Practical Use of a Remote Movable Avatar Robot with an Immersive Interface for Seniors

Masahiko Izumi[1], Tomoya Kikuno[1], Yutaka Tokuda[2], Atsushi Hiyama[1], Takahiro Miura[1], and Michitaka Hirose[1]

[1] Graduate School of Information Science and Technology, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8656, Japan
{masa,kikuno,atsushi,miu,hirose}@cyber.t.u-tokyo.ac.jp
[2] Graduate School of Engineering, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan
ytokuda@cyber.t.u-tokyo.ac.jp

Abstract. Societal aging is an inevitable social problem in many developed countries. In order to manage this issue, it is necessary to drastically change the welfare security and labor system of relevant societies. We have proposed "mosaic-type work" in which a single "virtual worker" is synthesized based on individual workers and seniors through seamless information sharing. In this paper, we specifically focus on the spatial part of the mosaic, namely, the "spatial mosaic," for enabling seniors to work without the burden of movement. Towards the actualization of this concept, we aimed to develop senior-friendly mobile avatar robots that can fit into seniors' daily lives. First, we analyzed problems of seniors' telecommunication with telepresence robots, and two elemental interfaces, the "physically operable interface" and the "acoustic zooming interface." The former is a senior-friendly interface that the seniors can manipulate with their motion, and the latter enables the user to listen to sounds in a specified area that the user can adjust. In order to discuss integrated interface designs, we conducted two exploratory experiments to evaluate the performance of these systems.

Keywords: Mosaic-type work, seniors, information communication technologies (ICT), avatar, interface.

1 Introduction

Population aging is one of the most significant global issues due to decreasing domestic productivity and increasing social security costs. In Japan, seniors accounted for 24.1% of the population in 2013 [1], a number estimated to reach 40.5% by 2055 [2]. The ratio of the young population to the elderly in 2009 was 2.81, which is expected to decline to 1.26 by 2055 [2],thus making it unreasonable to continue the conventional welfare security model, in which multiple young people support one senior. However, 83.6% of all seniors in Japan are healthy enough to not require care [3]. More than half of them are aware of the social contributions they have made and retain a strong desire to work [4]. These

C. Stephanidis and M. Antona (Eds.): UAHCI/HCII 2014, Part III, LNCS 8515, pp. 648–659, 2014.

facts indicate that the present social system needs to be—and can be—changed drastically.

According to a report by Fukushima, seniors have four types of work needs: to work without strenuous effort, to make themselves useful to others, to build personal relationships, and to earn extra money [5]. In particular, ex-white collar workers work for social connections and health enhancement [6]. They have knowledge, experience, and skills that young people do not have. Effective utilization of their ability can gradually change the conventional social system and revitalize work environments. The Silver Human Resource Centers in Japan attempt to meet the various needs of seniors [7]. However, it is difficult to meet all of their needs and effectively allocate seniors to the workforce, because these centers are chronically short-handed and limited to allocating work within their designated area.

Thus, we proposed the concept of a "mosaic-type work" system in which a single "virtual worker" is synthesized from individual work resources using information and communication technologies (ICTs) to supply an autonomous and stable labor force [8,9]. In order to actualize this concept, it is necessary to consider and quantify the workers' time, skills, spatial limitations, and other characteristics, as well as to develop a user-friendly interface for seamless information sharing. In this paper, we specifically describe our development and evaluation of "spatial mosaic" as an adapted model for mosaic-type work in senior employment contexts, with the purpose of exceeding seniors' limitations in daily available work area.

Our paper first discusses concepts pertinent to mosaic-type work and the spatial mosaic system (Section 2), followed by related work regarding telecommunication technologies and their problems (Section 3). Then, we provide an overview of a proposed telecommunication system along with the concepts of social telepresence and acoustic zooming (Section 4). We implemented a prototype interface and conducted two exploratory experiments for these two concepts (Section 5). With these results, we discuss the design of an integrated interface (Section 6).

2 Mosaic-type Work and Spatial Mosaic

2.1 Mosaic-type Work

Mosaic-type work can ensure compatibility between flexible work opportunities for seniors and a stable work force for employers by synthesizing virtual workers from a workforce of seniors. A "mosaic" is defined as a unit of one synthesized worker from multiple workers, as shown in Fig. 1. In this figure, as an example, senior A has expertise in field A, but not enough skill in field B, and also has mild motor and cognitive impairment. Senior B also has motor and cognitive difficulties in addition to insufficient skill in field A, but is highly skilled in field B. A young worker has good motor and cognitive functions, but is poor in fields A and B. Ideally, when their advantages can be synthesized into a virtual worker,

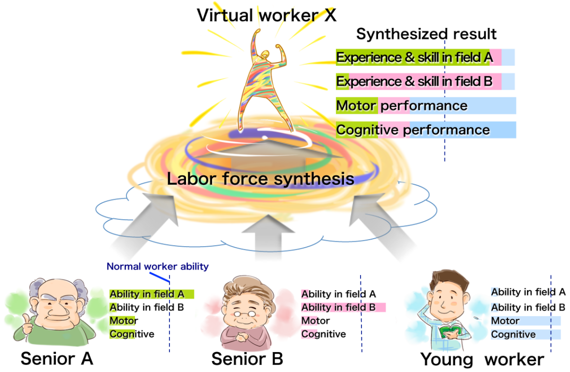

Fig. 1. The concept of "mosaic-type work" that generates a stable virtual worker from a workforce of seniors and young people

this virtual worker can demonstrate stable and superior performance in fields A and B.

In order to actualize this concept, it is necessary to consider and quantify the workers' time, skills, spatial limitations, and other characteristics. In this study, we focused specifically on the generation of a "spatial mosaic."

2.2 Spatial Mosaic

A spatial mosaic involves gathering remote workers in workplaces through ICTs to create a virtual workforce concentration. This concept is illustrated by Fig. 2. In this concept, remote workers can access workplaces in both the virtual and real environments, regardless of the worker's location. In the case of virtual workplaces, as shown in the left part of Fig. 2, various jobs are divided into task elements, and then are allocated to remote workers. This procedure results in co-operative work among a massive number of remote workers. On the other hand, real workplaces, as illustrated in the right part of Fig. 2, are realized by remote systems. In this case, workers remotely access the remote robots installed at the workplaces, and then conduct their activities. Current remote work systems include teleconference systems.We assume that telepresence robots can overcome the movability limitation on remote workplaces and enable remote workers to carry out various business activities. The robots can be useful to smoothly migrate from conventional work styles, in which workers gather in a particular place and work, to the proposed work style. Our concept can also be employed in the workplace using a combination of real and virtual workplaces.

The style of work created by the spatial mosaic enables seniors with movement problems to work more actively. Crowdsourcing is one relevant example that institutes the framework shown in the left part of Fig. 2. Kobayashi et al. reported a crowdsourced proofreading system of digital books for people with visual impairments [10]. In this platform, senior volunteers conducted phrase, ruby, and

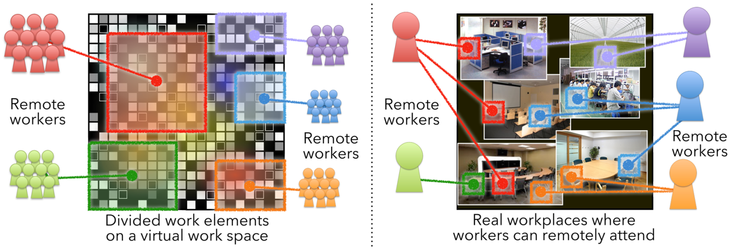

Fig. 2. The concept of a spatial mosaic that overcomes the limitations of real workplaces. [Left]: A workplace is generated in a virtual space. Jobs in the virtual workplace are divided into work elements, and then remote workers complete these tasks in the virtual environment. [Right]: Workplaces are located in real space. Workers can remotely access multiple workplaces in the real environment.

character corrections of the scanned books that were processed through optical character recognition. On the other hand, in order to implement the framework illustrated by the right part of Fig. 2, it is necessary to leverage various support systems for remote workers. Conventional teleconference systems are effective when the participants concentrate on speaking and do not move. The current telepresence robot is useful for moving smoothly in remote real spaces, but remains in the research and development phase. This telepresence robot, released by Double Robotics, is a reasonable robot that can be easily manipulated. However, for seniors, this kind of system should be equipped with an intuitive interface that does not disturb regular communication and creates highly realistic sensations in remote environments for promoting mutual understanding.

The ultimate objective of this study was to develop a support system for enabling seniors to remotely work on the real environment. In this paper, specifically, we focus on a mobile avatar robot telepresence system for remote work, and then propose a manipulation interface for promoting smooth remote communication.

3 Related Work on Telecommunication Technologies

In this section, we discuss relevant work on conventional telecommunication technologies and summarize the problems for implementing the spatial mosaic society-wide.

3.1 Mobility in Telecommunication

In the above section, we discussed the "spatial mosaic" concept, which generates a virtual workforce by gathering remote workers virtually within real and virtual workplaces through ICT. When the spatial mosaic is implemented in our society, the forms of spatial mosaic and its issues are summarized as follows:

1. **Participation in real-world workplaces:** To insert remote workers into the real world using ICT. It is important that the virtual worker integrated into real world appropriately. If there are any abnormalities, these can act as a distraction for on-the-spot workers. Examples: remote lecture, remote interview.
2. **Meeting in a particular space:** From distant locations, many people can meet in a particular place simultaneously It is important that on-the-spot information is correctly interpreted. Furthermore, equal attention to various information is also an important factor. Example: disaster site.
3. **Connect remote spaces in the virtual world:** From distant spaces, many people meet in a virtual world simultaneously. All participants are virtualized; therefore, it is important to create new and appropriate communication methods. Example: remote meeting.

As the first approach to the spatial mosaic, we focused on the participation model, which is the simplest model of telecommunication. Our first step was to insert one remote person into the real world without any abnormalities. Of course, this is a critical factor for telework towards maintaining fluent communication between remote and on-the-spot individuals. For interactions with on-the-spot workers, conventional television conference systems have limitations as an information medium. These technologies can only convey sound and video. They cannot express extensity, which is an important component of face-to-face communication. Therefore, through television conference systems, it is difficult to gauge the physical and psychological distance between on-the-spot workers and remote workers. Thus, proper communication between them is repressed.

If the remote user can move around freely in the remote office, there are benefits of the communication context for the remote user as follows:

– look around the remote space freely
– get up close with on-the-spot workers
– search for static information in the remote space

In addition to these merits for the remote user, there are certain benefits for the on-the-spot worker.

– estimate what the remote user is interested in
– no need to move toward the remote user
– extend presence of the remote user

For these reasons, we focused on avatar robots as a medium of telecommunication. We use "avatar robot" to describe a robot that can freely navigate the real world freely and convey auditory and visual information. In the following section, we discuss communication through avatar robots.

3.2 Telecommunication through Avatar Robots

The telepresence robot is a representation of oneself that enables a user to be anywhere and have an artificial face-to-face conversation with remote people without considering spatiotemporal or economic constraints. Research on telepresence robots began more than ten years ago, and recently, a variety of telepresence robots have been produced by various enterprises [11,12]. Telepresence robots came into widespread use for teleconferences, remote lectures, telemedicine, and so forth [13]. For example, InTouch Health and iRobot developed a telepresence robot RP-VITA for telemedicine [14]. By using RP-VITA, a doctor can ask hospitalized patients detailed questions about their symptoms while sitting in his own home or on a business trip. Thus, telerobotic technology has been rapidly developed. There are currently numerous low-priced products and program libraries for developing telepresence robots.

Under these circumstances, for real-world use—not in controlled conditions—we regard small mobile avatar robots as the most suitable robot for telecommunication. Therefore, we designed a telecommunication method using a mobile avatar robot.

3.3 Problems of Conventional Avatar Robots

Mobile robotics technologies progress daily. There are many considerations involved for using such technology in the real workforce (e.g., remote lectures). We conducted several remote lectures using a mobile avatar robot, named double (double robotics) and extracted the following problems.

- narrow eyesight
- operation complexity
- lack of presence
- difficulty of sound localization
- inviolable area

The narrow eyesight problem will be solved by hardware improvements. Therefore, we will not discuss this problem further. Operation complexity and lack of presence facilitated unsatisfactory social telepresence, which caused ineffective communication. The definition of "social telepresence" is discussed in the following section. Difficulty of sound localization and inviolable area appear to be independent problems, but if we use selective information browsing in remote space, these problems can be solved simultaneously. We call this selective browsing idea 'zooming.' This idea will be discussed further in the next section.

4 Interface Concepts

4.1 Social Telepresence

Social telepresence is defined as the level to which face-to-face communication is simulated. Social telepresence is high when the user feels as though they are in a face-to-face situation. This allows the user to experience more natural talk. Generally speaking, social telepresence become higher with visuals as opposed to sound alone. Returning to the conventional avatar problems, operation complexity is relevant to remote users, whereas lack of presence is relevant to on-the-spot users, which is experienced as a decline in social telepresence. If we can resolve these problems, mobile avatar robots can be the medium to facilitate improved social telepresence over visuals.

First, we discuss operability improvements. A complex operation interface forces users to concentrate exclusively on robot operation, hindering communication with the remote speaker. This issue is akin to putting the cart before the horse. The robot operation interface should be easy to understand and as intuitive as possible. To simplify the operation interface, we propose using body motion as the operation input. With a body motion operation system, in addition to simplifying operation, we believe that users will feel as though their individuality is better projected into the remote avatar robot. This increases the feeling of being in the remote place, which will thereby improve social telepresence.

Second, we describe presence enhancement. In face-to-face communication, nonverbal movement plays an important role in expressing the delicate nuances of thought. For example, eye gaze represents subtle nuances that cannot be expressed by means of language. With these intricate movements, humans infer customer reactions based on their words, movements, and so forth. In other words, customer movements represent the output based on our social input. With a conventional robot operation system, robots do not move unless users operate the robot with intention. Therefore, when remote users become heated in conversations with on-the-spot workers, the user's attention is directed toward speech alone. Attention is directed away from robot operation, such that robots do not show any movement. Under this condition, on-the-spot workers cannot detect any reflection of remote users' true mood, except for sound and visuals. Therefore, it is difficult for on-the-spot workers to know whether remote users can listen to or see on-the-spot information. Communication is reduced by reduced reflection of customer's feelings.

Therefore, we propose an interface that shows the remote user's unconscious movement to on-the-spot workers. By reflecting the operator's subtle movements in the robot's movements, we can simulate nonverbal aspects of communication. We believe that this increases feelings of robot presence among on-the-spot workers, and thus results in the improvement of social telepresence.

4.2 Acoustic Zooming

The word "zoom" means to magnify a section of a picture. Many studies have been conducted to support information understanding based on zooming effects. In this study, the word "zooming" means paying attention to a localized point and thereby understanding its characteristics. In particular, "acoustic zooming" is defined as the zooming of acoustic information.

Fig. 3 illustrates a schematic concept of acoustic zooming and an acoustic cubic volume. In Section 3, we discussed the difficulty of sound localization and inviolable areas as problems of conventional mobile avatar robots. We believe that acoustic zooming can solve these problems simultaneously.

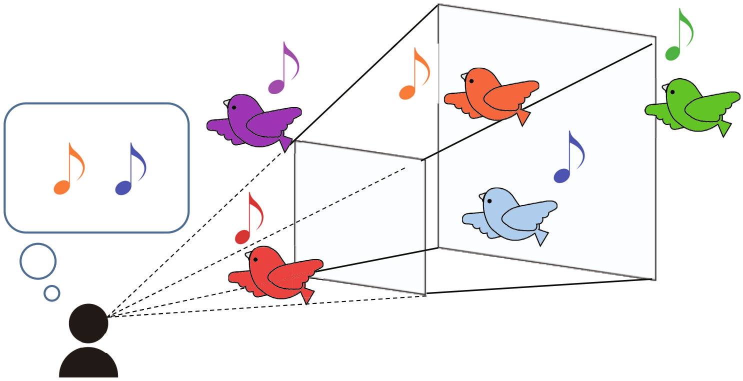

Fig. 3. The concept of acoustic volume. User can only hear sounds within the acoustic volume. In this case, users can hear the voices of orange and blue birds.

First, we discuss the sound localization problem. In the real world, there are many kinds of sound sources. Therefore, the sounds sent to remote users through avatar robots are mixed up with other sounds within the environment. From the perspective of communication, general sound—aside from the customer's voice—is primarily noise, which should be reduced. By targeting the talking partner by way of acoustic zooming, remote users can easily concentrate on talking. Furthermore, acoustic zooming can create a pseudo-cocktail party effect in terms of selective listening. Since the cocktail party effect is a unique effect under the context of face-to-face communication, this effect can enhance feelings of face-to-face communication, thus improving social telepresence.

Second, we discuss the inviolable area problem. In the real world, there are many immobile objects (e.g., posters on the wall and large, heavy monuments). Owing to mobile avatar robots, remote operators can move closer to and observed immobile information. However, the real world is not always suitable for mobile avatar robots. Thus, the area in which robots can move around is limited. For example, using Double [11], a small item in its route can be an obstacle for proper movement. Furthermore, conventional mobile avatar robots aim to enable users to engage in face-to-face-like talk with remote individuals. Thus, the main camera, which primarily streams the partner's face to the remote user, is typically placed level with a human face with a narrow field of view. Therefore, if there is a low-height obstacle, it eludes the avatar's eyesight, which causes the avatar to bump into the obstacle, thereby prohibiting forward movement. Cluttered routes are therefore unsuitable routes for mobile avatar robots. However, there might be intriguing information in the inviolable areas. Under these conditions, users can observe interesting information without any movement using acoustic zooming. This is a natural condition in face-to-face interactions. For example, when users are spoken to from a distance and there are obstacles between the conversational partners, users naturally react to their partner in face-to-face situations. However, with a mobile avatar, it is difficult to converse from a distance. By using acoustic zooming to target a conversational partner, users can converse from a distance.

5 Exploratory Experiments

In this section, we conducted fundamental experiments on the communication-promoting effects of our ideas.

5.1 Social Telepresence

First, we again list factors for the enhancement of social telepresence.

- physically operable interface
- unconscious movement synchronization

We implemented an interface that can have these features for robot operation on a tablet computer. There are two reasons for using a tablet computer. First, tablet computers have sufficiently large monitors to enable experience of authentic audio-visual sensations. Second, tablet computers provide several easy physical operations, such as pinch and rotate operations. These gestures provide more intuitive manipulation and interaction than simple buttons. Holding a tablet computer and wearing headphones, users can explore various objects in the virtual space. The system enables them to experience virtual space as if they were looking through the display.

As a physically operable interface, we introduced the operator's direction into the avatar manipulation interface as shown in Fig. 4. Operators could move the robot by tilting the tablet back and forth, and turn the robot by turning the tablet right and left. The concept of this interface was designed to reflect users' spatial movements into robot movements. In addition to this manipulation method, we implemented a mechanism for reflecting unconscious movements of users. We extracted hand jiggles, which constituted high-frequency data measured by an accelerometer with a high-pass filter. Back-and-forth jiggles were measured for back-and-forth movement, and right and left jiggles were measured for clockwise and counterclockwise turns, respectively. Drastic movements occurred when jiggles induced feelings of strangeness, and thus, we utilized limited threshold gap values.

Using this system, we conducted interface evaluation experiments. Participants were asked to conduct short presentations through the avatar robot. Then, we administered a brief questionnaire on the presenter and listener to understand participants' impressions. Among presenters, feelings of self-projection were enhanced by physically movable operation. This indicates that the feelings of presence in the remote place were also enhanced. Among listeners, presence of presenter was enhanced by the reflection of the presenter's unconscious movements. This fact indicates that listeners paid more attention to presenters when there was unconscious movement projection. Thus, it appears that, under these conditions, social telepresence was enhanced bidirectionally.

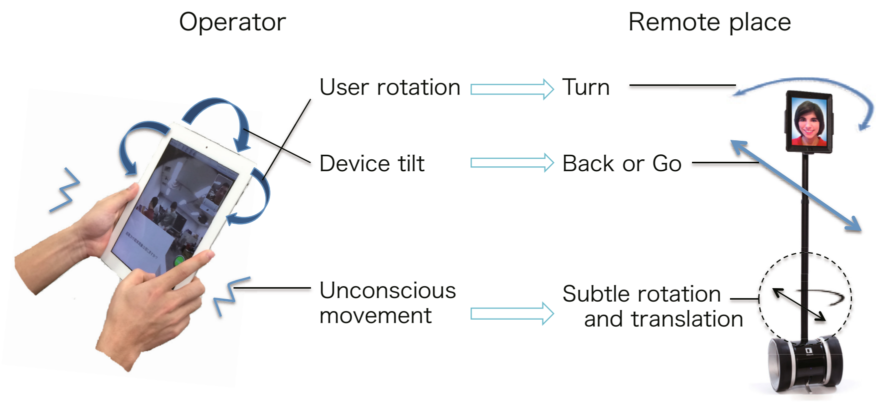

Fig. 4. Design of the physically operable interface. Users can move the avatar robot by tilting the tablet and turn the robot by user rotation. The system senses subtle movements—hand jiggles—and reflects these in the robot as unconscious, nonverbal movements of the operator.

5.2 Acoustic Zooming

Regarding acoustic zooming, we already have conducted a plain experiment [15]. We have also implemented an acoustic zooming system on a tablet computer. With the system, we conducted an experiment to evaluate the efficiency of acoustic zooming. To compare the effect of acoustic zooming with conventional interfaces, we implemented three volume calculation modes, namely, "normal," "direction," and "zooming." Fifteen sound sources—voices of different words—were recorded in advance. These were located within the virtual space, and participants were asked to find specific words.

The zooming mode produced significantly higher performance than the direction mode and significantly more accurate responses than the normal mode. According to the participants' comments, switching functions of the calculation modes and resizable acoustic volumes are important implementations for more effective browsing in various situations.

For these reasons, acoustic zooming is effective as a type of information interaction. Therefore, it is implied that avatar interaction with acoustic zooming can diminish stress related to information seeking in the remote environment. From the perspective of mobile avatar manipulation, it is important that the operation interface provides operation of robot movements and targeting of acoustic zooming, simultaneously. Notwithstanding substantive robots, it is unrealistic to produce fast-moving robots due to safety concerns. If the target is constantly moving, chasing it manually is difficult for these robots. Therefore, when implementing a zooming system into the avatar operation interface, the system should lock and chase the target automatically. This saves the user from unnecessary concentration.

6 Discussion

In this section, we discuss our next steps, namely, integration of the physical UI and acoustic zooming. These two methods are both effective in building an avatar operation interface. Thus, if we can effectively integrate these methods, our avatar operation interface can have enhanced long-distance communication abilities.

The most significant issue in need of consideration is a mapping method between user action and robot motion. This time, we mapped user's body motion to robot movement, and the pinch motion on a tablet surface for acoustic zooming. However, motions that are most suitable for specific robot motions should be researched. In particular, operation preferences might differ across generations. For example, younger generations, who are proficient with smartphone usage, might be accustomed to using a touch display interface. However, older generations are less likely to be accustomed to this. In this case, the young have a mental model for using touch-panel devices, but the elderly do not. The interface should be easy to make using image for the user who have no proper mental model.

The most appropriate interface will change dynamically change across scenes. The purpose of the present study, however, was to evaluate our ideas. Thus, we did not focus on particular work scenes. Hereafter, we will aim to use mobile avatar robots in a daily life context in order to clarify patterns of work using the spatial mosaic method, and design an interface for robot operation in specific scenes.

7 Conclusion

In order to make full use of the strengths of the senior workforce, such as its rich knowledge, experience, and skill, the spatial mosaic formation system and its fundamental technologies were discussed. We discussed the mosaic and spatial mosaic systems as well as conventional avatar robot communication problems. We proposed two interaction concepts to resolve the stated problems: "physically movable operation interface" and "acoustic zooming interface." We conducted an exploratory experiment for each interface. Results verified the efficacy of these interfaces. The physically movable operation interface enhanced social telepresence as experienced by the presenter and felt by the listener. This indicates that body movement is an important factor, as we predicted in Section . The acoustic zooming interface is useful for sound information browsing. We plan to conduct a similar experiment in real-world scenarios. Future studies will integrate these two concepts into one interface for avatar robot operation. Then, using this interface, our research target will be the difference in feelings and proficiency between the young and the aged.

Acknowledgments. This research was supported in part by S-innovation (Strategic Promotion of Innovative Research and Development) funding under Industry Academia Collaborative R&D Programs administered by the Japan Science and Technology Agency (JST).

References

1. Cabinet Office: Annual Report on the Aging Society 2013, http://www8.cao.go.jp/kourei/whitepaper/w-2013/gaiyou/s1_1.html (in Japanese) (last viewed February 7, 2014)
2. Kaneko, R., Ishikawa, A., Ishii, F., Sasai, T., Iwasawa, M., Mita, F., Moriizumi, R.: Population Projections for Japan: 2006-2055 Outline of results, methods, and assumptions. The Japanese Journal of Population 6(1), 76–114 (2008)
3. Ministry of Health, Labour and Welfare: Change in the number of people who are certified to be in need of long-term care by care category, http://www.mhlw.go.jp/seisaku/2009/03/01.html (in Japanese) (last viewed February 26, 2013)
4. Cabinet Office: White Paper on the National Lifestyle 2006, http://www5.cao.go.jp/seikatsu/whitepaper/h18/06_eng/index.html (last viewed February 26, 2013)
5. Fukushima, S.: Analysis of desires to work with elderly people (this translated title was prepared by the authors). Japan Labor Review 558, 19–31 (2007) (in Japanese)
6. Japan Organization for Employment of the Elderly, Persons with Disabilities and Job Seekers: Report of an investigation on the work and life of Japan's baby boomers (this translated title was prepared by the authors), http://www.jeed.or.jp/data/elderly/research/dankai22.html (in Japanese) (last viewed February 26, 2013)
7. Osada, H., Suzuki, T., Takata, K., Nishishita, A.: Social activity and relevant factors in the elderly: Focus on members of senior manpower center and senior citizen's club. Japanese Journal of Public Health 57(4), 279–290 (2010)
8. Hiyama, A., Sano, M., Kobayashi, M., Hirose, M.: Senior Cloud: ICT platform for improving society by circulating the experience, knowledge, and skills of the elderly. In: Proc. 17th VRSJ 2012, pp. 157–160 (2012) (in Japanese)
9. Nakayama, M., Hiyama, A., Miura, T., Yatomi, N., Hirose, M.: Support system of time-mosaic formation in flexible work styles for seniors. IPSJ Journal 55(1), 177–188 (2014) (in Japanese)
10. Kobayashi, M., Ishihara, T., Itoko, T., Takagi, H., Asakawa, C.: Age-based task specialization for crowdsourced proofreading. In: Stephanidis, C., Antona, M. (eds.) UAHCI 2013, Part II. LNCS, vol. 8010, pp. 104–112. Springer, Heidelberg (2013)
11. http://www.doublerobotics.com/
12. https://www.suitabletech.com/beam-plus/
13. Tsui, K.M., Desai, M., Yanco, H.A., Uhlik, C.: Exploring use cases for telepresence robots. In: 2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 11–18. IEEE (2011)
14. http://www.intouchhealth.com/products-and-services/products/rp-vita-robot/
15. Izumi, M., Hiyama, A., Miura, T., Tokuda, Y., Hirose, M.: Acoustic information zooming interface. In: Proc. of Asiagraph, vol. 8 (2013)