. However, there are several ways to describe the same piece of information, which decreases retrieve abilities between other RSSs.

Spatiotemporal informatics in area studies has almost accomplished its initial step. To make next progresses, the H-GIS is trying to set up interdisciplinary research projects to collect, organize, analyze and integrate a variety of data sets using RSSs and spatiotemporal tools. Evaluations, requests and opinions are being fed back from researchers, which will be the hints for further developments. Concurrently, the H-GIS is preparing for compiling thesauri of several disciplines related to area studies, which will be used for intelligent data processing.

Acknowledgement. This research was conducted by the H-GIS research group. Special thanks are due to Prof. Mamoru Shibayama, Prof. Tatsuki Sekino and Prof. Masatoshi Kubo for helpful discussions. This research has been supported partly by Grant-in-Aid for Scientific

Research (A) (19201051) and two projects of NIHU, “Resource Sharing System for the Humanities” and “Integrated Study of Water and People in Human Asia”.

References

1. The Humanities GIS Research Group (H-GIS), http://www.h-gis.org
2. The Center for Integrated Area Studies (CIAS) Kyoto University, http://www.cias.kyoto-u.ac.jp
3. Hara, S., Shibayama, M., Kishi, T. (eds.): Special Issue on Area Informatics, Intriguing ASIA, vol. 113 (2008) (in Japanese)
4. Hara, S., Shibayama, M.: Development of Area Informatics and Spatiotemporal Analysis Tools. In: Proceeding of the 2007 Symposium on Computers and the Humanities (JINMONKON 2007), pp. 71–78 (2007) (in Japanese)
5. Ishikawa, M., Hara, S., Shibayama, M.: On-site Data Entry Tools. In: Proceeding of the 2007 Symposium on Computers and the Humanities, JINMONKON 2007 (2007) (in Japanese)
6. Ishikawa, M., Kaneko, K.: Design of a Map Annotation System Using a Digital Pen for Field Work. In: Proceeding of Society for Information Technology & Teacher Education (SITE 2009), CD-ROM (2009)
7. Hara, S., Yasunaga, H.: Resource Sharing System for Humanity Researches. In: Proceedings of the Third International Conference on Language Resources and Evaluation, pp. 51–58 (2002)
8. Hara, S.: Inter-institutional Databases Integration by Z39.50 and Metadata. Information Processing IPSJ 43(9), 968–974 (2002) (in Japanese)
9. Sekino, T., Kubo, M.: T2Map – A tool for temporal information system. In: Proceeding of the 2007 Symposium on Computers and the Humanities (JINMONKON 2007), pp. 183–188 (2007) (in Japanese)
10. NIHU: Resource Sharing System, http://www.nihu.jp/kyoyuka/tougou/index.html
11. CIAS: Resource Sharing System, http://area.net.cias.kyoto-u.ac.jp/GlobalFinder/cgi/Start.exe
12. Dublin Core Metadata Initiative (DCMI): Dublin Core Metadata Element Set, Version 1.1. (2008), http://dublincore.org/documents/dces/
13. Hara, S.: Studies on Resource Sharing System and Metadata for Area Studies. Southeast Asian Studies 46(4) (2009) 2009 (in Japanese)
14. Library of Congress: Encoded Archival Description Version 2002 Official Site, http://www.loc.gov/ead/eaddev.html
15. Library of Congress: Metadata Encoding & Transmission Standard Official Site, http://www.loc.gov/standards/mets/mets-schemadocs.html
16. Library of Congress: MODS Metadata Object Description Schema Official Web Site, http://www.loc.gov/standards/mods/
17. Johnson, I.: Time Map: Time-based Interactive Mapping, The Archaeological Computing Laboratory. University of Sydney, Sydney, http://www.timemap.net/index.html.TimeMap
18. Massachusetts Institute of Technology, SIMILE Timeline, http://www.simile-widgets.org/timeline/SMILE
19. Oketani, I.: The Database of Topographical Maps and Place Names, SIG Technical Reports, Vol.2009-CH-83 No.3, pp. 1-7, (2009) (in Japanese)
20. The Electronic Cultural Atlas Initiative (ECAI), http://www.ecai.org/
21. The Pacific Neighborhood Consortium (PNC), http://www.pnclink.org/

Author Index