2013 International Conference on Advances in ICT for Emerging Regions (ICTer): 272 - 276

Multimodal + Multimedia + Sensors = Pleasant Interfaces

Masahito Hirakawa
Interdisciplinary Graduate School of Science and Engineering, Shimane University
Nishikawatsu 1060, Matsue 690-8504, Japan
hirakawa@cis.shimane-u.ac.jp

***Abstract*—If the computer is a machine to help humans, its functions should be completely useful, otherwise it is ineffectual, despite containing several powerful functions. Thus, interface development is an essential issue. We discuss the idea of a pleasant interface in this paper. The pleasant interface does not require much effort while using the computer and operates in the background. In addition, this interface helps humans do a thing which is worth doing. In this study, trials conducted in a laboratory are explained as a step toward realizing pleasant interfaces, which include multidip aquatic interface, spatial auditory, and gait interface systems.**

***Keywords*— Pleasant interface, multimodal, multimedia, sensor, aquatic interface, spatial auditory system, gait interface**

I. Introduction

Divergent thinking is considered essential for creativity, which is defined as flexible ideation to generate numerous responses to open-ended and multifaceted problems [1]. It is often used as a parallel to convergent thinking, which involves limiting the number of possible responses to reach correct solutions and works best with well-defined problems that have a clear, defined response.

In their book, Land and Jarman [2] presented that capacity for divergent thinking deteriorates with age. In fact, in one of their experiments, 1,600 children aged between three and five were given divergent thinking tests, which were used by NASA to measure the potential for creative work, and 98% of them scored in the top tier called creative genius. Five years later, when they were aged between eight and ten, 32% scored in the creative genius category. Ten years later, only 10% of the children scored in this category. In tests of over 200,000 adults over 25 years old, those in the creative genius category dropped further to 2%.

Progress in information and communication technology (ICT) may have spurred this trend. Big data [3] enables us to analyze petabytes of data collected from a variety of digital equipment, indicates social trends, and offers specific recommendations. The information extracted is definitely useful and beneficial, but it is obvious that software systems for big data management are neither omniscient nor omnipotent. If marketers become dependent on these systems, it might hamper their ability to think divergently.

Next, we focus on the history of user interface (UI) development. Since its inception through Xerox ALTOs in the 1970s, graphical user interface (GUI) still retains its position as the most conspicuous invention in UI development—it is best suited for desktop PCs. A touch user interface (TUI), or a gestural interface in general, is becoming common as mobile computing technologies evolve, that leads us into another stage of fulfilment in natural interfaces.

The size of a computer device dictates the choice of suitable user interface, while, when humans are used to the style of a certain interface (e.g., GUI and TUI), they do not replace it with another interface unless there is a substantial reason to do so.

In the near future, computers will be embedded completely in conventional equipment, such as TV, and in houses to the extent that we will no longer recognize the existence of computers; i.e., they will become invisible [4]. Computers are subject to the demands and, therefore, should support a variety of sensing channels to simplify communication with the human. Sensing technology will become much more important.

In this study, we discuss the idea of a pleasant interface which is beyond the invisible interface. Computer usage should not require much effort and the interface system should run in the background. This is necessary to ensure invisibleness, but these factors may not provide us with a perfect solution. The pleasant interface also helps humans do a thing which is worth doing. Note that we have the initiative to create actions; therefore, the term "user" is not apt.

In this study, trials that have been conducted in our laboratory are explained as a step toward realizing pleasant interfaces, which include multidip aquatic interface, spatial auditory, and gait interface systems.

II. Pleasant Interfaces

The purpose of a computer is to help humans accomplish their tasks with less effort; thus, it is imperative that its functions are fully useful. Therefore, user interface development is essential.

While the notion of invisible interface [4] is an ideal goal in user interface research, we maintain that a pleasant interface is preferable, which enables computers to help humans accomplish tasks without effort, and do a thing which is worth doing.

This idea is similar to intellect augmentation proposed by Douglas Engelbart [5], [6], which aims to increase human abilities and not just replace human effort. Note that Engelbart discussed the relationship between humans and technology under the concept of co-evolution. To enable technology to be more pleasant, we should consider the relationship between design and technology.

In the past, William Morris led an internationally recognized design movement, the Arts and Crafts Movement, in the late 19th century and called for social reform through integration of art and craftsmanship. He observed that nothing of value could be produced by machinery since mass production brought mass degradation [7]. In his own words, *Have nothing in your houses that you do not know to be useful, or believe to be beautiful.*

Bauhaus, which is a school founded by Walter Gropius in Germany, had many goals in common with the Arts and

978-1-4799-1276-6/13/$31.00 ©2013 IEEE

Crafts Movement. A manifesto of the Bauhaus school is given in the following [7]:

There is no essential difference between the artist and the craftsman. The artist is an exalted craftsman. In rare moments of inspiration, moments beyond the control of his will, the grace of heaven may cause his work to blossom into art. But proficiency in his craft is essential to every artist. Therein lies a source of creative imagination.

Let us create a new guild of craftsmen, without the class distinctions which raise an arrogant barrier between craftsman and artist. Together let us conceive and create the new building of the future, which will embrace architecture and sculpture and painting in one unity and which will rise one day toward heaven from the hands of a million workers like the crystal symbol of a new faith.

Obviously, a variety of studies based on different backgrounds are necessary to attain realization of pleasant interfaces. The primary areas are multimodal, multimedia, and sensor approaches.

Considering a computer is used by humans to perform specific tasks, specification of their needs should be made as direct as possible. Multimodality in communication between humans and computers is essential. In addition, the ability to sense the status of the human is helpful for the computer to decide on an appropriate response. This enables the computer to invoke a certain action without having any explicit command issued by the human. Furthermore, the human is freed from tedious explanations about the context because the computer infers it.

On the other hand, the computer must provide responses to the human, else the computer is ineffectual. Multimodal interfaces work well again in this situation. Because they provide a framework for communication between the human and computer, the design of the contents to be delivered through interfaces is also important. An effort should be made to increase the quality of multimedia content, which include factors such as graphics, video, and sound.

Figure 1 explains the conceptual schematic of pleasant interface development. Technology and design are the two wheels of a computing vehicle that humans drive.

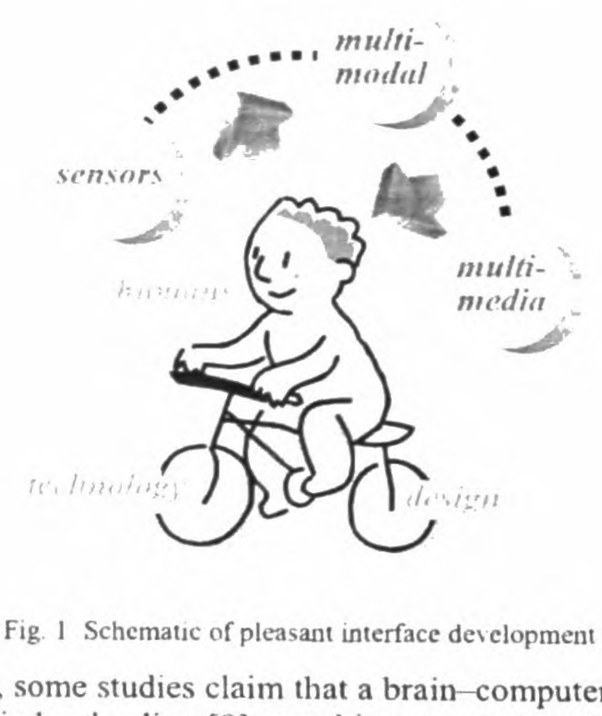

Fig. 1 Schematic of pleasant interface development

Finally, some studies claim that a brain–computer interface [8] and mind uploading [9] are ultimate user interfaces. Here the focus is on engineering without straying into neuroscience matters as multiple ethical issues must be discussed in parallel with their research.

In the following sections, the trials for realizing pleasant interfaces are explained. These trials include multidip aquatic interface, spatial auditory, and gait interface systems. All three trials deal with multimodal interaction, whereas the trials involving the spatial auditory and gait interface systems deal with multimedia and sensor approaches, respectively.

III. Aquatic Interface System

While GUI has remained the de facto standard for many years, the multimodal interface is considered to be a new, powerful approach. Among these, gestural interfaces are becoming increasingly popular; a touchscreen-based interface is common in smartphones and tablets, and 3D gestural tracking devices such as Microsoft's Kinect are available and used as a tool for developing interesting application prototypes. A tangible interface (e.g., Reactable [10]) is another promising approach.

These trials provide powerful and flexible means for human–computer interaction, but interaction will not be limited to things on a display or a tabletop. We have proposed adopting water as an interaction medium [11]-[13] because water is indispensable to humans.

There are several trials using water as an interaction medium, such as Floating Words [14] and Unconscious Flow [15], in which a human interacts or plays by putting a hand or ladle into water. Tangible Humidifier [16] provides a simple interface that allows a human to turn the humidifier on and off by touching the water surface in a bowl. WaterTouch is an aquatic interactive multimedia sensory table [17] and captures the position and movement of fingers dipped in water.

We first presented the idea of using water in interaction with a computer in a previous study [11]. The system is capable of scooping up a computer-generated object projected on water by the hands, allowing the human to enjoy selecting the object and trigger the associated computer application/command. However, hand recognition is performed by referring to the human's skin color, and its recognition performance depends largely on environmental conditions.

We developed a second version of the system [12], which adopts the frustrated total internal reflection (FTIR) technique [18] to solve the drawbacks observed in the first version. On each side of a transparent acrylic tank, blue LEDs are arranged. The light is internally trapped because of total internal reflection. When an object (e.g., a hand) is dipped in water, frustration of the total internal reflection occurs, causing light to scatter. Detecting this light using cameras located under the tank enables the computer to find the object position in a 3D tank space. Figure 2 shows the snapshot of the system demonstrated at an exhibition.

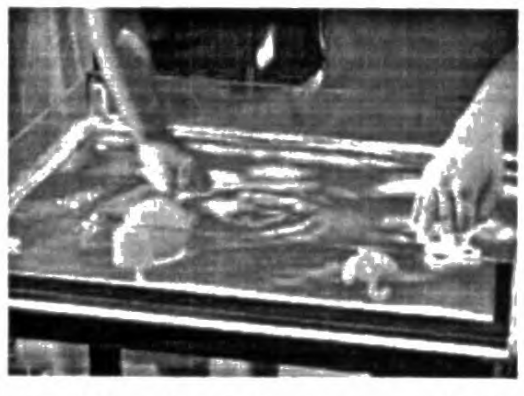

Fig. 2 Snapshot of the system using FTIR

The above two versions use a computer vision technique for object detection and tracking. One drawback of this approach is the need for a large space to set up the system. This leads to a bigger problem for some applications such as a foot bath (Ashiyu in Japanese) because the system must be set at a near-ground level. To address this drawback, we developed an alternative version of the system named SensorTank [13].

SensorTank is built with a transparent acrylic tank, whose size is 88 cm × 50 cm × 20 cm. An LCD monitor is placed below the tank. Combinations of a laser and a phototransistor (sensor unit) are placed on the side planes of the tank, facing each other. The size of the entire system, excluding the PC, is 99 cm × 80 cm × 37 cm in the current implementation. Figure 3 shows the system overview. The system captures not only the position of objects in the water but also their volume.

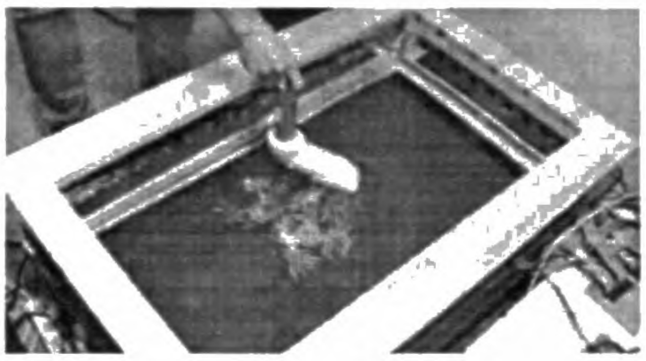

Fig. 3 SensorTank system

A total of 78 sensor units are mounted at distances of 5 cm (horizontal) and 3 cm (vertical) between modules on all four sides of the tank by using three mounting layers. This 5 cm spacing is not enough for finger gesture detection, but it is acceptable to detect some movements such as foot motion.

We have implemented some application prototypes, including footbath activity. Footbath activity has long been a social activity in Japan, where people come to meet and talk. The system calculates the centroids of objects being immersed and displays patterns of moving symbols, depending on the objects' information (Fig. 3). Incorporating interactive capabilities inside the footbath allows humans to enjoy and relax, with generated music tones creating a calm ambience.

In addition, an application of this system into rehabilitation is promising. A target symbol is presented to a patient to follow. The symbol's direction and speed can change depending on the status of rehabilitation, and the performance score can be provided for motivation. An interesting feature in using water (or liquid, in general) is that force feedback is provided whenever the patient moves his/her body part in the water. Moreover, the ability to adjust the water temperature could ease the physical strain on the patient.

IV. Spatial Auditory System

Multimedia is the basis of modern computers. Because of the fact that humans rely on vision as a sensory channel, the focus has been on the use of graphical or visual forms such as images and animation. Audio is another important media for expression of content. The idea of so-called earcon [19] was first proposed to present specific items or events by means of abstract patterns in loudness, pitch, or timbre of sounds.

Contents consist of not only objects appearing explicitly in the content but also their positions of representation in space. In audio, sound patterns are an issue, and the position of sound patterns is not fully utilized because realization of spatial sound generation or localization needs additional cost. Moreover, visual objects are explicitly perceivable in the sense that nobody misses their existence and positions, but this is not the case in audio. It is difficult for a system to locate sound at a certain position. This difficulty (or uncertainty from the viewpoint of a human) is generally considered a drawback, but an application scenario is presented later, where this characteristic works to its advantage.

We have investigated a multimedia framework in which an auditory table is a central component and designed to provide humans with effective visual and auditory feedback [20]-[22]. The system is capable of locating sound objects at any position on the table. Here there exist many surround sound systems for creation of a virtual sound space, but they aim to realize a "perfect" sound space as if the human can feel that he/she is there. One drawback is that the best spot for listening is predetermined. On the other hand, we are interested in creating an interactive collaborative environment where participants can work collaboratively sharing a common sound field—not only sound patterns but their positions.

Sixteen speakers are placed on a table (90 cm square) in a 4 × 4 matrix, and they are connected to a PC using two 8-channel audio interfaces. A video projector is mounted over the table and computer-generated graphic images can be projected onto its tabletop covered by a white cloth. We have implemented several input methods using Nintendo's Wii Remote, Microsoft's Kinect, and a shadow-based position detection technique to specify commands by the participants. Figure 4 shows the setup of this system.

Fig. 4 Spatial sound system

The speaker modules on the tabletop caused some restrictions in the system design. Therefore, we explored ideas to lower the restrictions [22]. The number of speakers was reduced from sixteen to four to increase flexibility in our system design. Instead, speaker modules are designed to be easily set at any position in a target working space. We also prepared a software tool for supporting sound localization control. These speaker modules were combined with a tabular-size multitouch device using the FTIR technique. A reconfigurable, fabric-type input sensing device was also developed.

We have implemented several applications that run in the systems explained above, including music mashup, reminiscence therapy, and education. Let us explain a Hyakunin-Isshu playing card application designed for educational purposes with an idea called sound hint [22]. Hyakunin-Isshu is a famous Japanese poetry anthology—one hundred waka poems by one hundred poets. In Japan, an experience in Hyakunin-Isshu is a step toward Japanese traditional culture and history, and it has long been used as a learning material in schools.

Hyakunin-Isshu cards are divided into two sets: one is called yomifuda, on which the entire poem is written with a picture, and the other is called torifuda, on which only the second half of the poem is written in Hiragana characters. Torifuda cards are spread out on a table or floor to play. The reciter selects one yomifuda card and reads the first half of the poem written on it. Players compete with each other to see how many torifuda cards are collected. To win the game, they have to memorize all 100 poems.

In our system, eight possible torifuda cards are presented on the table for a given yomifuda card, as shown in Fig. 5. If the student does not choose one of them within a certain time period, a sound cue is given at the position where the correct torifuda card appears. Because of uncertainty in recognition, the cue does not give him/her the exact position for the right card and works just as a "hint" in the trial.

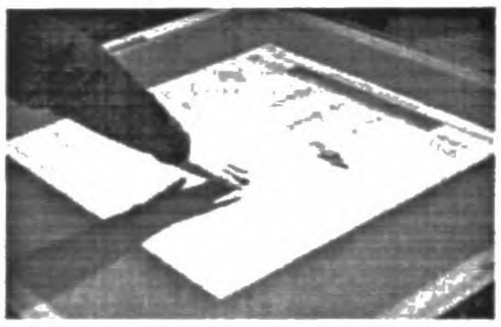

Fig. 5 Hyakunin-Isshu card application on the system

V. Gait Interface System

There are various trials to capture different human motions in existing gestural systems. However, a human is expected to stay within a limited area in his/her gesture demonstration. If we consider that the ability to walk upright is one of the hallmarks of being human [23], it is worth considering gestures with movements on the floor, or human gait, as a means to interact with a computer. Gait analysis research was performed in a variety of applications such as health diagnostics, rehabilitation, and sports [24]. In addition to such physical motion analysis, gait patterns convey complex states of mind, and there is a potential to develop a new application domain, where a human's emotion serves as a clue for interaction.

We have been conducting research in gait detection and analysis to recognize the physical and mental conditions of a human [25]-[27]. We adopted a floor mat sensor, the Xiroku LL sensor, as a motion capturing device, which adopts an electromagnetic induction mechanism. The device's sensing area size is 60 cm × 60 cm, whose spatial resolution is 10 mm. It is possible to connect up to 24 devices when a larger sensing area is needed. Figure 6 shows the snapshot of a working environment.

Fig. 6 System setup with three floor sensor devices

Footstep tracking is performed using a two-phase particle-filter-based technique, where Walker's alias method was applied to speed the execution [26], [27]. The system is capable of detecting and tracking footsteps of multiple persons who may wear shoes and identifying pairs of shoes (i.e., individuals) on the device. Furthermore, it is not necessary for them to walk in the same direction. They may turn in any direction and pass each other. Figure 7 shows a result of detecting footsteps.

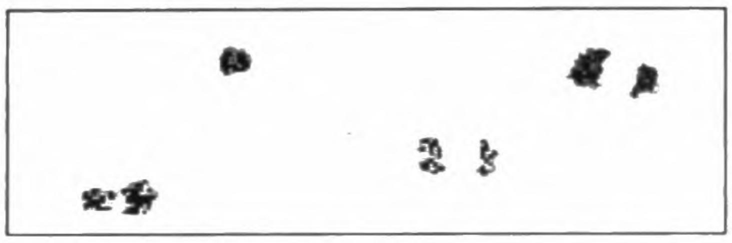

Fig. 7 Recognition of two walkers' footsteps

At present, we are conducting investigations on practical applications of this system. Healthcare and home security systems are required by an aging population to determine the state of health or aging of individuals by observing walking patterns. Moreover, it might be possible to detect a risk of falling, in addition to anticipating emergent adverse falling events. Furthermore, gait analysis can work in rehabilitation by helping the patient to compensate for deficits by understanding how well their rehabilitation program is progressing.

Another promising application of the system is digital signage, which could present specific contents to a pedestrian by considering his/her current emotion. Furthermore, existing digital signage systems (or control panels, in general) have been installed in many cases for adults of average build. Other groups of people, such as children, the elderly, or physically challenged individuals, may be forced to adopt an uncomfortable posture in such an interaction. If the system could recognize that the human is uncomfortable, it could change its interface adaptively, e.g., the position of texts, images, and selectable buttons, and size of fonts, so as to reduce the burden [25].

Figure 8 provides the ideal scenario at which content is chosen and presented in an optimum way; a guide map for an adult is presented at the upper position of the display, while an animated character is displayed at its lower position, adjusted for a child.

Finally, note that cameras are commonly used to detect human statuses and activities, but privacy is a concern. No one wants to be captured by cameras at home, even though sophisticated services can be expected. Our floor sensors do not collect privacy-sensitive characteristics data (e.g., face and body image). Moreover, unlike wearable sensor-based approaches [24], humans are not required to wear any special devices or clothing.

Fig. 8 Ideal application scenario in digital signage

VI. CONCLUSIONS

This study presented the idea of pleasant interfaces, which are preferable over invisible interfaces. A pleasant interface does not require humans to spend much effort to use the computer and operates in the background. In addition, it helps them do a thing which is worth doing. A framework to realize pleasant interfaces was described, in which technology and design can be coupled together to achieve a human-centered system containing three key elements: multimodal interactions, multimedia, and sensors. Three research trials were explained as examples.

Research on advanced user interfaces is essential to further promote world peace and trade using ICT, particularly in emerging regions, because it is an idea-centered discipline. We hope you have been inspired by human-centered computer development.

ACKNOWLEDGMENT

The author would like to thank the past and present students in his laboratory, who contributed to the development of the systems, especially Lasith Gunawardena, Koki Kimura, Kiryu Ibara, Yuya Shimomura, Ryuji Yamaguchi, Shinji Sugihara, Mika Oikawa, Mirai Oka, Hiroyuki Takai, Michihiro Ikeda, Norimasa Nagira, Takeshi Nakaie, and Chiemi Tanabe.

The work was partly supported by the Japan Society for the Promotion of Science (JSPS), Grant-in-Aid for Scientific Research (C) (23500149), and Japan Science and Technology Agency (A-STEP, Exploratory Research, AS231Z04815A).

REFERENCES

[1] C. Gibson, B. S. Folley, and S. Park, "Enhanced divergent thinking and creativity in musicians: A behavioral and near-infrared spectroscopy study," *Brain and Cognition*, vol.69, pp.162-169, 2009.

[2] G. Land and B. Jarman, "Breakpoint and beyond: Mastering the future today," *Leadership 2000*, 1998.

[3] K. Michael and K. W. Miller, "Big data: New opportunities and new challenges," *IEEE Computer*, vol.46, no.6, pp.22-24, 2013.

[4] M. Weiser, "The computer for the 21st century," *Scientific American*, vol.265, no.3, pp.94-104, 1991.

[5] D. Engelbart, "Augmenting human intellect: A conceptual framework," *SRI Summary Report*, AFOSR-3223, 1962.

[6] ACM Turing Award recipient, http://amturing.acm.org/award_winners/engelbart_5078811.cfm.

[7] G. Naylor, "The Bauhaus," London: Studio Vista, 1968.

[8] P. Konrad and T. Shanks, "Imprantable brain computer interface: Challenges to neurotechnology translation," *Neurobiology of Disease*, vol.38, pp.369-375, 2010.

[9] B. Goertzel and M. Ikle, "Special issue on mind uploading," *International Journal of Machine Consciousness*, vol.4, no.1, pp.1-3, 2012.

[10] S. Jordà, "The reactable: Tangible and tabletop music performance," *Proc. ACM CHI EA'10 Extended Abstract on Human Factors in Computing Systems*, pp.2989-2994, 2010.

[11] C. Tanabe and M. Hirakawa, "O-Key: A scoopable interface," *Proc. IEEE International Symposium on Multimedia*, pp.185-192, 2006.

[12] M. Ikeda, N. Nagira, and M. Hirakawa, "A multi-dip interface with water," *Proc. International Conference on Complex, Intelligent and Software Intensive Systems*, pp.169-176, 2009. [Revised *Journal of Information Processing*, vol.51, no.11, pp.2103-2111, 2010 (in Japanese)].

[13] K. S. L. Gunawardena, K. Kimura, and M. Hirakawa, "SensorTank: A water vessel for interaction," *Proc. 10th Asia Pacific Conference on Computer Human Interaction*, vol.2, pp.759-760, 2012.

[14] S. Moroi, S. Sasada, and R. Shibata, "The floating words for kids: Interactive installation for learning alphabet", *Proc. Conference on Computer Graphics and Interactive Techniques*, pp.115-116, 2002.

[15] N. Tosa and R. Nakatsu, "Interactive art for zen: Unconscious Flow", *Proc. IEEE International Conference on Information Visualization*, pp.535-540, 2000.

[16] Y. Shi, F. Ying, J. Yu, and C. Wei, "Tangible humidifier: Back to metaphor-based interaction," *Proc. 2010 International Conference on System Science, Engineering Design and Manufacturing Informatization*, pp.68-70, 2010.

[17] S. Mann, R. Janzen, and J. Huang, "WaterTouch: An aquatic interactive multimedia sensory table based on total internal reflection in water," *Proc. 19th ACM International Conference on Multimedia*, pp.925-928, 2011.

[18] J. Y. Han, "Low-cost multi-touch sensing through frustrated total internal reflection", *Proc. 18th ACM Symposium on User Interface Software and Technology*, pp. 115-118, 2005.

[19] M. M. Blattner, "Multimedia interface design", Addison-Wesley Publishing, 1992.

[20] T. Nakaie, T. Koyama, and M. Hirakawa, "Integrating spatial audio, visual, and gestural interfaces for supporting situated collaborative work," in *Ubiquitous Multimedia Computing*, CRC Press, pp.179-195, 2009.

[21] R. Yamaguchi, S. Sugihara, and M. Hirakawa, "Sound zone control in an interactive table system environment," *Proc. IEEE International Symposium on Multimedia*, 2011.

[22] R. Yamaguchi, A. Sakoi, and M. Hirakawa, "Multimedia framework for application of spatial auditory information to learning materials," in *Databases in Networked Information System*, Lecture Notes in Computer Science, Springer Press, pp.161-166, 2013.

[23] K. K. Hirst, "Bipedal Locomotion," http://archaeology.about.com/od/bterms/g/bipedal.htm.

[24] W. Tao, T. Liu, R. Zheng, and H. Feng, "Gait analysis using wearable sensors," *Journal of Sensors*, vol.12, pp.2255-2283, 2012.

[25] H. Takai, M. Oikawa, and M. Hirakawa, "Recognition of human gait using 2D continuous dynamic programming towards realization of invisible user interfaces," *Proc. 2010 IEEE Region 10 Conference*, pp.269-274, 2010.

[26] M. Oikawa, N. Tanijiri, and M. Hirakawa, "Estimation of lower body motion during walking on a floor sensor device," *ICIC Express Letters*, vol.6, no.12, pp.3027-3031, 2012.

[27] K. Ibara, K. Kanetsuna, and M. Hirakawa, "Identifying individuals' footsteps walking on a floor sensor device," *Proc. 2013 International Conference on Active Media Technology*, 2013 (to appear).