International Journal of Informatics Society

08/20 Vol.12 No.1 ISSN 1883-4566

Informatics Society

Aims and Scope

The purpose of this journal is to provide an open forum to publish high quality research papers in the areas of informatics and related fields to promote the exchange of research ideas, experiences and results.

Informatics is the systematic study of Information and the application of research methods to study Information systems and services. It deals primarily with human aspects of information, such as its qu ality and value as a resource. Informatics also referred to as Information science, studies t he structure, algorithms, behavior, and interactions of natural and a rtificial systems that store, process, access and communicate information. It also develops its own conceptual and theoretical foundations and utilizes foundations developed in other fields. The advent of computers, its ubiquity and ease to use has led to th e study of info rmatics that has computational, cognitive and social aspects, including study of the social impact of information technologies.

The characteristic of informatics' context is amalgamation of technologies. For creating an informatics product, it is necessary to integrate many technologies, such as mathematics, linguistics, engineering and other emerging new fields.

Guest Editor’s Message

Keiichi Abe

Guest Editor of Thirty-fourth Issue of International Journal of Informatics Society

We are delighted to have the Thirty-fourth issue of the International Journal of Informatics Society (IJIS) published. This issue includes selected papers from the Twelfth International Workshop on Informatics (IWIN2019), which was held at Hamburg, Germany, Sept. 8-11, 2019. The workshop was the thirteenth event for the Informatics Society, and was intended to bring together researchers and practitioners to share and exchange their experiences, discuss challenges and present original ideas in all aspects of informatics and computer networks. In the workshop 25 papers were presented in seven technical sessions. The workshop was successfully finished with precious experiences provided to the participants. It highlighted the latest research results in the area of informatics and its applications that include networking, mobile ubiquitous systems, data analytics, business systems, education systems, design methodology, intelligent systems, groupware and social systems.

Each paper submitted IWIN2019 was reviewed in terms of technical content, scientific rigor, novelty, originality and quality of presentation by at least two reviewers. Through those reviews 20 papers were selected for publication candidates of IJIS Journal, and they were further reviewed as a Journal paper. We have three categories of IJIS papers, Regular papers, Industrial papers, and Invited papers, each of which were reviewed from the different points of view. This volume includes seven papers among those accepted papers, which have been improved through the workshop discussion and the reviewers’ comments.

We publish the journal in print as well as in an electronic form over the Internet. We hope that the issue would be of interest to many researchers as well as engineers and practitioners over the world.

Keiichi Abe received his Ph.D in informatics from the Shizuoka University and a master's degree in Electrical Engineering from the Nippon Institute of Technology. Currently, he is professor of the Faculty of Creative Engineering, Kanagawa Institute of Technology. He has been working at development of semiconductor manufacturing equipment maker for about 6 years. After that, he has taught at the Organization for Employment of the Elderly, Persons with Disabilities and Job Seekers (vocational training) for 15 years. He has been working senior professor of the Polytechnic College until 2015. He changed to the Kanagawa Institute of Technology as associate professor in 2016. His research interests include Smart System using IoT Technology, Interactive System using Robot and CG, Victims Information Management System at Large Scale Disaster and Smart House. He is a member of IPSJ, IEICE, ACM and IEEE.

<u>**Invited Paper**</u>

Psychoacoustic Sonification as User Interface for Human-Machine Interaction

Tim Ziemer†, Nuttawut Nuchprayoon*, and Holger Schultheis†‡

†Bremen Spatial Cognition Center, University of Bremen, Germany
‡Institute for Artificial Intelligence, University of Bremen, Germany
* Faculty of Veterinary Science, Mahidol University, Thailand
{ziemer, schulth}@uni-bremen.de, nuttawut.nuc@mahidol.edu

Abstract - When operating a machine, the operator needs to know some spatial relations, like the relative location of the target or the nearest obstacle. Often, sensors are used to derive this spatial information, and visual displays are deployed as interfaces to communicate this information to the operator. In this paper, we present *psychoacoustic sonification* as an alternative interface for human-machine interaction. Instead of visualizations, an interactive sound guides the operator to the desired target location, or helps her avoid obstacles in space. By considering psychoacoustics — i.e., the relationship between the physical and the perceptual attributes of sound — in the audio signal processing, we can communicate precisely and unambiguously interpretable direction and distance cues along three orthogonal axes to a user. We present exemplary use cases from various application areas where users can benefit from psychoacoustic sonification.

Keywords: Human-Machine Interaction, Human-Computer Interaction, Navigation, Audio Interface, Auditory Display, Psychoacoustic Sonification

1 INTRODUCTION

Human-machine interaction (HMI) or human-machine collaboration is defined as coordinated interaction of a human and a machine that requires complex sensor-motor control abilities [18]. Simply put, HMI is a cycle with three stations as illustrated in Fig. 1. A human operator manually controls a machine via a Human Interface Device (HID), which modifies the location and/or orientation of the machine, and thereby its spatial relation to targets or obstacles in the environment. This modification is sensed by the machine's sensors. The sensor data is communicated to the operator via a User Interface (UI). Traditionally, UIs are visual displays. However, in this paper we introduce *psychoacoustic sonification* as an auditory UI for HMI.

Compared to conventional forms of sonification, the strength of psychoacoustic sonification is that multivariate or multidimensional data can be communicated to a user interactively via sound in an unambiguous way with high precision, low technical demands and low latency. This makes it suitable for situations in which vision is limited, or in which the visual scene is overloaded. Throughout this paper we will highlight the benefits of psychoacoustic sonification and present exemplary use cases in which these benefits become effective.

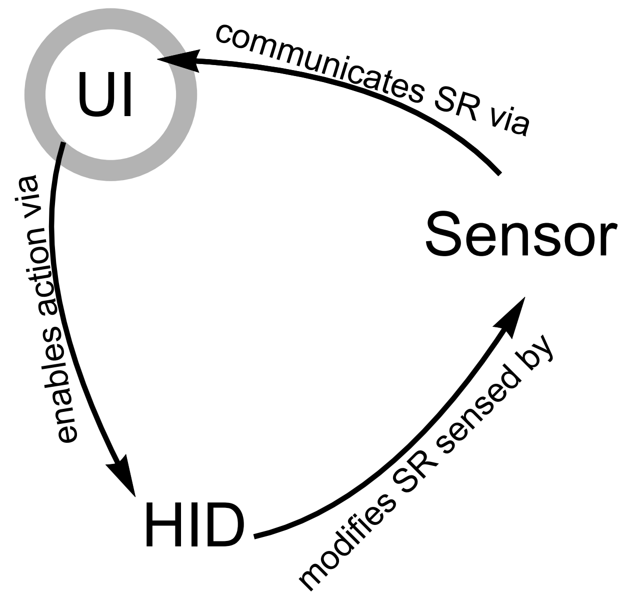


Figure 1: Cycle of HMI: Sensors sense information about Spatial Relations (SR), communicate them via User Interface (UI), which enables the operator to take appropriate action via Human Interface Device (HID), which again modifies the SR as sensed by the sensors. This paper deals with psychoacoustic sonification as a UI.

2 HUMAN-MACHINE INTERACTION

As illustrated in Fig. 1, Human-Machine Interaction (HMI) [18] is a cycle with three stations. Human interface devices (HIDs) include computer mouse and keyboard, gamepads, touchpads and touchscreens, knobs and sliders, joysticks, eye-tracking, voice control and gesture control via visual object recognition, motion tracking with optical markers or electromagnetic fields. These serve to operate a machine, vehicle, vessel, spacecraft, aircraft, missile or robot in real, virtual or mixed reality environments.

The machine is equipped with sensors that sense spatial relations between the machine and some objects of the outside world. Sensors for the outside world include radar, lidar, active and passive sonar, electromagnetic sensors, passive or active monoscopic or stereoscopic (infrared) cameras and, e.g., computer vision or motion capture systems [11]. Sensors for the location and orientation of the machine itself include tachometer, inclinometers, Hall sensors, accelerometers, gyroscopes, global navigation satellite system (GNSS), compass, or combined Inertial Navigation System (INS) [15]. Sensor data may be processed and interpreted by a computer processor and communicated to the operator via a User Interface (UI).

Typical UIs are visual displays [6], like lamps, seven segment displays, and monitors that show camera recordings, graphs, maps, charts, depictions, icons, text, crosshairs, meter needles, and/or tables, etc. Auditory displays [26] are less common UIs [7, 21]. They represent data, e.g., by means of speech, auditory icons, earcons, and the core element: sonification.

3 AUDITORY DISPLAYS

Auditory Displays (ADs), sometimes referred to as *tactical audio*, are sounds that inform a user during HMI. In that sense they serve as UIs that can communicate sensor data to the operator. An overview can be found in [16, 26]. ADs can have multiple elements.

Auditory icons are sounds that mimic a real-world sound that is causally or symbolically related to the represented data. For example the sound of slamming a door can represent closing a file. Their major advantages are that they can be intuitively understood and that they can be easily recorded or synthesized. The main drawbacks are that they carry only a very limited amount of discrete information, i.e., nominal scale of measure.

In contrast to auditory icons, *earcons* are the combinations of discrete sounds to short sound events that are not derived from the real world. For example some computer operating systems play a certain melody when attaching a device via USB, and a similar melody with the same instrument when detaching a device. The major advantage of earcons is that they can carry more information than auditory icons, like hierarchies, i.e., ordinal scale of measure. Their drawbacks are that their meaning needs to be learned, no continuous information can be presented and overlapping earcons may be indiscriminable.

Speech is self-explaining. Spoken words or sentences inform the user about the data. Famous examples of speech as a UI are screen readers. The major benefit of speech is that even complicated information and relations can be communicated. The most crucial drawbacks of speech are that it takes much time, and that it is cognitively demanding to interpret language. Only a small number of data points per second can be expressed by words, no continuous data streams.

In contrast to earcons and speech, *sonification* is not discrete in nature. It can therefore communicate continuous data. A famous example of sonification is auditory pulse oximetry, where the oxygen concentration in the patient's blood is communicated to the anesthesiologist by means of the pitch of a complex tone that is triggered every heartbeat. The main advantage of sonification is that it can potentially present the highest density of information. It can carry much more information than auditory icons and earcons, much faster than speech. Sonifications can represent data on an interval or even ratio scale of measure. The main disadvantage is that the meaning of each individual sonification needs to be learned by the user. Furthermore, sonifying a high amount of data often lead to ambiguous or confusing results.

Auditory displays are typically employed in situations in which visual displays are (partly) occluded or lie outside the visual focus or completely out of the visual field [7, 21]. Furthermore, they are useful when vision is limited, e.g., due to environmental conditions, like darkness, fog, smoke, muddy waters, sandstorm. They also complement visual displays to overcome shortcomings of visual UIs, like the missing depth on monitors, a visual overload due to too many depictions, or missing detail, e.g., due to graphics overlay. Auditory displays can be useful alternatives to graphical displays for visually impaired or blind users. Auditory displays are also useful in situations in which visual perception is overloaded, e.g., by the environment or by a high number of visual displays, or when the visual channel is needed for tasks other than HMI, like exploring the environment. Last but not least, reaction times to sound are typically much shorter compared to visual cues [20] which is why audible alarms [12] are ubiquitous. Not only can auditory displays serve as an alternative to visual displays, they can also complement or augment visualization in terms of a multi-modal user interface.

4 PSYCHOACOUSTIC SONIFICATION

Sonification is the communication of data by means of sound. The translation from input data to output sound is called *mapping*. In contrast to auditory icons, earcons, and speech, sonification is able to communicate one continuous data stream. Hence, it helps for data monitoring of stock markets and vital functions, and for navigation aid in driving, piloting and surgery [26, 40].

Multivariate or multidimensional sonification can communicate one coordinate in a multi-dimensional space, i.e., a vector, at any point in time. The visual counterpart of multidimensional sonification is illustrated in Fig. 2. One location is three-dimensional space (black dot) as well as its x-, y- and z-component (gray dots) are communicated at any point in time. Achieving this by means of sound is not straightforward. In the remainder of this section we explain how psychoacoustic sonification can serve for this purpose. As the example of a data point in three dimensional space is rather abstract, we give a more comprehensive explanation of multiple practical use cases in Section 5.

4.1 Problems With Conventional Sonification

For tasks that require complex information, i.e., multivariate or multidimensional data, conventional auditory displays [1, 5, 17, 19] could be demonstrated to be helpful in supplementing visual UIs. However, when trying to sonify multiple data streams, no conventional auditory display can satisfy all demands, which are

1. interpretability
2. linearity
3. continuity
4. high resolution

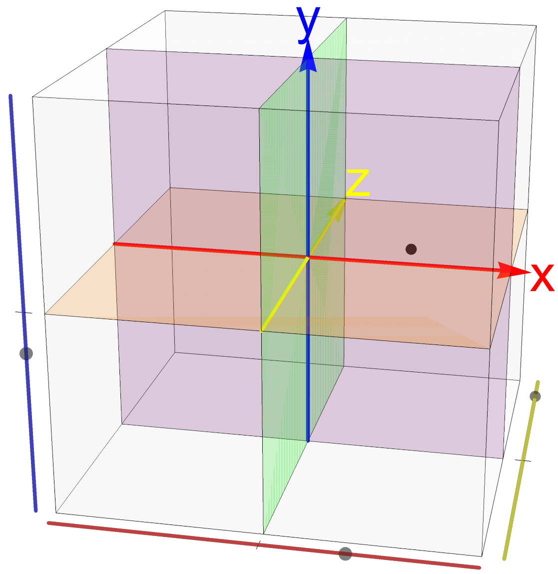


Figure 2: Abstract example of the psychoacoustic sonification. One sound informs about the location of a data point in three-dimensional space (black dot). The single sound attributes inform about its individual components, i.e., the direction and distance along the three orthogonal dimensions (gray dots on the colored lines). The data point can be static, dynamic, or interactively changed.

5. absolute magnitudes (interval or ratio scale of measure)

of each single data stream and

1. orthogonality
2. integrability

of multiple data streams.

Figure 3 illustrates common issues in multidimensional sonification. First of all, the x-axis is partly chaotic, which makes the sonification **uninterpretable** in that region. At the same time, the y-axis is **discontinuous**. This creates audible artifacts instead of a smooth transition when raising y continuously. Furthermore, the y-axis has **no negative** part. This means, it is at best half a dimension, since it only represents a distance in one direction, not in two. The low number of ticks on the y axis represents a **low resolution**, which is a typical issue in sonification. The two axes are no lines but curves. This curvature is a metaphor for the circumstance that the axes are nonlinear. This nonlinearity implies **unequal intervals**, i.e., doubling x does not necessarily sound twice as intense. The two axes do not cross, i.e., the coordinate system has **no obvious origin**. The two axes are **not right-angled**, i.e., the axes are not orthogonal. This means the axes can at most be considered as fractal dimensions. Due to the curvature of the axes, and their lack of orthogonality, many points $\vec{X}$ in the coordinate system are not unique, i.e., they are **ambiguous**. This means, when moving the vector $\vec{X}$, it is impossible to recognize whether the x- value, the y-value, or both has/have changed.

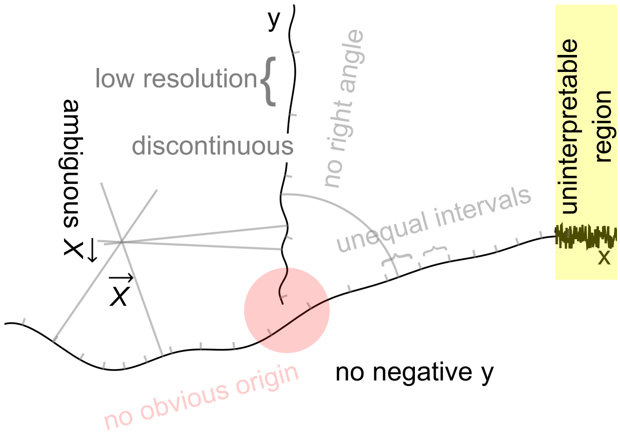


Figure 3: Typical perception issues in conventional, multidimensional sonification.

4.2 Advantages of Psychoacoustic Sonification

Psychoacoustic sonification [27, 29, 30] is the first sonification technology that satisfies all of the above-listed demands. Psychoacoustics describes the relationships between the physical, acoustic quantities and the perceptual, auditory qualities of sound by means of nonlinear equations [28, 39]. Physical quantities include amplitude, frequency and phase over time and space, auditory qualities include pitch, beats, roughness, sharpness, fullness and tonalness. These qualities have been shown to be able to accurately capture human listening listening experience.

Practically all physical sound parameters can affect each auditory quality. Consequently, physically quantities are not perceptually orthogonal, i.e., independent auditory qualities. In contrast to conventional sonification, psychoacoustic sonification does not map input data to physical sound quantities, but to perceptual auditory qualities. Psychoacoustic considerations in the signal processing improve the interpretability, linearity and resolution of sonification and make orthogonal and integrable multivariate or multidimensional sonification possible.

Figure 4 illustrates psychoacoustic sonification. Here, the sound impression is clear and non of the above-mentioned issues occur.

4.3 Exemplary Psychoacoustic Sonification

4.3.1 Properties

An example of a psychoacoustic sonification is illustrated in Fig. 5 for three-dimensional data. The sonification is a monophonic, tonal sound. Its *attributes* or *characteristics* represent different directions. The *magnitude* or *intensity* of each attribute represents the distance along that direction. The target lies at the origin of the coordinate system at $\vec{X}_{\text{target}} = (0, 0, 0)$, a vector $\vec{X}_{\text{user}}$ describes the location of the user.

The direction of the target relative to the user along the x-axis is mapped to the direction of the chroma aspect of pitch. At $x = 0$, pitch is steady, $x > 0$ creates a falling pitch, $x < 0$, creates a rising pitch. The distance along the direction is mapped to the speed with which pitch rises or falls. The larger

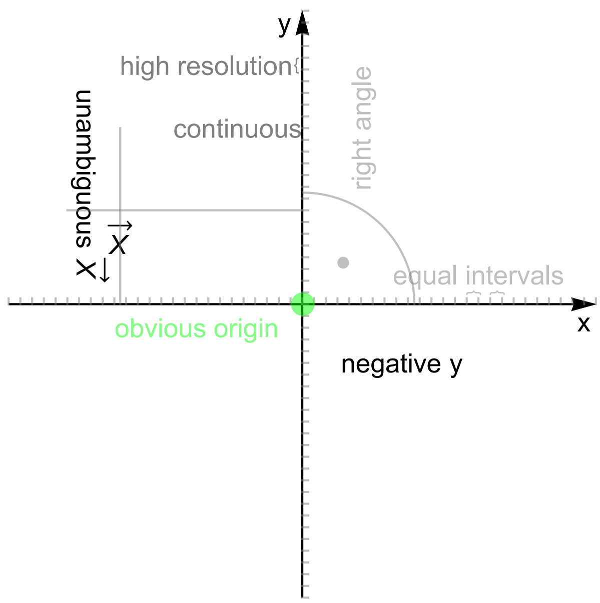


Figure 4: Perception of psychoacoustic sonification.

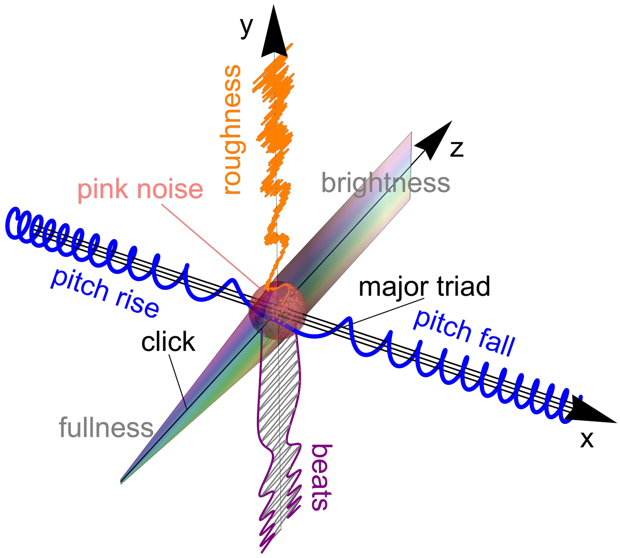


Figure 5: This psychoacoustic sonification uses cyclic pitch change as x-dimension, speed of beats and degree of roughness as y-dimension and degree of sharpness and fullness as z-dimension. A click indicates surpassing the x-y plane, a major triad surpassing the x-z plane, pink noise indicates that the target region has been reached.

the absolute value, the faster the pitch rise or fall.

The y dimension is divided in two. Positive y-values are mapped proportionally to the intensity of roughness. Negative y-values are mapped to the speed of beats, i.e., regular loudness fluctuations. Only at $y = 0$ the sound exhibits neither roughness nor beats.

Some simple trajectories through this two-dimensional sonification space can be found on our YouTube channel at https://www.youtube.com/watch?v=WaCzHU3-OYE &list=PLVv3BMS8IIXGOhYG0l1NCbAnRNCmoAe21.

The z-dimension is also divided in two. Positive z-values are mapped to the degree of brightness, negative values to the degree of fullness. At $y = 0$ the sound is dull but full.

Some exemplary trajectories through a three-dimensional psychoacoustic sonification can be found at https://www.youtube.com/watch?v=7EeB7AGJnpQ &list=PLVv3BMS8IIXGo-SkwwD9rSUQKCPLy89kK. The example videos slightly deviate from the mapping principle explained above.

In a navigation task the origin of the coordinate system represents either the target to reach, or the obstacle to avoid. The sonification itself informs the operator whether the origin has been reached or not. Only at the origin the sonification has a steady pitch and loudness, sounds smooth, dull and full. No reference tone is needed. At all other locations the sound attributes inform the operator where the target or obstacle lies. One important characteristic of this sonification is that it is intuitive on a qualitative level. Parameters like brightness and roughness, have an innate valence to a human listener. Near the origin, the sound is comparably pleasant and calm. However, towards the extreme values, when the operator is far away from the target, the sound becomes more and more obtrusive, unpleasant, and urgent, forcing the operator into action. For example, quick chroma fluctuations sound like a siren, a high degree of roughness sounds scratchy and distorted.

The urgency of extreme values in the chroma- and beats/roughness-dimension can be heard in a YouTube demo video: https://youtu.be/Z_Iu575erI4.

In addition to the interactive sonification, earcons are triggered. A click informs about surpassing the target height, a short major triad informs about surpassing the target depth. Reaching an extended zone around the ideal target point triggers subtle pink noise.

Together, the psychoacoustic sonification and the additional sounds of the auditory display interactively inform an operator about the location of a sensed or defined target in three-dimensional space. For example, the location of the machine can be the origin of the coordinate system. Then, the sound attributes inform the operator about the direction along the three spatial dimensions (x, y and z) and about the distance along each direction. So the operator gets informed whether the target has been reached, and if not, where the target is located.

4.3.2 Evaluation

We have experimentally validated the first two dimensions in multiple passive and interactive studies [27, 29–31, 33–35, 38, 40]. Passive listening tests with the x-y dimensions revealed that novice users can interpret the direction and magnitude of the two dimensions unambiguously after just 5 minutes of training [31]. Experiments with interactive sonification where the sound is continuously modified by the action of the operator revealed that interaction improves interpretability and precision a lot.

The signal processing for an interactive psychoacoustic auditory display including sonification and earcons is described in [38]. Interactive experiments revealed that novice users ac-

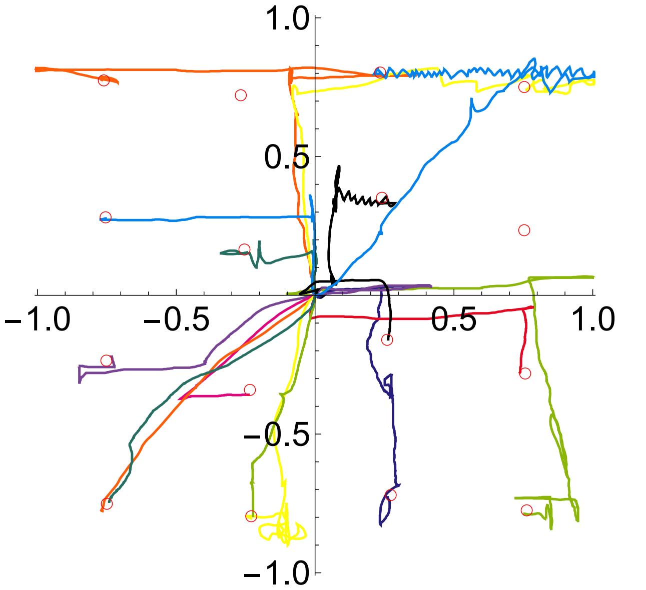


Figure 6: Typical mouse trajectories from novice users of psychoacoustic sonification. Some users approach the target axis-by-axis, whereas others aim for the direct path. Insecure users trigger the click repeatedly to confirm that they are still at the target height.

curately find invisible sonified targets that have a diameter of just 4 mm on a two-dimensional computer screen with an area of 4000 mm^2 [33, 40]. They only needed 30 minutes of explanation and exploration of the sonification to achieve this. Typical mouse trajectories are illustrated in Fig. 6. Practically all targets have been found by each participant. The six trajectories on in the lower right quadrant approach the target almost on the direct path. Direct paths highlight that users were able to integrate the two auditory dimensions and derive the respective angle from the two Cartesian axes. Other participants approached the target axis-by-axis, like the four examples in the lower right (x first) and the three examples in the upper left quadrant (y first). This strategy underlines that the two dimensions are orthogonal, i.e., unambiguously interpretable and distinguishable, even though the two dimensions are presented at the same time. Some participants sometimes surpassed the target height over and over, like the three examples in the upper right quadrant. This action triggers the click repeatedly and confirms that the users are still at the target height.

In [36] we presented the signal processing for the third dimension. Initial results can be found in [32], a preprint of the complete results is available at [37].

In a game-like environment we are validating the 3D psychoacoustic sonification. Screen shots of this serious game are shown in Fig. 7. People can play the game for fun, while we collect their motion trajectories to receive statistics from a large population and longitudinal data about learning effects. Game elements, like high scores, levels, missions and power-ups motivate the players to keep learning the sound metaphors, improve their skills and play for a long time. At higher levels less visual cues are given, while more information is presented auditorily to the player. At some

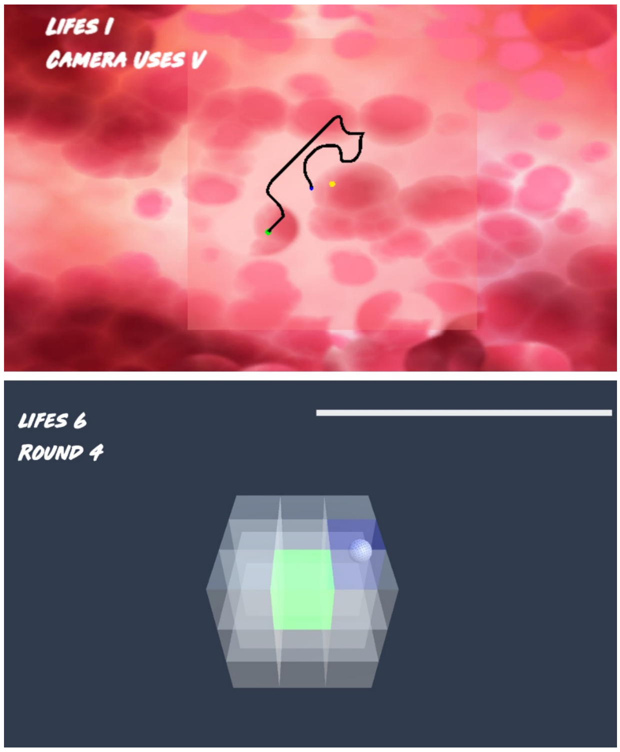


Figure 7: CURAT visual game environment to evaluate the 3D psychoacoustic sonification.

point the player depends completely on the sound to navigate through three-dimensional space. Details on the game design and progress are available on http://curat.informatik.uni-bremen.de/.

Figure 8 shows how we test the 3D psychoacoustic sonification in a surgical scenario. The planned location of burr holes for a craniotomy are found by means of sound. In a craniotomy, a part of the skull is removed, e.g., to insert an ablation needle that destroys a brain tumor. Here guidance is necessary, because a skull hardly provides visible landmarks that help for orientation. At the same time it is crucial to remove exactly the planned part of the skull to ensure that the needle can be inserted at the planned location, so that the tumor can be reached via the planned insertion angles. However, guidance by means of visual UIs has a number of drawbacks, which are discussed in detail in Section 5.4.

So far, we have implemented three dimensions and are in the process of evaluating them. However, psychoacoustic sonification is not limited to three dimensions. The current 3D sonification is completely tonal and largely harmonic. As we have already demonstrated, noisy and impulsive elements, i.e., pink noise and clicks, can be added without degrading the 3D sonification. We believe that at least three additional dimensions are implementable, with proper psychoacoustic sound design of noise and/or pulses.

In the next section, we describe application areas and potential use cases where psychoacoustic sonification can act as interface for human-machine interaction.

Figure 8: Tim Ziemer testing the 3D psychoacoustic sonification in a surgical setting.

5 USE CASES

Psychoacoustic sonification can serve as a UI to communicate the position of a point in a coordinate system. The point in the coordinate system may be a location in space or a vector in any other multi-dimensional or multivariate space. The position is either sensed by means of sensors, or it is predefined. The point can either be a natural limit, a target point to reach, or a critical point to avoid. The point and the origin of the coordinate system can either be fixed, or dynamic. As mentioned above, the psychoacoustic sonification can serve as a complementary or substitutional UI for HMI in situation in which visual UIs are impractical or insufficient.

As these conditions seem rather abstract, we present some exemplary use cases from different domains that highlight, how machine operators can benefit from psychoacoustic sonification.

5.1 Drone Flying

In manual drone flying, the HID is a remote control that has two sticks to control pitch, throttle, yaw and roll, and buttons that trim pitch to move back and forth, throttle to increase or decrease, and yaw or roll to move sideways. Practically all the above-mentioned sensors are available for drones, including GNSS and INS to track self-location and orientation and radar, lidar and sonar to track obstacles in the environment.

Visual UIs for manual drone navigation are often included in the remote control. They show the camera view overlain by depictions of compass, GPS location on a map and other graphics that indicate the location of obstacles, as illustrated in Fig. 9.

Manual Drone flying without sensors is restricted to situations in which the drone is in the visible field of the operator, or in open territories. Uncertainties occur, for example, in fog, as illustrated in Fig. 10.

Navigation by the drone's camera is only safe for basic flying, if the drone flies in the direction of the camera view, since obstacles outside the camera view are overseen. Furthermore, transferring the camera view to the remote control introduces a latency. This enforces the operator to slow down and wait

Figure 9: Typical drone remote. A visual UI shows the camera view overlain by graphics that indicate spatial relations, like drone coordinate and orientation, and distance and direction of obstacles.

Figure 10: In fog, visual navigation is almost impossible. Here, sensors can scan the environment and UIs can inform the operator about the location of obstacles.

for the picture when approaching an obstacle. For advanced maneuvers, some spatial relations need to be sensed. Autonomous drones are largely capable of flying without user control. However, in the case of uncertainty, e.g., due to environmental conditions or tasks that seem unsolvable to the machine's artificial intelligence, human control is necessary, and the operator can fly, while the drone supports her with the available sensor information. In such cases, e.g., in regions with a high density of buildings, trees, transmission lines and other structures, as in urban areas, forests, factory premises, and inside hangars and other buildings, UIs that inform the operator about sensed obstacles help avoiding a crash.

In visual UIs, overlays can occlude important aspects of the camera view. When visualizing the camera view and the location of obstacles on two different screens, the operator has to decide which screen to consult. Here, psychoacoustic sonification may be preferable over graphical UIs in manual

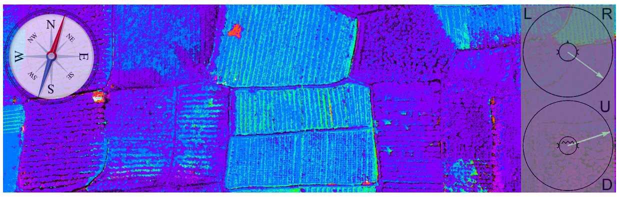


Figure 11: Near infrared spectrometry view from a drone in agricultural planing and monitoring. Overlays indicate the location of obstacles, but they partly occlude the view.

drone navigation.

In agricultural planing and monitoring with drones, near infrared spectrometry is used to identify the healthiest spots (red) and unhealthy crops (yellow). Here, a visual UI is realized with overlays that allow the operator to avoid hitting obstacles, like overhead transmission lines or trees. However, the overlays occlude some of the spectrometry visualization, as illustrated in Fig. 11. Here, psychoacoustic sonification allows the operator to concentrate on the visualization while being able to avoid sensed obstacles.

Drones are often used for photography, as illustrated in Fig. 12. Here the drone operator visually concentrates on finding the perfect spot for a camera shot. Autonomous drones might prevent the photographer from finding the perfect spot, if it lies dangerously close to an obstacle. However, for manual navigation the camera only helps to avoid obstacles in front. But it is likely that the operator will fly the drone backwards, up, down, left and right, while focusing the camera on the motif. Here, visual UIs with overlays of sensed obstacle positions are disruptive. Two-screen solutions without overlay are dangerous, because the visual focus lies on the camera view and not the navigation aid. In this situation, psychoacoustic sonification allows the operator to fly safely while finding the right spot.

Visual inspection is a task similar to photography. Visual inspections are carried out, e.g., for airplanes in hangars or roofs in urban areas or factory premises. Figure 13 shows an exemplary drone view of a bridge inspection. Here, obstacles are omnipresent, like abutments, piers, girders, trusses, cables. But manual navigation may become inevitable when the autonomous drone mode cannot find a proper perspective. The visual focus needs to lie on the camera view. Other graphical UI elements are distracting. Again, psychoacoustic sonification allows a single operator to both navigate the drone and perform the visual inspection at the same time.

Even though we discussed practical use cases for psychoacoustic sonification in the field of drone navigation, the principle transfers to all kinds of unmanned aerial vehicles, rockets and missiles, and even to other remotely operated vehicles (ROVs), like unmanned surface vessels and unmanned underwater vehicles. Figure 14 shows an underwater robot.

Underwater robots [2] are mostly used for survey work, like installation, inspection, welding and maintenance of subsea structures and production facilities. Especially offshore oil and gas installations are almost exclusively serviced by ROVs [2]. Here, proximity sensors detect obstacles based on active sonar, while a so-called *short baseline acoustic positioning system* provides absolute location information, just like GNSS, which are not working underwater. Underwater, vision can be limited due to darkness and muddy waters, so redundant multisensor systems are employed to navigate safely to the target structure. Here, psychoacoustic sonification can give a low-latency feedback on the location of obstacles.

Figure 12: Finding the perfect camera view without distracting UI graphics, while avoiding obstacles, like trees or overhead power lines.

5.2 Manned Aircraft

Aircraft tend to have a very high number of visual displays. This is true for civil and military airplanes, jets and helicopters, as illustrated in Figs. 15 to 17.

These displays serve to fly the aircraft, referred to as *instrument flying*. Instruments include altitude meter, airspeed indicator, vertical speed indicator, turn and slip indicator, compass or another heading indicator, and attitude indicator, indicating pitch and roll. Instrument flying is particularly necessary in bad visual conditions, e.g., in darkness, clouds, fog, smoke or heavy rains. Unfortunately, instrument flying imposes a “tremendous burden” on the eyes [14] and is cognitively demanding, since the information of multiple displays needs to be interpreted and integrated. During some maneuvers the pilot’s ability to interpret visual displays or match visual and balance perception may be impaired [8, 14, 40]. Consequently, spatial disorientation causes most aviation accidents [8].

Auditory displays have been developed to enable blind flight

Figure 13: The location and orientation of drones constantly changes in inspection tasks. Obstacles can be everywhere, but the visual focus needs to lie on the inspected structure.

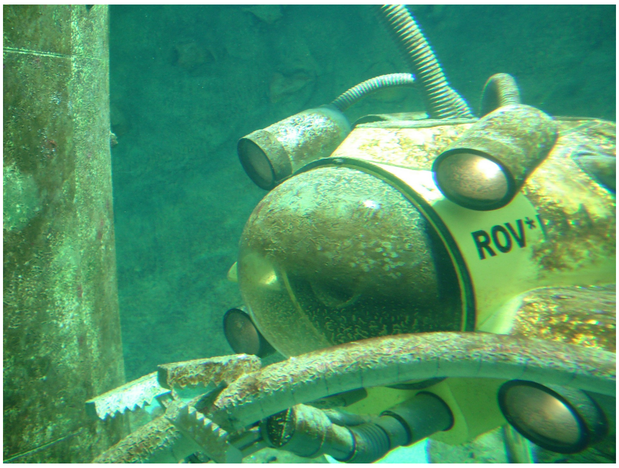


Figure 14: Underwater robot for maintenance work. Psychoacoustic sonification can help reaching the target structure and avoiding potentially moving obstacles despite limited vision.

[8, 9, 14], i.e., stable flight without visual cues. These studies conclude that it would be highly desirable to have a system that communicated pitch and roll with a high resolution on a linear ratio scale of measure, i.e., with an obvious coordinate origin that represents the horizon. Psychoacoustic sonification fills exactly this gap. Here, the coordinate origin represents the horizon, two dimensions represent pitch and roll of the attitude indicator, which is the primary instrument for flying. The third dimension may represent the air speed. Additional earcons can serve to represent altitude. This even enables pilots to stay on a pre-defined course or carry out maneuvers safely, without the need for visual feedback.

5.3 Autonomous Vehicles

Autonomous vehicles, like self-driving cars, are classified into six different levels of autonomy by the Society of Automotive Engineers (SAE) [22]. Only levels 4 and above allow the driver to be inattentive. At lower levels the environment may be scanned by means of sensors as illustrated in Fig. 18. The driver needs to stay aware of the vehicle's actions to take over control if necessary.

Figure 15: Cockpit of an airliner. Not all visual displays lie in the visual focus. Some may even be occluded by the pilot or the copilot.

Figure 16: Cockpit of a jet fighter. Some maneuvers cause a mismatch of vision and balance and may even impair the ability to read visual displays.

Here, hand-over and take-over commands and alarms are the minimum requirement. However, psychoacoustic sonification can inform the driver continuously about the actions of the vehicle. For example, the automotive can spontaneously decide to deviate from the planned route if it receives information on a traffic jam or recognized a "Road closed!" sign. Here, one sonification dimension could indicate the pending deviation, i.e., that the car plans to turn right instead of driving straight ahead. Another sonification dimension could indicate the absolute deviation from the planned route. The presence of the sound capzures attention and increases driver awareness. Furthermore, the driver is informed about the decision of the vehicle and its consequence for the route. These pieces of information help the driver decide whether it is acceptable and necessary to turn right, and if the new route seems appropriate or causes too much deviation. If preferred, the driver can take over control. Here, one characteristic of the above-described psychoacoustic sonification becomes effective: the larger the deviation from the planned path and route, the more obtrusive and unpleasant the sound becomes for the listener. This makes the sonification intuitive on a qualitative level.

In a similar fashion, the psychoacoustic sonification could

Figure 17: Helicopter cockpit. The high number of visual displays and the inevitable use of headphones pave the way for psychoacoustic sonification for navigation.

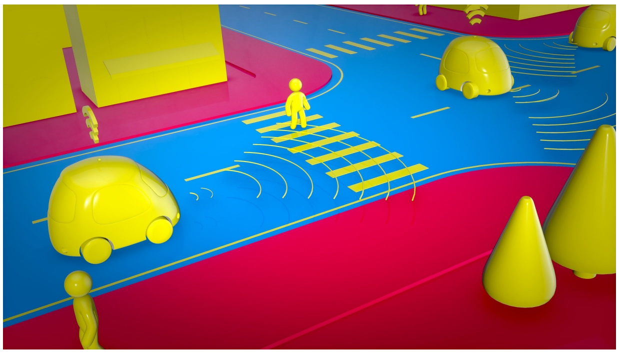


Figure 18: Autonomous vehicles of level 0 to 3 require the driver to stay aware of the car action and readily intervene when called upon by the vehicle.

inform the driver about the confidence level of the implemented artificial intelligence. For example, if road signs or road markings are recognized with a low confidence, the sound could pass this information to the driver. Then, the driver is aware about the general confidence and confidence change but can still concentrate on other things until he or she decides to take action. This is the advantage of psychoacoustic sonification over speech messages, which require more attention and more cognitive resources to be interpreted.

5.4 Image-Guided Surgery

Image-Guided Surgery (IGS) enables surgeons and other clinicians to carry out interventions that are too risky without guidance. Examples are biopsy of tissues or resections of tumors near critical structures, like large arteries, important nerves, or sensitive membranes.

In image-guided surgery (IGS), the human or veterinary patient is scanned, e.g., by means of computed tomography or magnetic resonance imaging. From the scan, an augmented reality 3D model is created. It shows the anatomy, where associated structures are grouped, like vessels, nerves, bones, healthy tissue and tumor. Each group can receive an individual color and opacity level. This helps to localize the exact location of the tumor, the center and the extent. With this model clinicians plan the surgery, e.g., where to make incisions and how to approach the tumor past critical or impenetrable structures. These plans, may include insertion points, and the angles and depth of an ablation needle from the surface to the center of the tumor. Or they include the cutting trajectory for a scalpel blade.

During the procedure, the location and orientation of the surgical tool is tracked in relation to the patient anatomy. Tracking devices are electromagnetic sensors or optical markers that are sensed by stereoscopic infrared cameras. Tools include scalpels, cauterizers, biopsy or ablation needles. On a screen clinicians see an overlay of the surgical tool and the augmented anatomy. This mixed reality situation guides the clinician, who can see the current location of the tool, the target location, and the position of obstacles, i.e., critical and impenetrable structures. In terms of HMI, the tool is the HID, the tracking system is the sensor and the screens show the UI.

The major drawback of IGS is the high cognitive demand that mostly originates in the UI design. Typical IGS setups (see Fig. 19 for an example) give rise to a number of spatial challenges such that mentally processing spatial information becomes a crucial ability for successfully performing IGS [24]. Often, installing the monitors has a low priority, so they might be located a little far away or even behind the surgeon, which is rather unergonomic. The screen shows a pseudo 3D model of the patient together with the three anatomic planes of the patient, i.e., sagittal, coronal and axial. All four graphics are overlain by a depiction of the surgical tool. However, none of the graphics' perspectives coincides with the perspective of the surgeon. So he or she has to mentally integrate the four graphics, mentally rotate, translate and scale the them to navigate properly. This is quite demanding. For example, the surgeon might move the tool to her right. This might cause a motion of the overlain tool to the upper left in the pseudo 3D-model, a small motion to the upper right in the median view, a motion far to the right and slightly up in the coronal view and a motion to the right and slightly down in the axial view.

In addition to that, 3D graphical processing can cause latencies of hundreds of milliseconds. This forces clinicians to operate so slowly that they can always anticipate their proximity to target and obstacles, despite the delayed feedback.

Here, psychoacoustic sonification offers a huge advantage. Instead of looking at a screen to derive spatial information from integrating four seemingly conflicting views, the surgeon can visually focus on the patient while navigating by means of sound. The sound can either communicate the location of the tool in relation to the planned trajectory, location of the nearest obstacle, or the location of the target in relation to the current tool position.

5.5 Anesthesiology

In the medical sector, not only surgeons can benefit from psychoacoustic sonification. Anesthesiologist monitor vital parameters of patients in the operating room, the postanesthesia care unit and the intensive care unit. Vital parameters include heart rate, blood pressure, end-tidal carbon dioxide (capnography), respiration rate, tidal volume, oxygen concentration in the blood, and body temperature [25]. These tend

Figure 19: Typical IGS situation: The monitor is too far away to see details. It shows an overlay of the surgical tool in a pseudo-3D graphic, and three 2D graphics that show the three anatomic planes. None of the graphics represents the first-person perspective of the surgeon of the assistants. Still, they have to take their eyes off the patient to navigate during the surgery.

Figure 20: User interface for monitoring a patient's vital functions. A monitor is combined with an auditory display to serve as a multi-modal user interface.

to have an ideal value, a healthy range, and critical thresholds that depend on factors, like age, type of surgery, disease state, and anesthetics [4].

A typical user interface for vital parameter monitoring is illustrated in Fig. 20. The current state and a short history of multiple vital functions are plotted over time and displayed on a screen. Often, integrated loudspeakers play additional pulse-oximetry sounds.

Auditory pulse oximeters are among the most widely used sonifications. They inform the anesthesiologist about the heart rate and the relative oxygen concentration in the blood. The strength of this auditory adjunct to the visual UI is that it enables anesthesiologists to keep monitoring and staying aware of two essential parameters without looking at a screen, e.g., while counteracting adverse reactions, intubating, or repositioning the patient. There are attempts to add information of even more vital functions [13, 25]. The main drawback of conventional sonification for patient monitoring is that the sounds fail to communicate precise quantitative information about the relation between current state, target value and/or critical thresholds.

Figure 21: While drinking water, an obtrusive and unpleasant sound is an obvious warning. Instructed people can extract even more information about the water quality.

In terms of HMI, sensors measure the vital functions over time. Visual displays, i.e., plots on a screen, serve as UI, together with pulse-oximetry sounds. Psychoacoustic sonification can serve to communicate the state of one or more vital functions in relation to the target value and/or critical thresholds. This has been demonstrated for a pulse-oximetry scenario of neonates [23]. Here, users are able to identify whether the oxygen saturation lies safely within the target region (near 92.5%, near or within the critical region below 90% or above 95%).

5.6 Water Quality Monitoring

In regions with chlorinated water distribution systems, households and pubic drinking fountains can be equipped with electrode-based pH meters, digital chlorine sensors and thermometers to measure pH, residual chlorine and thermometers to monitor the quality of the tap water in real-time. Such sensors serve as a fallback solution in addition to the monitoring that is required for distribution systems of municipal water systems.

Raising instant awareness of these measures can help to prevent from water-borne diseases and protect children from tap water scald burns. However, as illustrated in Fig. 21, the typical drinking posture prevents you from looking at a screen. Here, psychoacoustic sonification is the ideal UI. As described above, it can be equipped with an innate quality, where large deviations from the coordinate origin sound unpleasant. Here, a clean sound represents optimal drinking water conditions at the coordinate origin. Deviations sound sound either rough or siren-like; both intuitively sound alarming. Brief instructions on the fountain can inform the user about the detail of the different sound qualities.

Water quality monitoring is just one out of a number of monitoring tasks that can be assisted by psychoacoustic sonification. We can also imagine to monitor temperature and flow rate of a cooling circuit, visits, downloads and responds on a web-server or pressure, temperature and wind velocity of a weather station.

Figure 22: On rescue missions inside burning buildings, firefighters use sensors to recognize obstacles in their environment or holes in the ground. Psychoacoustic sonification can warn them about the nearest danger to guide them safely.

5.7 Firefighters

On a rescue mission, firefighters may enter burning or smoke-filled building. An exemplary training scenario can be seen in Fig. 22. Here, firefighters can hardly see their environment. Approaches exist to equip firefighters with sonar [3] or radar [10] sensors. These localize obstacles or holes in the ground in smoke-filled buildings. Due to the limited vision, even hand-held monitors cannot be employed as a UI to pass the sensor information to the firefighter. Therefore, haptic devices [3] [10] have been suggested. These inform the firefighter about the presence of an obstacle around or a hole in the ground to increases their spatial awareness.

In this scenario, psychoacoustic sonification can not only inform them about the presence of an obstacle or hole, but even about its location, i.e., direction and distance. This additional information allows them to explore the building faster and to distribute more widely compared to haptic 'presence only'-feedback and compared to conventional exploration along a guide-rope.

Note that firefighters often use drones to get an overview, to monitor and coordinate operations in disaster areas, like flood, wildfire and large-scale structure fires, and to search and rescue victims. Likewise, they may fly firefighting planes or helicopters. One sound, i.e., one type of psychoacoustic sonification could serve as a UI for all these application areas of firefighters.

5.8 Freight Container Cranes

In container ports, freight container cranes unload containers from container ships. These containers are densely piled. Due to continual growth of container ships, the containers need to be piled faster and more densely every year. A photo of a container port is shown in Fig. 23.

Piling cranes based on vision is very demanding. Due to the large distance between the crane operator and the container, it may be difficult to estimate whether spaces are sufficiently large for a container to get through. Especially depth perception and estimation of spatial constellations in partly occluded scenes is demanding. Furthermore, the steel cabin as well as other containers and cranes may party occlude the visual field. To assist crane operators in piling the containers without accidental collisions, cranes became equipped with proximity sensors, such as laser rangefinders. If an obstacle is dangerously close, an alarm will sound. In this situation psychoacoustic sonification could offer much more than just an alarm. The sound could inform the operator not only about the fact that an obstacle is close. It even informs the operator where, i.e., in which direction and how far, the obstacle lies.

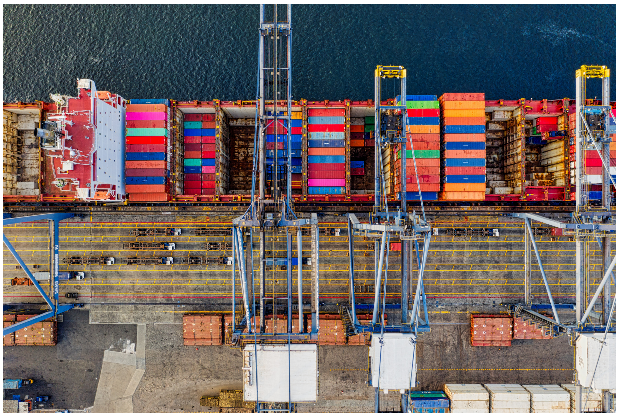


Figure 23: Freight container cranes in a container port. The large distance between operator and container as well as the high density of piles make visual navigation demanding and dangerous. Psychoacoustic sonification can inform crane operators about the distance and direction of the nearest container to avoid collisions.

6 CONCLUSION

In this paper we gave an extensive introduction to *psychoacoustic sonification*. It is allocated at the user interface part of the human-machine interaction circle. The psychoacoustic closed loop can serve as auditory user interface or as the acoustic part of a multimodal user interface. Psychoacoustic sonification takes the largely interfering and nonlinear relationships between acoustics and auditory perception into account in the digital signal processing for the sound design. This ensures interpretability, linearity, continuity, a high resolution and absolute magnitudes of each single dimension in an uni-or multidimensional sonification. Furthermore, psychoacoustic sonification ensures orthogonality and integratability of the axes. This means, psychoacoustic sonification can carry more information than conventional sonification, without losing any clarity.

We referred to our previous publications on the theory, implementation and experimental evaluation of psychoacoustic sonification and reported the state of the art and ongoing research. Moreover, we explained what benefits psychoacoustic sonification can offer for human-machine interaction.

Psychoacoustic sonification is beneficial whenever vision is the bottleneck of performance, be it due to limited vision, visual overload, demanding visual interpretation or the limited focal point of visual attention. This is the case in various scenarios in all kinds or application areas.

Researchers and practitioner of many disciplines are unaware of sonification and the benefit that auditory displays can offer to them. Hence, we presented a number of practical use cases from multiple application areas where the psychoacoustic sonification can take full effect. Of course, our listing is not exclusive, but exemplary.

7 OUTLOOK

We are still in the progress of evaluating the benefit of the psychoacoustic sonification as a user interface for human-machine interaction in both abstract scenarios and a practical use case, namely surgical navigation. Pretests by the authors, demo sessions on conference and the initial results of the experiments provide evidence that the 3D psychoacoustic sonification enables users to navigate through three-dimensional space after just a couple of minutes of training. Note that the precise sonification of three orthogonal and continuous dimensions is not a natural upper limit. It is possible to include three new dimensions not based on a tonal but a noisy sonification signal. Together, such a six-dimensional sonification could allow an operator to navigate to a target location in three-dimensional space while avoiding obstacles. Or the six dimensions could represent all six degrees of freedom of an object, including three-dimensional location and orientation. For blind flight, new dimensions could sonifiy the information of the turn and slip indicator, enabling pilots to carry out a number of complicated maneuvers blindly.

Note that the psychoacoustic sonification might be optimized for each individual application. Ideally, as little information as needed is provided aurally, to shorten the learning time.

References

[1] J. E. Anderson and P. Sanderson. ”sonifica-tion design for complex work domains: Dimen-sions and distractors”. *Journal of Experimental Psychology: Applied*, Vol.15, No.3, pages 183–198, Mar (2009). doi:10.1037/a0016329.

[2] G. Antonelli, T. I. Fossen, and D. R. Yoerger. ”underwater robotics”. In B. Siciliano and O. Khatib, editors, *Springer Handbook of Robotics*, pages 987–1008. Springer Berlin Heidelberg, Berlin, Heidelberg, (2008). doi:10.1007/978-3-540-30301-5_44.

[3] C. Bertram, Mathew H. E., M. Javaid, T. Stafford, and T. J. Prescott. ”sensory augmentation with distal touch: The tactile helmet project”. In N. F. Lepora, A. Mura, H. G. Krapp, P. F. M. J. Verschure, and T. J. Prescott, editors, *Biomimetic and Biohybrid Systems. Second International Conference, Living Machines 2013 London, UK, July/August 2013 Proceedings*, pages 24–35, Berlin, Heidelberg, (2013). Springer. doi:10.1007/978-3-642-39802-5.

[4] B. Bindu, A. Bindra, and G. Rath. ”temperature management under general anesthesia: Compulsion or option”. *Journal of anaesthesiology, clinical pharmacology*, Vol.33, No.3, pages 306–316, (2017). doi:10.4103/joacp.JOACP_334_16.

[5] D. Black, C. Hansen, A. Nabavi, R. Kikinis, and H. Hahn. ”a survey of auditory display in image-guided interventions”. *International Journal of Computer Assisted Radiology and Surgery*, Vol.12, No.9, pages 1665–1676, (2017). doi:10.1007/s11548-017-1547-z.

[6] K. Blankenbach. ”what is a display? an introduction to visual displays and display systems”. In J. Chen, W. Cranton, and M. Fihn, editors, *Handbook of Visual Display Technology*, pages 1–22. Springer, Cham, (2016). doi:10.1007/978-3-319-14346-0_201.

[7] S. Brewster. ”nonspeech auditory output”. In J. A. Jacko and A. Sears, editors, *The Human-Computer Interaction Handbook*, pages 220–239. Lawrence Erlbaum Associates, London, (2003). doi:10.1201/9781410606723.

[8] D. S. Brungart and B. D. Simpson. ”design, validation, and in-flight evaluation of an auditory attitude indicator based on pilot-selected music”. In *14th International Conference on Auditory Display*, page 8 pages, Paris, Jun (2008).

[9] D. S. Brungart, B. D. Simpson, R. C. Dallman, G. Romigh, R. Yasky, and J. Raquet. ”a comparison of head-tracked and vehicle-tracked virtual audio cues in an aircraft navigation taask”. In *Proceedings of the 13th International Conference on Auditory Display*, pages 32–37, Montréal, Jun (2007).

[10] A. Cassinelli and C. Reynolds. ”augmenting spatial awareness with haptic radar”. In *Proceedings of the 10th IEEE International Symposium on Wearable Computers*, pages 61–64, (2006). doi:10.1109/ISWC.2006.286344.

[11] G. Cook. *”Mobile Robots. Navigation, Control and Remote Sensing”*. IEEE/Wiley & Sons, Hoboken, NJ, (2011). doi:10.1002/9781118026403.

[12] J. Edworthy. ”medical audible alarms: a review”. *J Am Med Inform Assoc*, Vol.20, pp. 584–589, (2013). doi:10.1136/amiajnl-2012-001061.

[13] W. T. Fitch and G. Kramer. ”sonifying the body electric: Superiority of an auditory over a visual display in a com-plex multivariate system”. In Gregory Kramer, editor, *Auditory Display: Sonification, Audification and Audi-tory Interfaces*, volume 18 of *SFI Studies in the Sciences of Complexity*, pages 307–325, Reading (MA), (1993). Perseus Publishing.

[14] L. de Florez. ”true blind flight”. *Journal of the Aeronautical Sciences*, Vol.3, No.5, pp.168–170, (1936). doi:10.2514/8.176.

[15] P. D. Groves. *"Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems"*. Artech House, Boston, MA, 2nd edition, (2013).

[16] T. Hermann, A. Hunter, and J. G. Neuhoff, editors. *"The Sonification Handbook"*. COST and Logos, Berlin, (2011).

[17] T. Komatsu and S. Yamada. "can monaural auditory displays convey directional information to users?". In A. Papafragou, D. , Grodner, D. , Mirman, and J.C. Trueswell, editors, *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, pages 930–935, Austin, TX, (2016). Cognitive Science Society.

[18] Jörg Krüger. "human-machine collaboration". In S. Chatti, L. Laperriére, G. Reinhart, and T. Tolio, editors, *CIRP Encyclopedia of Production Engineering*, pages 922–925. Springer, Berlin, Heidelberg, (2019). doi:10.1007/978-3-662-53120-4_300320.

[19] T. Lokki and M. Gröhn. "navigation with auditory cues in a virtual environment". *IEEE MultiMedia*, Vol.12, No.2, pages 80–86, April (2005). doi:10.1109/MMUL.2005.33.

[20] R. W. Morris and S. R. Montano. "response times to visual and auditory alarms during anaesthesia". *Anaesth. Intensive Care*, Vol.24, No.6, pp. 682–684, (1996).

[21] S. C. Peres, V. Best, D. Brock, C. Frauenberger, T. Hermann, J. G. Neuhoff, L. V. Nickerson, B. Shinn-Cunningham, and T. Stockman. "auditory interfaces". In Kortum P., editor, *HCI Beyond the GUI. Design for haptic, speech, olfactory and other nontraditional interfaces*, chapter 5, pages 147–195. Morgan Kaufman, Burlington, MA, (2008).

[22] SAE International. "taxonomy and definitions for terms related to on-road motor vehicle automated driving systems", January (2014).

[23] S. Schwarz and T. Ziemer. "a psychoacoustic sound design for pulse oximetry". In *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)*, Newcastle, June (2019). doi:10.21785/icad2019.024.

[24] T. Vajsbaher, H. Schultheis, and N. K. Francis. "spatial cognition in minimally invasive surgery: A systematic review". *BMC Surgery*, Vol.18, No.94, (2018). doi:10.1186/s12893-018-0416-1.

[25] M. Watson and P. Sanderson. "sonification supports eyes-free respiratory monitoring and task time-sharing". *Human Factors*, Vol.46, No.3, pages 497–517, (2004). doi:10.1518/hfes.46.3.497.50401.

[26] D. Worrall. *"Sonification Design"*. Springer, Cham, (2019). doi:10.1007/978-3-030-01497-1.

[27] T. Ziemer. "two-dimensional psychoacoustic sonification". In Franziska Olbertz, editor, *33. Jahrestagung der deutschen Gesellschaft für Musikpsychologie (DGM)*, pages 60–61, Hamburg, Sep (2017).

[28] T. Ziemer. *"Psychoacoustic Music Sound Field Synthesis"*, chapter Psychoacoustics, pages 73–122. Springer, Cham, (2020). doi:10.1007/978-3-030-23033-3_4.

[29] T. Ziemer and D. Black. "psychoacoustic sonification for tracked medical instrument guidance". *The Jour-nal of the Acoustical Society of America*, Vol.141, No.5, pages 3694, (2017). doi:10.1121/1.4988051.

[30] T. Ziemer and D. Black. "psychoacoustically motivated sonification for surgeons". In *International Journal of Computer Assisted Radiology and Surgery*, Vol. 12, (Suppl 1):1, pages 265–266, Barcelona, Jun (2017).

[31] T. Ziemer, D. Black, and H. Schultheis. "psychoacoustic sonification design for navigation in surgical interventions". *Proceedings of Meetings on Acoustics*, 30, (2017). doi:10.1121/2.0000557.

[32] T. Ziemer, T. Höring, L. Meirose, and H. Schultheis. "monophonic sonification for spatial navigation". In *International Workshop on Informatics (IWIN 2019)*, pages 95–99, Hamburg, September (2019).

[33] T. Ziemer and H. Schultheis. "a psychoacoustic auditory display for navigation". In *24th International Conference on Auditory Displays (ICAD2018)*, Houghton, MI, June (2018). doi:10.21785/icad2018.007.

[34] T. Ziemer and H. Schultheis. "perceptual auditory display for two-dimensional short-range navigation". In *Fortschritte der Akustik — DAGA 2018*, pages 1094–1096, Munich, March (2018).

[35] T. Ziemer and H. Schultheis. "perceptual auditory display for two-dimensional short-range navigation". In *Programm der 44. Jahrestagung für Akustik*, page 306, Munich, Mar (2018).

[36] T. Ziemer and H. Schultheis. "psychoacoustical signal processing for three-dimensional sonification". In *25th International Conference on Auditory Dis-plays (ICAD2019)*, Newcastle upon Tyne, jun (2019). doi:10.21785/icad2019.018.

[37] T. Ziemer and H. Schultheis. "three orthog-onal dimensions for psychoacoustic sonifica-tion". *Preprint*, (2020). arXiv:1912.00766, doi:10.13140/RG.2.2.36037.19683.

[38] T. Ziemer, H. Schultheis, D. Black, and R. Kiki-nis. "psychoacoustical interactive sonification for short range navigation". *Acta Acustica united with Acustica*, Vol.104, No.6, pages 1075–1093, (2018). doi:10.3813/AAA.919273.

[39] T. Ziemer, Y. Yu, and S. Tang. "using psychoacoustic models for sound analysis in music". In *Proceedings of the 8th Annual Meeting of the Forum on Information Retrieval Evaluation*, FIRE '16, pages 1–7. ACM, (2016). doi:10.1145/3015157.3015158.

[40] Tim Ziemer and Holger Schultheis. Psychoacoustic auditory display for navigation: an auditory assistance system for spatial orientation tasks. *J. Multimodal User Interfaces*, 13, Special Issue: Interactive Sonification, (2018).
doi:10.1007/s12193-018-0282-2.

(Received April 14, 2020)

Tim Ziemer is a researcher in the field of applied psychoacoustics. He received his doctorate in Systematic Musicology at the University of Hamburg in 2015. He worked as a researcher at the Institute of Systematic Musicology at the University of Hamburg, at the National Institute of Informatics, Tokyo, and in the Medical Image Computing Group and the Bremen Spatial Cognition Center at the University of Bremen. His research interests include sound field synthesis, sonification, microphone array technology and music information retrieval. He received several scholarships and awards, e.g., by the Acoustical Society of America (ASA), Claussen Simon Foundation, Association for Computing and Machinery (ACM) and Special Interest Group on Information Retrieval (SIGIR), and by Hyundai Motors and the International Community for Auditory Display (ICAD).

Nuttawut Nuchprayoon is a PhD candidate and equine practitioner at the Faculty of Veterinary Science, Mahidol University, in Nakhon Pathom, Thailand. He has started working as equine clinician in 2014. His research interests are equine biomechanics and equine dentistry. He got an intern abroad scholarship for short-term studying in Germany during June to August 2019 at Bremen Spatial Cognition Center, University of Bremen.

Holger Schultheis has a background in Computer Science and Cognitive Psychology and is currently a senior researcher at the University of Bremen. His research aims to bridge the gap between natural and artificial intelligence. By combining insights from cognitive science and methods from artificial intelligence he works towards more sophisticated artificial agents and a deeper understanding of human cognition. Among other things, he applies the insights gained from his basic research to build intelligent tutoring and cognitive assistance systems that flexibly adapt to the current needs of the human user.

Industrial Paper

Proposal on Victims Information Management System

Kosuke Akasaka [*], Takayo Namba [**], Masao Isshiki[*], and Keiichi Abe [*]
[*] Kanagawa Institute of Technology, Atsugi City, Kanagawa, Japan
[**] Tokyo Metropolitan University, Hachioji City, Tokyo, Japan
abe@he.kanagawa-it.ac.jp

Abstract - In this paper, we propose a victims information management system that collects and transmits information on disaster victims and relief needs in evacuation refuges even when the lifelines such as communication and electricity are disrupted, assuming a large-scale disaster occurrence at the Tohoku Earthquake level. In addition, we propose the management of the health status of the victims living in the refuges, and also the function that can centrally manage the refuges around the country. Since the system scale of the disaster victim information management system proposed by us is large, this paper describes the realization of a system to manage the information and health condition of victims in refuges. Also, we will describe in detail the prototype development and evaluation of the proposed system.

Keywords: Disaster, Refuge Management System, Victims, Persons requiring special care, ICT, Solar Panel

1 INTRODUCTION

At the time of past large-scale disasters such as the Hanshin-Awaji Earthquake in 1995, and the Tohoku Earthquake in 2011, it took a considerable amount of time to collect information on the victims [1]. In particular, refuge administrators couldn't capture relief needs such as the presence of refugees who need special care and their allergy information. Also, due to the disruption of the infrastructure at the time of disaster occurrence, managers were unable to make the most of the Information and Communication Technology (ICT) and its power, and they could not collect and transmit information on the victims promptly [1] [2].

In addition, past measures put more emphasis on disaster prevention (such as collapse of buildings) rather than preparation for disasters (such as securing disaster supplies). Therefore, the quality of life (QoL) of the victims has not been sufficiently considered in the long-term refuge life after the occurrence of a disaster. In fact, in the Tohoku Earthquake, there were many victims who could not cope with the refuge environment and suffered from health problems [3] [4].

Therefore, we developed the Refuge Management System (RMS) that consistently manages the health condition of victims by collecting information on disaster victims and creating an evacuation list using ICT even at the time of the disruption of lifelines such as communication and electricity [5] [6]. The final goal of this study is to develop a system that installs RMS in refuges all over the country, and uses cloud servers to centrally manage the disaster situation and relief needs in each refuge area. We call this system the “Victims Information Management System (VIMS)” and are developing it. Since the development scale of the VIMS is large, in this paper, we describe the contents of the research and development project and the realization of the system to manage personal information and health condition of the victims in the refuge currently under development.

2 RELATED RESEARCH

After the Tohoku Earthquake, research on victims’ support systems using ICT is underway by many researchers, companies and universities in various places [7].

Similar representative existing technologies include "Earthquake disaster recovery support system" provided by Microsoft [8] and “information center victim support system" provided for free by Nishinomiya city [9]. These systems collect information such as the number of victims, the proportion of men and women, safety information, etc. On the other hand, these systems are not designed to pick up and manage the needs information of persons requiring special care (patients with allergic diseases or intractable diseases, disabled people, elderly people who need nursing care, pregnant women, etc.) within the refuge. In addition, these systems have the problem that they cannot detect and manage victims who have become ill due to a long-term refuge life.

Therefore, the problems of the various disaster victim supporting systems are as follows.

① The existing disaster victim support systems cannot operate when there is a power and communication infrastructure disruption immediately after the disaster occurs.
② The existing systems cannot adequately collect the needs of persons requiring special care.
③ It is difficult to manage the presence of victims among multiple refuges.
④ The QoL of the victims declines in refuge life.
⑤ It is impossible to centrally monitor damage situation of each refuge and relief needs in real time immediately after the disaster.

The VIMS proposed in this paper aims to solve problems ① to ④.

3 VICTIMS INFORMATION MANAGEMENT SYSTEM (VIMS)

In refuge support, the measures to be taken in refuges change with the passage of time from the occurrence of the disaster. For example, even if the power and communication infrastructure are cut off immediately, it is necessary to quickly grasp the situation of the victims in the refuges and inform external organizations of the situation after the occurrence of a disaster. On the other hand, there are many victims who are forced to live in refuges even after the restoration of infrastructure is completed and a long time has passed since the disaster occurred. For such victims, refuge managers need to manage their health status and prevent infections, economy class syndrome and so on. Furthermore, it is also necessary to grasp the damage situation regarding the collapse of the houses owned by the victims with the passage of time.

Therefore, in this research, as shown in Fig. 1, research and development are being carried out in the following four phases aiming for the realization of the final goal, VIMS. We assume that Phase 1 of VIMS will collect victim information, send it to Disaster Countermeasures Headquarters, and manage it in shelters immediately after a disaster occurs. We suppose in Phase 2, there are determined the transported destination and refuge spaces of the victims appropriately from the disaster situation and medical condition of them. We assume that in Phase 3 we will manage the health status of victims in long-term refuge life in real time. In Phase 4, we will devise a system that is able to monitor centrally the disaster situation and relief needs in a large area refuge. The main feature of VIMS is the support and management of victims in single system, from the immediately after a disaster occurs to the long-term shelter life. This feature is not found in the existing technology, so it can be said to be a great strength of this research.

Specifically, we will develop the Refuge Management System (RMS) corresponding to Phase 1 to Phase 2 shown in Fig. 2 and aim to install the proposed RMS at evacuation sites all over the country. The final goal is to realize a system that visualizes the disaster situation of refuges nationwide and the relief needs information of the victims by listing the victim information collected by the RMS as a cloud server by wireless communication etc. and mapping it with map information. In this paper, we describe the research results of Phase 1 and Phase 2 of this research project and the overall concept of the information management system of the victims in the refuge.

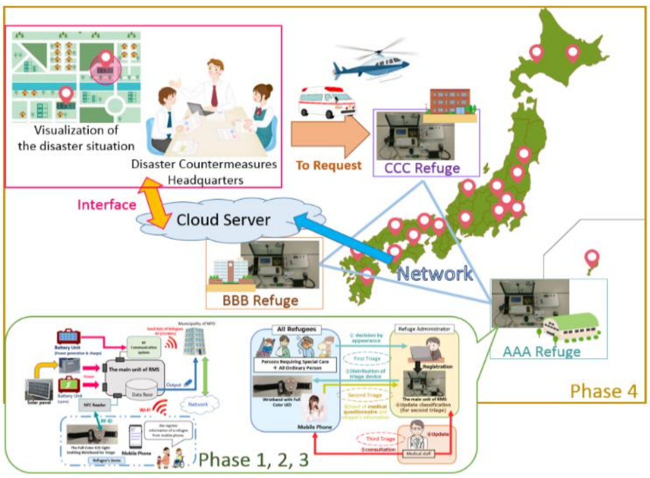


Figure 1: Outline diagram of VIMS

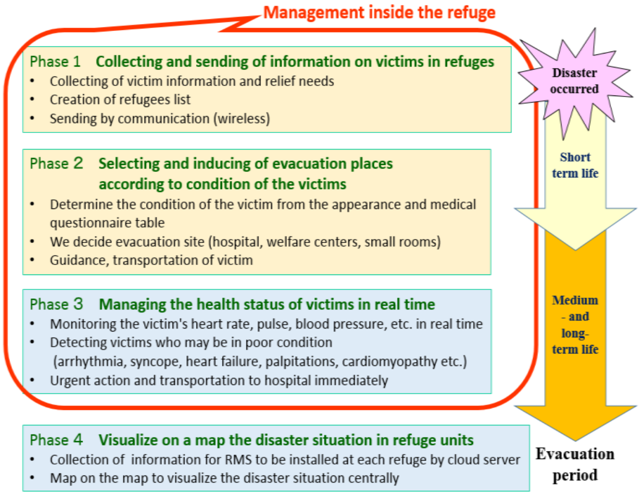


Figure 2: Phased timeline

4 REFUGE MANAGEMENT SYSTEM (RMS) DEVELOPMENT CORRESPONDING TO PHASE 1

Chapter 4 describes the design concept and system development for the RMS corresponding to Phase 1 of the VIMS research and development.

4.1 Concept of RMS for phase 1

We designed RMS with the following concept.

I. The system user

The administrator of the RMS is preferably a civil servant etc. Because, the administrator deals with victim information.

II. Collection and management of victim information

We assume that the installation site of the RMS has a capacity of about 1000 individuals per refuge, such as schools and community centers.

III. Power supply of RMS

Lead battery units charged with electric current generated by solar panels are used as the power source of the RMS. this makes the RMS available even when the infrastructure is not working.

IV. Data collection content

The contents of data to be collected shall be the name, address, gender, age, nationality, emergency contact information, damage situation, and stockpile status of evacuation supplies, items requiring long-term care (pregnancy, chronic illness, and allergy). These items of information concerning the necessary nursing care were added to the questionnaire because they were actually used at the refuges of the Tohoku Earthquake. Additional items to be entered include a User Identifier (UID) to identify the victim and a questionnaire to confirm the health condition. These information is input by victims themselves or staff members who support refugees.

V. Identification for victims

As a method to identify each victim, we use the UID of NFC (Near Field Communication) or RF-ID information with a full-color LED light-emitting wristband distributed in the refuge. By using these data to manage entry and exit in the refuge, we can grasp the situation of the victim's presence in the refuge and track transfer of other victims to other refuges.

VI. How to send victim information

We consider two patterns of transmission methods for victim information when the communication infrastructure is not working. The first method is to connect a wireless unit (simple type radio, amateur radio) to the RMS. By this method, we can send information to the national disaster response headquarters using character data in JSON format or CSV format through any communication protocol. Another way is to save it in a USB flash memory and take it out. After the infrastructure is restored, the system can acquire the victim information through the Internet.

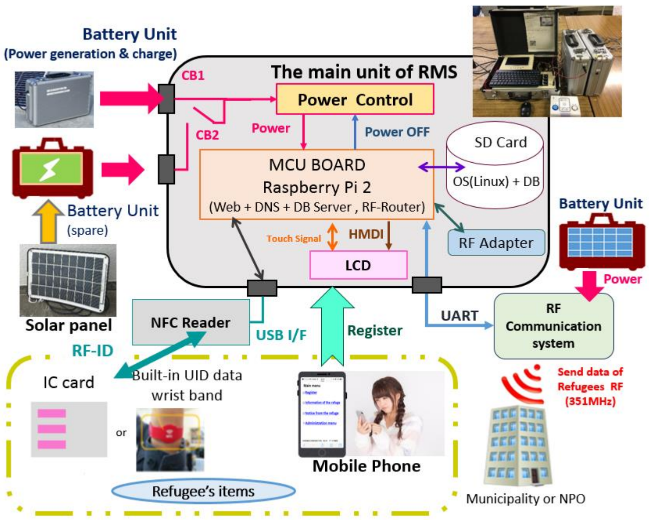


Figure 3: RMS corresponding to Phase 1

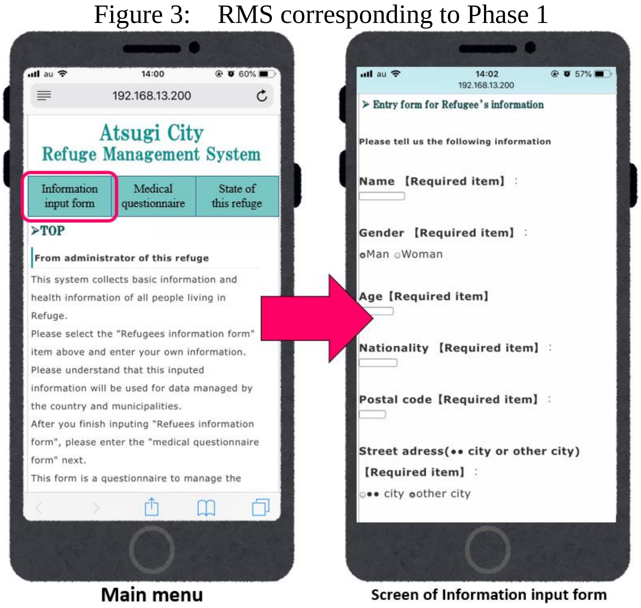


Figure 4: RMS prototype

VII. Communication method between RMS and various information terminals

When inputting victim information, Wi-Fi is used for the communication between the information terminal such as a smartphone and the RMS. If the number of connections with information terminals is insufficient, we can increase the number of repeater devices.

4.2 Refuge Management System (RMS)

In R & D for Phase 1, we prepared an outline design of the refuge management system. The outline is shown in Fig. 3. A commercially available microcomputer board Raspberry Pi Model B (RPI) was used for RMS hardware. We installed an embedded Linux OS (Raspbian) on this hardware and designed a system that can launch Graphical User Interface (GUI) applications. The reason for adopting embedded Linux on hardware is to save power of the RMS itself. The power consumption of an ordinary notebook PC is about 40 ~ 50 [W], which can be reduced to about 1/10 by using RPI. In addition, with regard to the information input of evacuees described in Design Concept Ⅳ, it is necessary to allow for the input of refugee's information from a web browser on a smartphone or mobile phone owned by the victims. Therefore, we constructed a web server using apache2 and developed a PHP application for a disaster victim information input screen. As shown in Fig. 3, the refugee's information registration application was developed using a server-side application that does not depend on the client environment so that victim's information can be input from the WWW browser installed on the client side.

Victims are requested to input personal information such as the name, address, gender, and age using the input screen shown in Fig. 4. In addition, we have added selective buttons asking whether or not the respondent is a pregnant woman, a person with a disability, or requires nursing care, etc., as well as check boxes to input food allergy determined by law [13]. (Wheat, egg, milk, buckwheat, shrimp, crab, peanut, etc.) In consideration of a more detailed description and answer which doesn't correspond to any items, we have also provided a text box for free description. By summarizing such information, it is possible to create a victims roster including the relief needs within the RMS. Based on the above, we will realize a system that picks up the relief needs of the affected people, including those with special needs, and will eventually solve Problem ② in Chapter 2.

By the way, the system of Phase 1 assumes operation at an actual refuge. Therefore, the system must be able to operate even when the power infrastructure is not working. Therefore, we will develop a battery unit with a solar panel and a lead storage battery as the power supply of the RMS, and use two or more such battery units to operate the system. The operation method is described in the next section.

In addition, we developed a method for managing the evacuation using the RMS in preparation for the case where victims irregularly enter or exit from the refuge. We conceived the idea of using NFC cards which the victims usually carry with them. Typical examples of the NFC cards include Edy and Suica etc., and the total number of NFC cards issued in Japan was 33.346 million as of January 2017 [11].

From such background, we thought that we could use an NFC card effectively as a victim's personal identification. Entering and exiting of refuges for victims is done by holding the NFC card over the NFC card reader. By associating the UID of NFC with the already registered victim information in the RMS, it is possible to quickly manage the entry and exit of each victim.

The refugee's roster created by the RMS is organized as an electronic file, and can be saved in a CSV format and a JSON format on an electronic medium such as a USB flash memory. As a result, the data list of evacuees can be taken out to external facilities such as the disaster response headquarters. In addition, amateur radios and simplified radios which are available for the communication even during a large-scale disaster can be connected to the RMS with a serial cable to send a relief request to the disaster response headquarters.

4.3 RMS Prototype to Phase 1

Figure 5 shows an outline of the proposed RMS prototype system. The RMS consists of an RMS server, a battery unit, an NFC reader panel, and a wireless communication device. The main power of the RMS server is supplied from the battery unit. By connecting this unit to a solar panel, the generated electricity can be stored in the lead battery. The lead battery unit is made of an attachment case containing a lead storage battery and a thin solar panel, making it easy to carry. The battery unit requires a minimum charging time of 5 hours under fine weather condition. In order to operate the RMS power supply in the long term, we developed two lead battery units which should be used alternately.

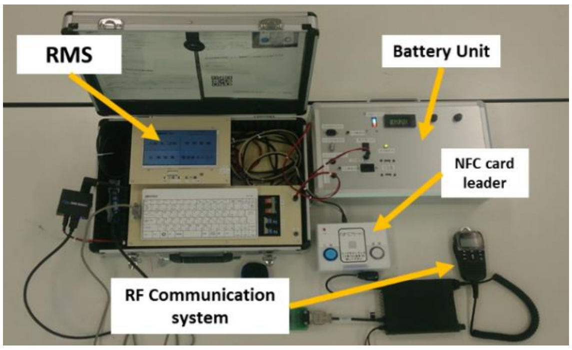


Figure 5: RMS prototype

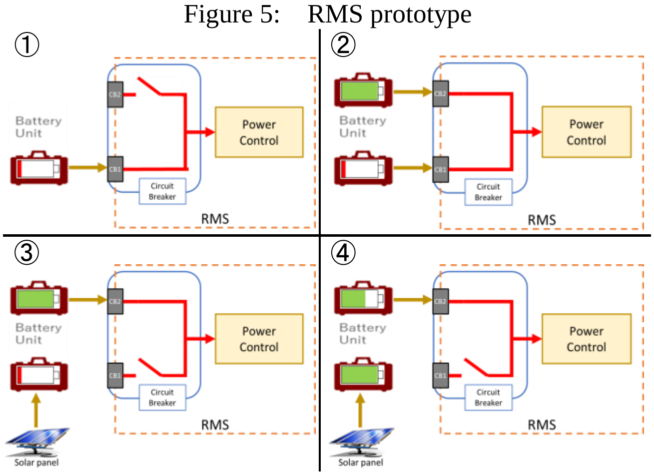


Figure 6: Replacement method of the battery unit

Table 1: Criteria and classification of triage for persons requiring special care [12] [13]

color classification	Class	Decision Criteria	Area
1 (Orange)	Person who needs urgent treatment	Injuries with bleeding Fever, Diarrhea, Nausea, Vomiting, Dialysis patient , Oxygen suction patient	Hospital
2 (Pink)	Person who needs full assistance in daily life	Person who can't eat, walk or excrete independently Bedridden person	Welfare center
3 (Purple)	Person who needs some assistance in daily life	Pregnant women Infants under 3 years of age Half paralyzed person People with intellectual disabilities Wheelchair user	Small room (classroom etc.)
4 (Blue)	People who can act without problems	Person who can walk independently	Large room

Figure 6 shows the procedure for replacing the RMS lead storage battery. The system user confirms that the voltage of the battery unit in use has fallen by looking at the indicator lamp of the battery (Fig. 6 ①). The user then connects the charged battery and performs manual circuit breaker replacement. This makes it possible to replace the battery without shutting down the server power (② and ③) . At the same time, the user charges the low voltage battery using the solar panel (④). By alternately performing these operations, we enabled the RMS to operate for a long time. We actually operated the system using this method for a week. As a result, there was no problem with the RMS MCU. Also, we confirmed that the operating time can be extended without shutting down the system power.

This will allow the RMS to operate even when the power and communication infrastructure are disrupted, and we believe that we can solve Problem ① in Chapter 2.

5 RMS DEVELOPMENT CORRESPONDING TO PHASE 2

Chapter 5 describes the contents corresponding to Phase 2 of the VIMS research and development. Specifically, we developed a system to perform color coding according to the physical and mental condition of the affected people, including persons requiring special care. In addition, we will explain the method to guide victims to appropriate living spaces in the refuge based on judgment from their physical conditions.

5.1 Triage for Persons Requiring Special Care

As prior research corresponding to phase 2 of this paper, M. Ohara. *et al*. proposed "Triage for persons requiring special care" [12] [13]. Table 1 shows the criteria of triage for persons requiring special care proposed by M. Ohara. The table provides decision criteria for the leader of the residents to decide the room allotment for persons requiring special care when entering the evacuation refuge and the priority of transfer to the welfare evacuation refuge. It is different from the Simple Triage and Rapid Treatment method. The proposed triage method uses four colors: orange, pink, purple and blue. Because of this, the colors do not mix, so the two triage methods can be used simultaneously.

5.2 Problem of the proposed triage and solution

This time, we considered an RMS that incorporated the "triage method for persons requiring special care " proposed by M. Ohara et al. In small-scale refuges, medical staff may not be present. Even under such circumstances, we consider it is necessary to use the proposed triage to make classification decisions according to the condition of the affected person smoothly. So, we have devised a system that will determine color classification automatically to some extent even without medical staff, by having the victims respond directly to an electronic questionnaire.

Considering persons requiring special care to change their condition, our proposed electronic questionnaire form (e-questionnaire) is a system that can respond anytime during the life of a refuge after a large-scale disaster. This makes it possible to constantly monitor and check the health status of the victims.

A full-color LED light-emitting wristband is used as a method for identifying the condition of the victim. As a result, the target person in victims can be instantly identified by the color of the LED light, so refuge staffs can quickly navigate them to the appropriate refuge space. However, if the subject is color-coded by triage, the surrounding eyes may be worrisome, which may lead to mental and psychological stress. For this reason, we have developed a system that allows victims to identify triage colors by emitting light only when necessary using a full-color LED light-emitting wristbands distributed by the refuge staffs.

5.3 Positioning of Related Research and Our Proposal Methods Concerning Triage

By the way, various companies and universities are currently studying electronic triage devices for use in emergency medical environments during disasters [14] [15]. A typical example is the "Electronic Triage Tag" developed by T. Higashino et al [16] [17]. This is a system in which a sensor device with a small CPU is installed in a device that measures the pulse, respiratory rate, blood oxygen concentration (SpO2), etc. of injury persons. This enabled START (Simple Triage and Rapid Treatment Triage) method assessment to be performed on the device. In addition, it is a system that can monitor the position and medical condition of the victim in real time by transferring their data to the server via the ad hoc network.

Compared to these studies, our proposed system has many similarities in terms of managing the health of victims using triage. However, the proposed system has a different use from the traditional electronic triage system. This is based on the "triage method for persons requiring special care" proposed by M. Ohara, and is intended for all victims living in refuges for a long time after the disaster.

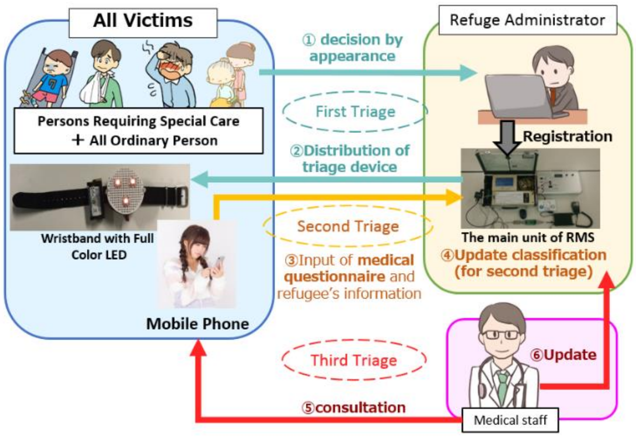


Figure 7: Victims Management Triage

Therefore, in the phase-2 system of this proposal, we have devised a system of additional functions combining the " triage method for persons requiring special care " proposed by M. Ohara with the RMS of phase 1 and a full-color LED light-emitting wristbands described above.

5.4 Victims Management Triage（VMT）

The method proposed by M. Ohara is a two-step classification method including the "primary triage" to judge by appearance, followed by the "secondary triage" based on the decision of medical staff [13]. On the other hand, the triage we propose is a three-stage system, which we call the Victims Management Triage (VMT). Figure 7 shows the VMT system. Section 5.3 describes how to use the VMT. Figure 7 shows the management method in the refuge using the VMT. This system should be used at least three times to perform three types of triages which we call "First Triage", "Second Triage", and "Third Triage". "First Triage" is an initial classification by refuge administrators and staff. "Second triage" is done by the evacuees themselves by entering the questionnaire. After that, the method of making a judgment based on medical examination by a doctor etc. is called "Third Triage". We describe the method of each triage below.

In the first triage, the refuge manager and staff will determine the triage classification from the appearance for all victims entering. At this time, the staff will distribute a full-color LED light-emitting wristbands (VMT wristbands) to all victims. The reason for judging the status of evacuees only by appearance in the first triage, is to prevent congestion at the entrance and reception of the refuge.

After several hours from the start-up and looking at the people inside the refuge, the second triage is performed at an arbitrary timing. In the second triage, victims can enter mental and physical information by selecting the item of the e-questionnaire table from the RMS registration application shown in Fig. 8. Concretely, they register their personal information (name, address, disaster situation etc.) in the RMS, and at the same time, make an inquiry by self-report which inputs the health condition such as the presence or absence of injury or disease.

The third triage is to re-determine the classification from the information obtained by the second triage and from the consultation of victims by a patrol doctor and medical staff. Because the conditions of all victims change with the long-term life in the refuge, the e-questionnaire table can always be changed by the victims themselves, and the color of the triage can be instantly updated according to the condition of the person requiring consideration or the judgment of the traveling doctor.

At this time, the refuge administer can easily find the person who corresponded to the 4 triage colors by turning on the switch of their VMT wristbands. As a result, the staff will be able to quickly guide the victims to the appropriate evacuation space. On the other hand, when the transport destination such as a hospital or a welfare facility is not ready, the victim turns off the light of the VMT wristbands and stands by at a waiting place or the like until the preparation is completed.

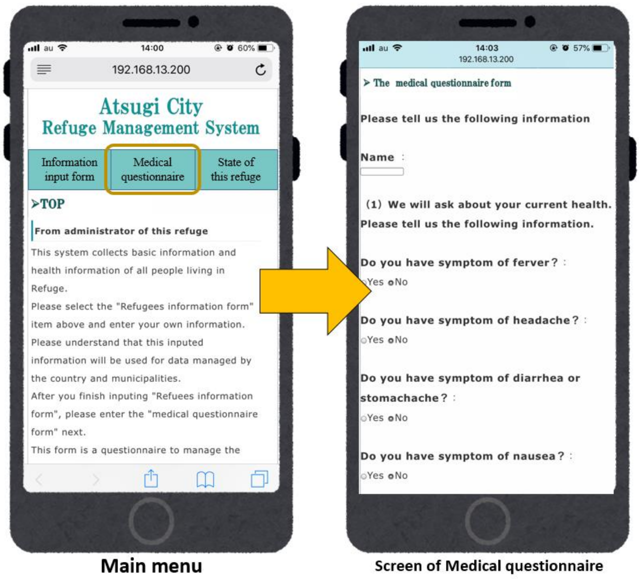


Figure 8: Screen of Medical questionnaire

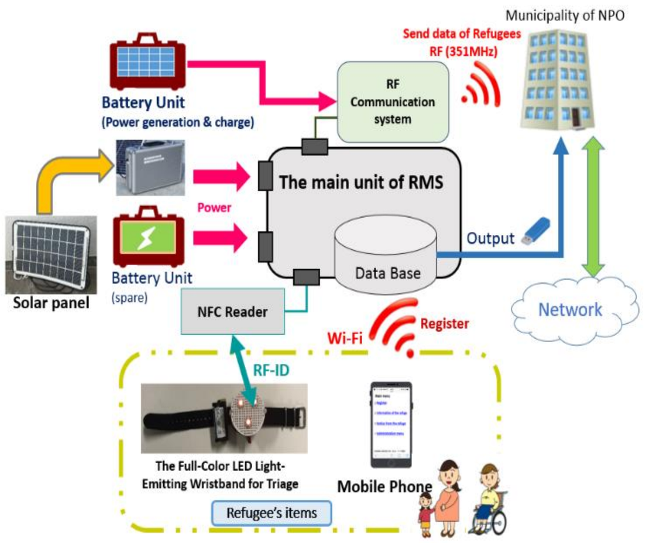


Figure 9: RMS for Phase 2

Table 2: Classification colors for VMT and handling method

Medical Questionnaire	Class (Color)
・ MQ1: Two or more items selected from fever, headache, abdominal pain (diarrhea), and nausea. ・ MQ1: A total of four or more items selected ・ MQ4: "Treatment required within 3 days" selected	1 (orange)
・ MQ2: "Inoperable without assistance" selected for two or more items	2 (pink)
・ MQ1: 1~3 items selected from the 7 items ・ MQ2: "Some actions require assistances" selected for two or more items (Accepting up to one caregiver) ・ MQ3: Pregnant woman (9 months ~) or a family with an infant ・ MQ4: "No treatment required within 3 days" selected ・ The person who should move to a private room according to judgment by the administrator based on the contents of the description column	3 (purple)
・ MQ3: Pregnant woman (~8 months) ・ Person who can stay in a large room according to decision by the administrator or medical staff based on the content of description	4 (blue)

5.5 e-questionnaire table from the RMS

Figure 8 shows the registration screen of the electronic questionnaire. By answering this electronic questionnaire, the RMS automatically performs the VMT. Table 2 shows triage classification criteria for the answers of the electronic questionnaire.

As a specific criterion for the triage from e-medical questionnaire, the e-medical questionnaire form has four question items: "MQ 1: health information", "MQ 2: presence / absence of assistance in daily behavior", "MQ 3: prepartum / postpartum women", and "MQ 4: existing medical treatment (such as a chronic illness) ". We will explain each check item in detail. In MQ 1, the respondent is asked to select any symptom that applies to him/her from the seven items of fever, headache, abdominal pain (diarrhea), nausea, dizziness, cough (sputum), and runny nose. MQ 2 is a question about whether the respondent requires assistance in any of the three items of walking, meals, and excretion. The answer is selected from three options: "no problem", "some actions require assistances", and "inoperable without assistance". MQ 3 is an item that asks whether or not the respondent is pregnant or has any infant, and if applicable, whether the pregnancy period has exceeded nine months. MQ 4 is an item that asks whether a medical treatment such as dialysis and medication is necessary, and whether the treatment is necessary within 3 days. Also, in consideration of detailed descriptions and answers that do not correspond to any items, a text box for free description is provided. The RMS automatically determines the color of the division into the triage based on the criteria of Table 2 from the answer result of this e-examination table.

It is thought that this will solve Problem ④ " The QoL of the victims declines in refuge life " mentioned in Chapter 2.

5.6 RMS and triage wristbands Prototype System of VMT

An outline of the VMT prototype proposed in this paper is shown in Fig.9. The VMT is a system that adds the proposed VMT operation functions and the LED wristband to the RMS of Phase 1. The information input screen of a person requiring consideration and the questionnaire for confirming the health condition can be accessed from a WWW browser installed on an information terminal such as a smartphone. In addition, the administrator distributes VMT wristbands to all evacuees. In order to identify persons who are color-coded, we devised VMT wristbands so that it indicates the color code of each person by LED emission and helps to smoothly guide the person to an appropriate refuge spaces. We adopted IEEE802.15.4 standard wireless communication for communication between LED wristband and RMS. The reason for adopting this standard is that it can be connected to more than 40,000 nodes theoretically in addition to lower power consumption than the Wi-Fi standard.

Figure 10 shows an overview of the LED wristband system developed this time. We used a commercially available microcomputer AT mega328P for the wristband Micro Control Unit (MCU). Also, the development of the AT mega328P microcontroller was done in Arduino-like C language. We also used IEEE802.15.4 standard XBee for the wireless communication module of this system. We communicated by defining the Xbee on the RMS side as the coordinator and the Xbee on the full color LED wristband side as the device.

Figure 11 shows the communication procedure between the RMS and VMT wristbands. We designed the RMS to transmit a request command to VMT wristbands so that the emission color of the LED can be changed as needed. We defined the command length of this system as 10-byte character data. The 10-byte character data includes 7 bytes as the device number, 1 byte as the color code, and the remaining 2 bytes as the delimiter (combination of CR and LF). The device number is the unique ID of VMT wristbands and is treated as hexadecimal characters. The color code is O, P, M, or B in order from class 1 to class 4 for VMT in Table 2, and the system lights the LED in the color corresponding to the input code.

Figure 12 shows a wireless communication system between the RMS main unit and a VMT wristband. Wireless communication between devices is performed by the RMS main unit acting as the coordinator and a VMT wristband acting as the end device. Also, if data on a VMT wristband is difficult to reach, the refuge manager can use a router device for relaying. This enables the system to communicate with all victims.

6 QUESTIONNAIRE EVALUATION OF VIMS

At Kanagawa Institute of Technology Academy Festival and open campus, we conducted a questionnaire evaluation of the prototype system for the RMS which is the solution method of Phase 1, and the VMT which is the solution method of Phase 2. We conducted this questionnaire after introducing the outline of this research and the demonstration of the prototype. Chapter 6 describes the contents and results of the questionnaire evaluation regarding them.

6.1 Questionnaire Evaluation on Phase 1 Prototype

Questions in the Phase 1 prototype (RMS) correspond to questions Q1.1 to Q1.3 shown in Table 3. The questionnaire target was 38 individuals who visited the RMS exhibits. In this questionnaire, Question 1.1 (Q1.1) and Question 1.2 (Q1.2) are four choices. Also, we asked the respondents to describe and answer Question 1.3 (Q1.3) arbitrarily. For the results for Q1.1 and Q1.2, the number of answers to the question is treated as a score, and the result of them is shown as a column graph with each average score (the full score is 4.00 points) in Fig. 13.

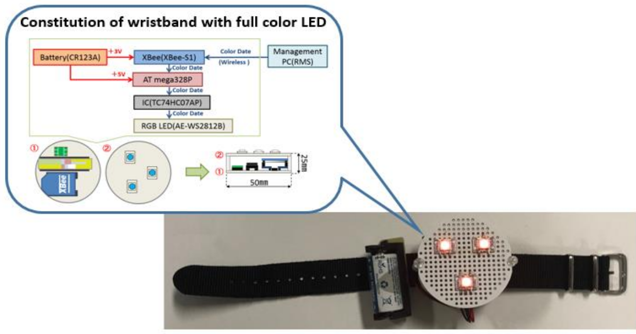


Figure 10: A full-color LED light-emitting wristband system

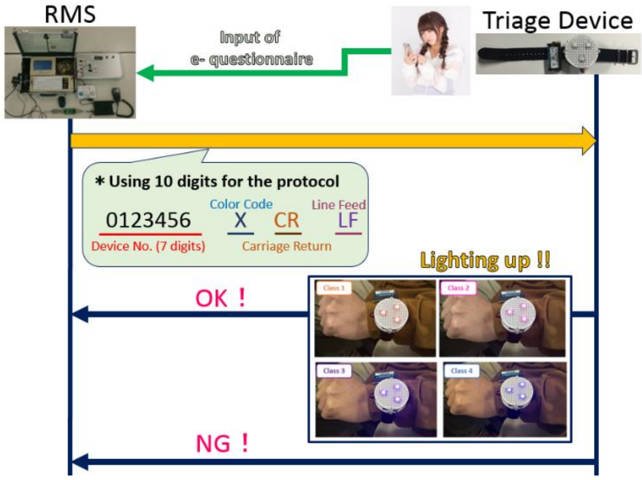


Figure 11: Communication procedure between RMS and VMT wristbands

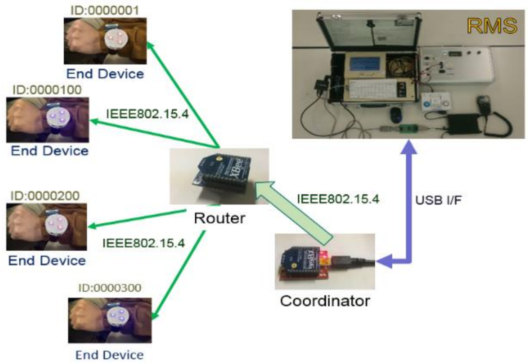


Figure 12: Wireless communication system between

Table 3: Questionnaire questions on VIMS prototypes

	Contents
Q1.1	Did you understand the outline of RMS? 4 : Good 3 : Neither 2 : Poor 1 : Very poor
Q1.2	Do you feel like using this system when a large-scale disaster occurs? 4 : Good 3 : Neither 2 : Poor 1 : Very poor
Q1.3	Please tell us about opinions and requests about RMS. (descriptive expression)
Q2.1	Was it easy to input evacuation information and fill in *e-medical questionnaire form* on the RMS? 5 : Very good 4 : Good 3 : Neither 2 : Poor 1 : Very poor
Q2.2	Do you think that VMT is necessary after listening to the explanation of the electronic triage for persons requiring special care? 5 : Very good 4 : Good 3 : Neither 2 : Poor 1 : Very poor
Q2.3	Do you think the quantity of items in the VMT evacuation information and e-medical questionnaire form is appropriate? ・ Too much ・ Much ・ Appropriate ・ Less ・ Too less
Q2.4	What other kinds of people do you think need consideration other than the consideration items listed in "Decision Criteria" in Table 1 (of this paper)?

In Q1.1 "Do you understand the outline of RMS?" 29 answered "Very good" and 8 answered "Good". It was found that 97% of the visitors understood the system. Also, the average score for this question was 3.74 points.

In addition, for Q1.2, "Do you feel like using this system when a large-scale disaster occurs?", 21 visitors responded "Very good", and 9 responded "Good". On the other hand, it was thought that about 20% of the visitors could not decide to use this system because there were 7 non-responders. The average score for Q1.2 was also 3.64 points, and it was found that more people felt dissatisfaction than Q1.1. We asked the reason for one visitor who answered "Poor" in this question. As a result, he said "Because RMS and battery unit are a bit heavy."

Furthermore, the following answers were obtained in response to the question Q1.3 "Please let me know if you have any feelings, opinions or requests about the refugee management system".

- Pet information should also be included in the RMS.
- It is better to reflect facial pictures on the RMS.
- The system should also be available at a time other than when a disaster occurs.

6.2 Questionnaire Evaluation on Phase 2 Prototype

Questions in the Phase 2 prototype (VMT) correspond to questions Q2.1 to Q2.4 shown in Table 3. The questionnaire target was 20 individuals who visited the RMS exhibits. The answering method was to distribute the questionnaires asking questions Q2.1 to Q2.4 and to have the visitors fill in the questionnaire after the demonstration, as in the method described in the previous section. Questions 2.1 (Q2.1) to 2.3 (Q2.3) are five choices. Also, we asked the respondents to describe and answer Question 2.4 (Q2.4) arbitrarily. For the results, the number of answers to the question is treated as a score, and the result of them is shown as a column graph with each average score (the full score is 5.00 points) in Fig. 14(Q2.1&Q2.2) and shown as a pie chart Fig. 15(Q2.3).
As for the medical questionnaire about Q2.1, 10 visitors responded "Very good", 6 visitors responded "Good", 2 visitors responded "Neither", and 2 visitors responded "Poor" about Q2.1 "Was it easy to input evacuation information and fill in e-medical questionnaire form on the RMS?". Also, the average score for Q2.1 was 4.20 points.

Next, about Q2.2 "Do you think that VMT is necessary after listening to the explanation of the electronic triage for persons requiring special care?", 14 visitors responded "Very good", 6 visitors responded "Good". There was no one who answered "Neither", "Poor" and "Very poor". Also, the average score for Q2.2 was 4.70 points.

On the other hand, about 50% of the respondents answered "appropriate" to Q2.3 "Do you think the quantity of items of the VMT evacuation information and e-medical questionnaire form is appropriate?" The remaining 50% answered "too much" or "much".

As for Q2.4 "What other kinds of people do you think need consideration other than the consideration items listed in "Decision Criteria" in Table 1 (of this paper, section 5.1)", there was an opinion that "children unattended by their parents also need consideration".

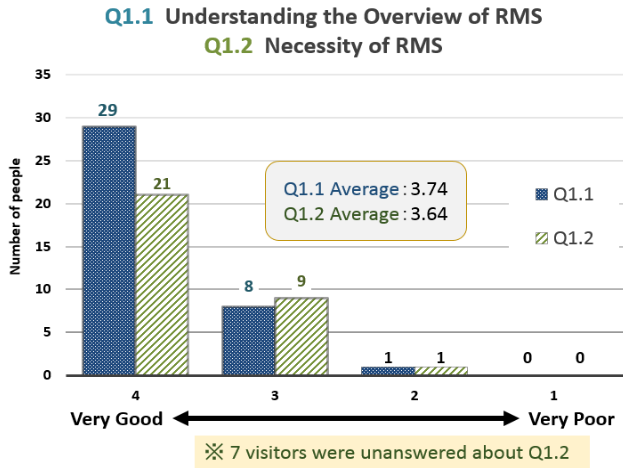


Figure 13: Results of Questionnaire (Q1.1 & Q1.2)

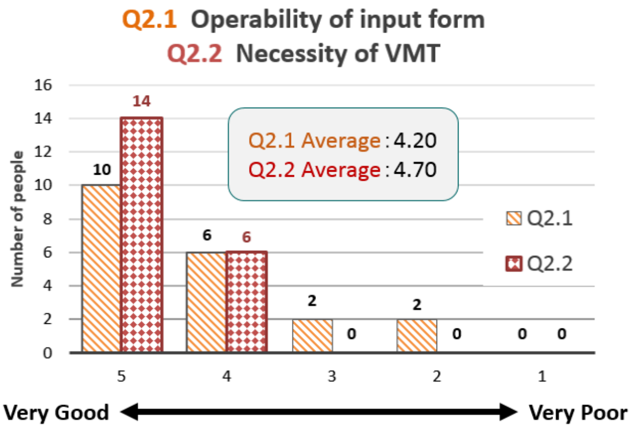


Figure 14: Results of Questionnaire (Q2.1 & Q2.2)

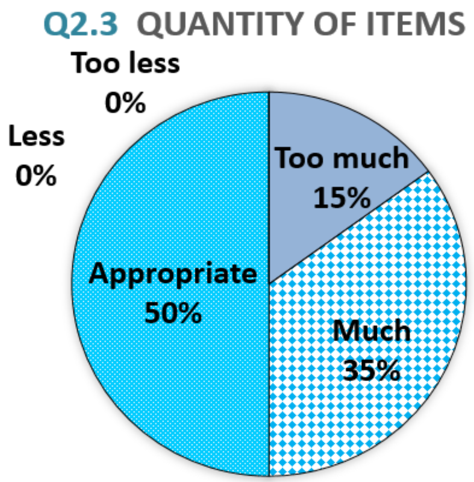


Figure 15: Results of Questionnaire (Q2.3)

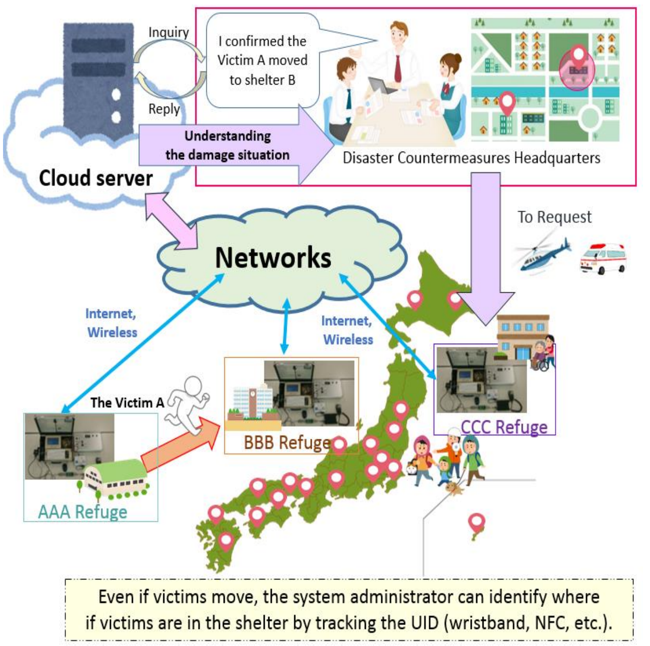


Figure 16: Attendance management method of victims in refuges nationwide

6.3 Conclusion of Questionnaire Evaluation

As a result of conducting the questionnaire about the prototype system of Phase 1 and Phase 2, it is thought that many people highly appreciated the significance of the proposed VIMS system. Also, the average score was high in both Phase 1 and Phase 2. Above all, in the questionnaire on VMT in Phase 2, it was found that 100% of people understood the necessity of this research.

On the other hand, as there were also suggestions regarding this system, we also felt that it is necessary to study and improve it in the future.

7 CONCLUSION

This paper described the realization of the Victims Information Management System (VIMS), which centrally manages the disaster situation and the relief needs (including the need for relief for persons requiring special care) in refuge areas around the country. Since this VIMS is a large-scale system, we are conducting research in four phases, and promoting development toward the final goal of realizing the VIMS. Therefore, in this paper, we described in detail Phases 1 and 2 completed as prototype development.

In Chapter 4, we proposed an RMS corresponding to Phase 1 (“The existing disaster victim support systems cannot operate when there is a power and communication infrastructure disruption immediately after the disaster occurs”), and a method for collecting and sending the relief needs in conjunction with a solar charging unit. We also proposed a system to manage entry and exit in refuges using UID management owned by the victims.

We proposed VMT for Phase 2 (“The existing systems cannot adequately collect the needs for persons requiring special care”) in Chapter 5. Specifically, we have proposed a system that mechanically determines the health status of victims from electronic questionnaires, and selects and guides the victims to evacuation spaces according to the patients' status.

In addition, a questionnaire survey on the prototype development of Phase 1 and Phase 2 of the VIMS research and development confirmed the effectiveness of this proposed system because many people answered that this proposed system is necessary.

Therefore, it is thought that Problems ① to ④ of the existing victim support systems mentioned in Chapter 2 can be solved by the contents of Phase 1 and Phase 2. The solution to Problem ③ was not described in detail in this paper, but as shown in Fig. 16, we consider that this can be solved by installing the proposed RMS at refuges around the country as well as connecting the RMS via a network and centrally managing it with a cloud server. By using the UID of the victims and the evacuation entry and exit management information, system administrators can track victims and check their presence.

In the future, we plan to study a system to acquire biological information of the victims in the refuge corresponding to Phase 3 and manage their health. In addition, we are developing an application that visualizes disaster information and disaster relief needs in a unified manner by mapping on a map the disaster victim information obtained from the RMS in various locations corresponding to Phase 4.

REFERENCES

[1] Ministry of Internal Affairs and Communications of Japan : "Survey result report on evacuation support of people who need assistance during disasters" , (Mar.2009). <http://www.shakyo.or.jp/research/2014_pdf/20140530_jisedai.pdf >

[2] Cabinet Office Government of Japan : "Survey result report on promotion of comprehensive measures on evacuation" <http://www.bousai.go.jp/ kaigirep/houkokusho/hinan_taisaku/houkoku.html>, (Sept.2016).

[3] Article 8, Paragraph 2 of the Disaster Countermeasures Basic Law of Japan, <http://www.japaneselawtranslation.go.jp/law/detail_main?re=01&vm=04&id=3092>

[4] Information and Communication Technical Committee in Japan : ""Security guideline for ICT tool (cloud service) useful for large-scale disaster response" , First Edition,

[5] K.Akasaka, K.Abe et al., "Development and evaluation of Refuge Management System using ICT at the time of Large scare disaster (In Japanese)" , Information Processing Society of Japan, CDS, Vol.7, No.3, pp.15 - 25, (Sept.2017).

[6] K.Akasaka, K.Abe et al., "Proposal on Refuge Management System for Large Scale Disaster for Persons requiring special care by using triage" , IEEE BigComp2019 proceedings, pp.525-531, (Feb. 2019)(ISBN:978-1-5386-7789-6).

[7] Microsoft of Japan : "About the earthquake disaster recovery support system" , <http://www.microsoft.com/ja-p/citizenship/disasterrelief/ default.aspx>

[8] A.Hamamura, N.Egusa et al., : "AkariMap : Evacuation Support System for Everyday Use in Offline Environment (in Japanese)" , IPSJ DICOMO 2014,pp.2070-pp.2078, (Jul.2014).

[9] Nishinomiya City Urban Improvement Corporation : "Outline of evacuees support system" , <http://www.nishi.or.jp/homepage/n4c/hss/>

[10] N.Kosaka et al., "Applicability Assessment of An Emergency Management Support System "KADAN"" , IEEE BigComp2019 proceedings, pp.532-537, (ISBN: 978-1-5386-7789-6), (Feb. 2019).

[11] Bank of Japan Settlement Agency Settlement Trends, <https://www.boj.or.jp/statistics/set/kess/release/2017/kess1703.pdf>, (March 2017).

[12] M.Ohara , et al., : "Residents in the evacuation center at the occurrence of a disaster Efforts of people who need assistance triage" , Japan Disaster Reconstruction Association Journal, Volume No. 10 Vol.6-No.1, pp. 19-24, (Jun. 2014).

[13] Cabinet Office Government of Japan : "Guidelines for securing and managing welfare evacuation shelters" , (April 2016). <http://www.bousai.go.jp/taisaku/hinanjo/pdf/1604hinanjo_hukushi_guideline.pdf>

[14] H.Kojima, K.Nagahashi, K.Okada, "Proposal of the Disaster-Relief Training System Using the Electronic Triage Tag" , Advanced Information Networking and Applications (AINA) 2011 IEEE International Conference on, pp. 256-263, DOI: 10.1109/AINA.2011.104 (2011).

[15] T.Massey, "The Design of a Decentralized Electronic Triage System" , <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839501/>, (2006).

[16] T.Imarura, T.Higashino *et al.*, : "Evaluation of Electronic Triage Tag to observe injured person's condition in Real Time *(In Japanese)*" , SIG Technical Reports, 2010-MBL-53 -No.5, pp.1-8, (Mar. 2010).

[17] N.Kiyama, T.Higashino *et al.*, : "An Advanced Electronic Triage System for Rescue Operations in Disaster *(In Japanese)*" , Information Processing Society of Japan, Vol.51-No.9, pp.1916–1926, (Sep. 2010).

(Received November 22, 2019)
(Revised February 5, 2020)

Kosuke Akasaka received the B. E. degree from Kanagawa Institute of Technology in 2018. He is currently engaged in research on ICT disaster support management systems at Department of Electrical and Electronic Engineering, Graduate School of Engineering, Kanagawa Institute of Technology.

Takayo Namba received Ph.D. from Nagano College of Nursing in 2009. She was a professor of the faculty of nursing, Kanagawa Institute of Technology from Apr.2017 to Mar.2019. She has been a researcher of the faculty Health Science since Apr.2019. Her research interests about Elderly Abuse and Co-dependency, Educational program and Visiting nurses. She is a member of the Japan Academy of Nursing Science and the Japan Society for Health and Welfare etc.

Masao Isshiki Ph.D Prof. Professor of Kanagawa Institute of Technology since 2011. Director of Smart House Research Center since 2012. JSCA HEMS-Task Force chairman since 2011 and also working for many chairmen of METI, MIC and MLIT. Project professor of Keio University since 2009 to 2015 responsible for W3C/keio site manager. Fellow of ECHONET Con-sortium. Masao holds a Ph.D. degree in heat transfer engineering from the Tokyo University of Agriculture and Technology, and a master's degree in mechanical engineering from the University of Tokyo Institute of Technology. He currently lives in Tokyo, Japan.

Keiichi Abe received Ph. D. in informatics from Shizuoka University in 2012. He has been an associate professor of the Faculty of Creative Engineering, Kanagawa Institute of Technology since 2016.His research interests include smart system using IOT technology, Interactive system using Robot and CG, and Victims Information Management System at Large Scale Disaster, HEMS. He is a member of Information Processing Society of Japan, the institute of Electronics, Information and Communication Engineers of Japan and a member of IEEE.

Industrial Paper

Proposal of IoT Communication Method for the Rice Field

Mikiko Sode Tanaka[*]

[*]Global Information and Management, International College of Technology, Japan
sode@neptune.kanazawa-it.ac.jp

Abstract -Social application of the internet of things (IoT) and commercialization of the low-power wide-area network (LPWA) by telecommunications carriers are progressing, and efforts to put IoT into practical use using LPWA are ongoing. LPWA is expected to be applied for automatic meter reading of water, gas, etc., and large-scale demonstration experiments have been conducted in Japan. Furthermore, application to fields requiring real-time performance is also being considered. The telecommunications carrier adopts a direct method, but the direct method is not suitable for a usage method in which data are transmitted from IoT nodes at the same time. To solve this problem, a method that gathers data in the shortest time is proposed to attain the capacity of the network fully and provide data at the time desired by the users. In this paper, we provide an example of the field server for rice fields and propose a method for gathering data in a short time. The method is effective with low power consumption. Comparison results with a conventional method using a simulation demonstrate that the proposed method can collect all data in the shortest amount of time. Moreover, the power consumption is lower than that of the conventional method. It was also confirmed that the rate of increase in the time necessary for the parent node to collect data due to an increase in the number of sensor nodes is lower than that of other methods. Therefore, this proposal makes it possible to transmit the situation of a rice field in a timely manner and can be considered to be a useful method for reducing the burden of agricultural work. In addition, it is effective for other applications that require real-time performance.

Keywords: LoRa, Rice field, Data gathering, Multihop routing

1 INTRODUCTION

Social application of the internet of things (IoT) is progressing. Especially, in factory management, IoT has become an indispensable basic tool [1]. In addition, the commercialization of the low-power wide-area network (LPWA) by telecommunications carriers is progressing, and efforts to put IoT into practical use using LPWA are ongoing [2],[3]. Figure 1 shows an illustration of the network provided by telecommunications carriers. With LPWA, it is possible to cover a wide range of areas and also to collect data in both homes and buildings and units of towns and cities.

LPWA is excellent for low-power and long-distance transmission and is expected to be applied for automatic meter reading of water, gas, etc., and large-scale demonstration experiments have been conducted in Japan [2],[3].

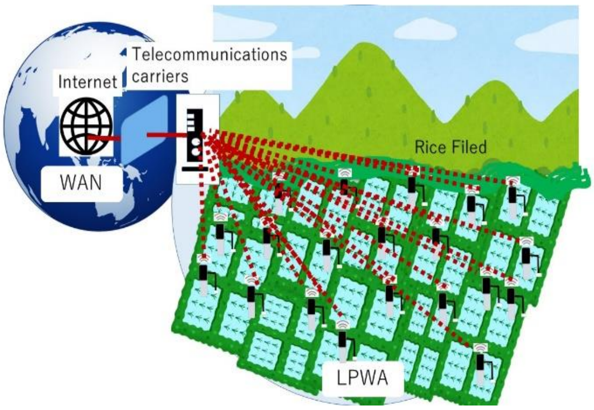


Figure 1: LPWA network for IoT

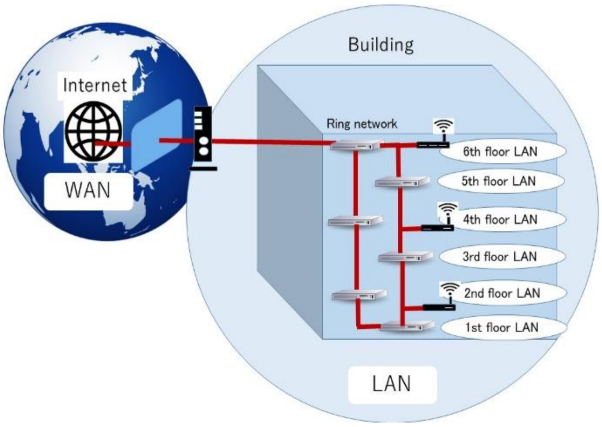


Figure 2: Wired LAN + Wi-Fi network for a building

Automatic meter readings such as those for water and gas are suitable for LPWA because the timing for acquiring and transmitting data is not strictly regulated, and data for a certain period can be sent within an arbitrary period. Furthermore, application to fields that require real-time performance is also being considered [4].

A field server for rice cultivation management system for rice fields is an application that requires real-time performance. The main function of the rice field server is water level management. The water level must be managed in units of several cm in the months immediately after rice planting, and real-time performance is required. However, the telecommunications carrier adopts the direct method. As a result, a conflict may occur and farmers may not be able to see the data in the time they want. Therefore, direct method is not suitable for a usage method in which data are transmitted from IoT nodes at the same time.

When there are many applications that acquire and transmit data at intervals of several minutes, data transmission concentrates at the same time and a conflict occurs. For this reason, to prevent data transmission from occurring at the same time, telecommunications carriers apply an operation method that reduces the probability of competition by providing a difference in usage fees. However, the conflict problem has not been solved.

There are many cases in which companies construct networks with a wired large area network plus wireless fidelity (LAN + Wi-Fi). Figure 2 shows an illustration of the LAN + Wi-Fi network in a building. This method is effective in terms of security and cost. Cost and security are also important items for rice cultivation management system. Thus, when considering a communication system for rice cultivation, there is a high demand for building a wireless network with the same structure as a wired LAN + Wi-Fi network in a company. The most important reason is cost. In addition, to prevent the water level management technique from being stolen, it is desired to manage the water level data locally. However, there is currently no local area wireless network system for rice fields that can realize this objective.

In this paper, we assume an application that acquires and transmits data at an interval of tens of minutes, particularly the system for rice cultivation. In addition, a wireless network that connects all rice fields locally is assumed. The wireless network is of the same type as a wired LAN + Wi-Fi network in a company. We propose a communication protocol to realize real-time performance with this wireless network.

Comparison results with a conventional method obtained using a simulation demonstrated that the proposed method can collect all of the data in the shortest amount of time. Moreover, it was confirmed that the power consumption is also lower than that of the conventional method. Furthermore, it was confirmed that the rate of increase in the time necessary for the parent node to collect data, due to the increase in the number of sensor nodes, is lower than that of other methods.

2 REQUIREMENTS FOR FIELD SERVERS SUITABLE FOR JAPANESE RICE FIELD

In Japan, rice produces a lower income than other crops. Table 1 shows the profit structure of rice production revenue. The unit of 1 pyo is equal to 60 kg, and 1 tan is 0.1 ha. The standard yield per tan is 9 pyo, and this amount of rice can be sold in the market for about 13,000 yen. If the cost is deducted, a profit of about 50,000 yen can be obtained per pyo [5]. In other words, in Japan, farmers can only make a profit of 50,000 yen per year from 1 tan (Table 1). On the other hand, field servers rent for 8,280 yen per month [6]. Thus, the rent for 6 months costs 50,000 yen. This calculation indicates very little profit. Therefore, the introduction of field servers has not progressed among rice farmers. From these facts, in order to introduce the field server practically, the selling price must be less than 10,000 yen. The manufacturing price can be assumed to be around 3,000 yen. Therefore, we assume a field server that can realize this price.

Field servers should not interfere with the farmer's work. Because large agricultural machines are used in the rice field, the field server needs to be easy to move. Therefore, its height must be less than 1 m, and it should be as small as a lunch box and compact. Additionally, because rice fields do not have a power source, it is necessary to operate the servers with batteries from rice planting to rice harvesting. Therefore, intermittent operation is necessary.

To realize an affordable field server, we decided to use a long-range (LoRa) network that does not require a communication line usage fee and can transmit a great distance. LoRa can send more data than Sigfox [7] and is considered more effective in an agricultural field having a relatively large amount of sensor data. In addition, the protocol and frame format can be created freely, so it is suitable for special applications that require performance and price. LoRa can build and operate all architectures including base stations by itself, and its specifications are open. For this reason, it is easy to build a system for special purposes.

Furthermore, to build a reasonable price field server for rice cultivation, it is necessary to reduce the number of parts to the minimum limit. For example, crystal elements and GPS, which are expensive and highly accurate clock sources, cannot be used. In addition, there must be a lightweight communication protocol that operates without expensive components. For this reason, the design must accept time errors, which is a fatal problem for communication systems. To solve the problem of time error, it is necessary to take a time slot longer to be able to tolerate time error. The time slot is a time interval for sending data.

Table 1: Profit structure of rice production revenue

	Breakdown		Total
Income	1 pyo (60 kg) Rice crop yield (1 tan)	13,000 Yen 9 pyo	117,000 Yen
Spending	Fertilizer (1 tan): Herbicide and pesticide (1 tan) Agricultural machinery fee (1 tan) Personnel costs (1 tan)	15,000 Yen 10,000 Yen 20,000 Yen 15,000 Yen	60,000 Yen

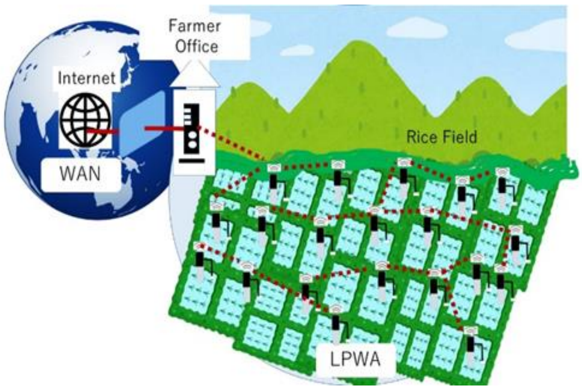


Figure 3: Proposed LPWA network for rice field

A rice cultivation management system using a field server focused on in this paper was developed to manage the field smoothly and flexibly. The configuration is shown in Fig. 3. The system consists of field servers, a parent node, and a cloud service. From the perspectives of cost and security, farmers desire a local wireless network like the wired LAN + Wi-Fi network in a building. Farmers want to check the water level in every hour within 5 minutes after sensor data are acquired.

The field server is installed in the rice field. After acquiring the sensor data, it transmits the data to the parent node via the LoRa wireless network. The parent node aggregates the sensor data from the field servers installed in each rice field and transfers these data to the cloud via a 3G line or Wi-Fi. Cloud services are provided by applications such as smartphones and tablets and web pages. The system provides functions such as water level warnings, work plan proposals, preservation of work records, etc., for the farmers.

The average amount of rice field held by a farmer in Japan in 2010 was 105.1a [8], and the effective transmission distance of LoRa is about 3 km. Therefore, the wireless communication protocol proposed in this paper assumed a rice field area of 3 km × 3 km and a maximum number of field servers to be installed of 100. In the farmer office, the parent node always wake up and have a power supply.

3 BASIC WIRELESS COMMUNICATION PROTOCOL

To operate for 6 months in an environment without a power supply, it is necessary to turn off the power when sensor data acquisition and communication are not being performed. To realize these operations, an intermittent operation communication protocol and time synchronization technology are required. Time synchronization technology has been studied extensively [9]-[12].

For example, a time synchronization method using a radio clock has been proposed [13]. This method requires about 3 minutes to set the time. For this reason, the operation time becomes longer and it is difficult to realize low power consumption. In addition, the received signal (time code) needs to be converted to a format that can be handled in the field server, but a program that performs this processing cannot be written to a low-cost microcomputer, such as a PIC microcomputer, with little memory. Therefore, it is difficult to use it for a rice field server.

A time synchronization method using GPS has also been proposed [10]. To use this method, it is necessary to install a GPS receiver module in each field server, which leads to an increase in the initial cost at the time of introduction. For this reason, it is difficult to use the method for field servers where there is a desire to lower the cost barrier of introduction.

Reference broadcast synchronization (RBS) has been proposed as a time synchronization method for wireless networks [11]. Time synchronization is performed by exchanging time information between field servers that receive a certain broadcast packet. However, RBS has the problem of increasing power consumption because the amount of information exchanged increases in proportion to the increase in the number of receiving field servers. Therefore, it is difficult to apply it to field servers for rice field.

The Timing-sync Protocol for Sensor Networks (TPSN) has been proposed as a method for improving RBS [12]. The TPSN forms a tree structure with the parent node at the top, and the parent node performs time synchronization with the field server. The field server synchronizes time with the field servers below it. In this way, time is synchronized throughout the system. However, because this method requires a long time for time synchronization, it is not suitable for field servers that require low power consumption.

We proposed a wireless communication protocol that achieves low cost [4]. Figure 4 shows a basic communication protocol sequence diagram of the proposed protocol. Communication is performed according to scheduling. In Fig. 4, field server A (FS-A) sends the sensor data to field server B. Then, field server B sends its own sensor data and the sensor data of the field server A to field server C. Field server C constructs a frame with the sensor data of field servers A and B and its own sensor data and sends these data to field server D. The time is transmitted as a signal indicating that data have been received. The time can be synchronized among field servers transmitting and receiving data, but it cannot be synchronized with other field servers. The field server does not include a crystal or other system to measure the exact time, so there is an error of about a dozen seconds between the field servers in the proposed wireless communication

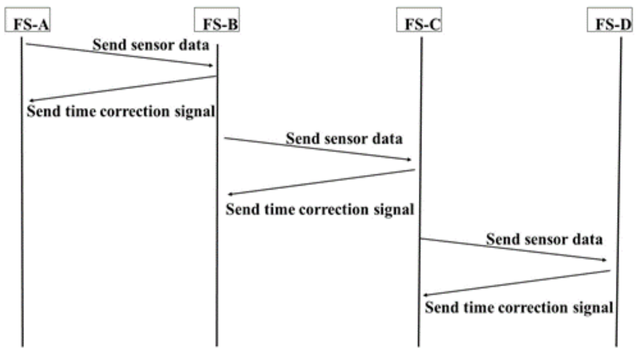


Figure 4: Communication protocol sequence diagram

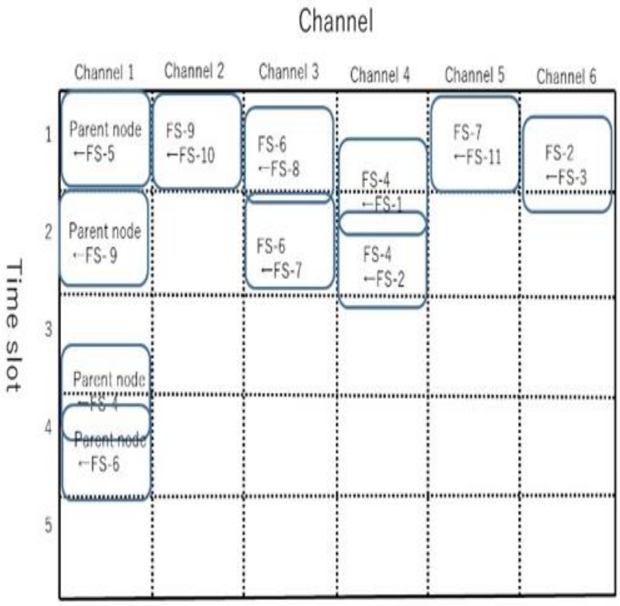


Figure 5: Time slot shift due to time error

protocol. Transmission / reception between field servers takes place within the time slot. The time slot is set to long, when a transmission or a reception fails, a retry is made after a random time.

The operation when a field server is newly introduced follows. The newly added field server first sends a wake-up signal to inform the parent node that it has been added. When the parent node receives the wake-up signal from the field server, the field server number is added to the library and rescheduling is performed. The scheduling result is sent to the field server at the same time. The field server synchronizes the time based on the reception time from the parent node. When a field server is newly introduced, the field server works by direct method at first.

The time at which the field server is installed is arbitrary. Therefore, when another field server is communicating with the parent node, the newly added field server may start communication. However, in this case, collisions may occur. In the event of collision, retransmissions are performed after a random period within a few seconds. This process is repeated until the field server receives the current time information from the parent node.

The proposed method allows time errors. Therefore, as shown in Fig. 5, the original time is deviated and a time slot is generated. It may also occur when time slots overlap. We decided to allow them by setting the time slot to about 7 or 8 times the time for a single transmission. Due to the time error, the transmission times slot executed in each channel are not unified.

4 SYSTEM CONFIGURATION OF RICE CULTIVATION MANAGEMENT SYSTEM

In the LPWA network, the time is often overlapped for the gathering sensor data and for sending them to the parent node that is the gateway. As a result, the user may not be able to confirm the data at a desired time. To solve this problem, a method that the gathering data in the shortest time is needed to fully extract the capacity of the network and the provide data at the time desired by the users [14]-[17].

Table 2: Frame format (Byte length)

Destination	Sender	Payload
1	1	variable

Table 3: Payload for sensor data (Byte length)

FS ID	Sensor data 1	Sensor data 2	Sensor data 3	Sensor data 4	sensor data 5
2	2	2	2	2	2

A rice cultivation management system monitors the air and water temperature of rice fields. This system consists of a field subsystem, a home subsystem, and a management subsystem. The field subsystem is installed in the field, acquires air and water temperature data and transmits the acquired data to the home subsystem via a wireless network. The home subsystem is a system that uploads the data sent from the field subsystem to the management subsystem via the Internet. The management subsystem is a system that analyzes and displays the collected data and allows farmers to check the data accumulated on the cloud via the Internet.

In the field server subsystem, a plurality of field servers are held; each field server acquires data once per hour, and transmits the acquired data to the home subsystem which is a parent node. This single process is called round one. Table 2 shows the frame format used for transmission. The frame is composed of a transmission destination, a transmission source, and a payload. Table 3 shows the sensor data transmission format stored in the payload of Table 2. In addition, the parent node can communicate with at most one field server at the same time, and the field server can communicate with at most one field server or parent node. In a single process, the power is turned off after system startup and after data acquisition, acquisition of a data transmission, and transmission are completed. The time required for this series of processes is defined as one round of operation time. This operation time is the time required to gather data to the parent node. In each field server, the data merging processing is performed as necessary. By this merging process, it is possible to shorten the time required for transmission.

Synchronization between the parent node and field servers is carried out by using the current time transmitted according to the transmission request sent from the parent node [18]. When the field server receives the current time, the time is updated to the current time.

5 LOW DELAY DATA GATHERING METHOD

Currently commercially available LPWA uses the direct method. However, since this method has a long data-gathering-time problem. When the number of field servers is n, and the time step T of direct method which is required can be expressed by the following equation.

$$T = n \tag{1}$$

In the direct method, the time required to transmit the data of all field servers to the parent node increases in proportion to the number of field servers. If one time slot is set to 30 seconds and there are 200 field servers, a time of 30 * 200 = 6000 seconds = 100 minutes is required, and a request of 5 minutes or less cannot be satisfied.

In this paper we propose a method that has a low power consumption and a short data-gathering time. The procedure of the proposed method is shown below.

① After the rice planting, the field servers are installed in the rice field. The field server starts in installation mode.
② Farmer sets the initialization mode to the parent node, and then the field server position confirmation procedure is executed in order to collect position information of the field server.
③ A transmission network graph is created on the parent node based on the field server position information, and the scheduling result is sent to the field servers.
④ Data gathering procedure on the local wireless network is performed based on scheduling result.

The newly added field server first sends a wake-up signal to inform the parent node that it has been added. When the parent node receives the wake-up signal from the field server, the field server number is added to the database and the time for transmitting data to the parent node is sent to the field server. Operation using direct method is started [19].

In this state, when initialization mode is set in the parent node, the field server position confirmation phase starts at the timing when data is sent from the next field server to the parent node.

The field server sends data to the parent node by direct method, then enters the reception mode. Next the field server receives data from other field servers, calculates the radio field strength, and creates a list of field server numbers in order of strength. At the next transmission timing, a list of field server numbers in order of strength is transmitted together with sensor data.

The parent node has a list of field server numbers in order of strength and a list of field server numbers in order of strength sent from each field server. Based on this information, a transmission network graph is created and the scheduling result is transmitted to the field servers by broadcasting. Each field server is turned off when it receives the scheduling result. Thereafter, each field server performs transmission processing with this scheduling result.

6 TRANSMISSION NETWORK GRAPH

We describe the procedure for creating a transfer network graph. A transmission network graph is a graph that describes the order, direction, and transmission timing of data transmission between a field server and a field server, or between a parent node and a field server. Data collection is performed according to this graph. The transfer network is defined by the effective graph G = (V, E). V is a set of the parent node and the field servers. E is a set of edge, and edges e = {i, j} indicate that data is transmitted from node i to node j. E_m is set of edges which data is sent in m steps.

Here, in order to simplify the following discussion, the symbols are defined as follows.

V: {N1, N2} node set
N1: set of parent node
N2: set of field servers
n_i: parent node
n_j: field server j where j=1, 2, 3, ...
l: the number of parent node
m: the number of field server
e={n_j, n_i}:transmission from n_i to node n_j
E_m: set of edges in the step m
E={E_1, E_2, … }: set of edges
F: set of transmission completion nodes
E: set of un-transmission completion nodes
Nearest(j, G): The node with the closest distance to node j in node set G for which transmission has not been completed
Farthest(j, G): The node with the farthest distance to node j in node set G for which transmission has not been completed
t: step number
k: tradeoff variable of processing completion time and power consumption

The transfer network graph creation flow is shown below.

```
Procedure Routing(V, E)
F={n_0};
G=N2;
t=0 ;
k=4;   /* Processing time and power trade-off factor*/
While( There are k or more nodes belonging to G) {

    /* Transmission to the parent node*/
    t=t+1;
    i= Nearest(n_0, G);
   /* Find the nearest field server to the parent. */
    Create branch {n_0, n_i} and put in set E_t representing processing at t time step
    Delete node i from un-transmit set G.
    Put the node i in the transmission complete set F.

    /* Data transmission between field servers */
    H=G;
    While(There is a node that belongs to H) {
            j=Farthest(n_0, H);
            Delate j from H.
            i=Nearest(j, T).
            Delate i from H.
            Generate edges {i, j} and put them in the set E_t representing the processing at t time steps
```

```
            Delete node j from un-transmit set G.
            Put the node j in the transmission complete
set F.
    }
}
While(There is a node that belongs to G) {
    /* Transmission to the parent node */
    t=t+1;
    i= Nearest(n_0, G);
  /* Find the closest field server to the parent node. */
    Generate edges {n_i, j} and put them in the set E_t
representing the processing at t time steps.
    Delete node j from un-transmit set G.
    Put the node j in the transmission complete set F.
}
```

The movement of the flow is explained using the example of one parent node and 11 field servers shown in Fig. 6. In the first step, the field servers 5 nearest to the parent node is selected, and an edge representing transmission to the parent node is made. Next, the field server 11 which is farthest from the parent node is selected, and an edge from the field server 11 to the field server 7 is made. Similarly, an edge from field server 3 to field server 2, an edge from field server 10 to field server 9, an edge from field server 8 to field server 6, an edge from field server 1 to field server 4 are created, and the first step is complete.

In the second step, as in the first step, an edge from the parent node to the nearest field server 9 to the parent node is created, an edge from the field server 7 to the field server 6, an edge from the field server 2 to the field server 4 is created. In the third step, an edge from the field server 4 to the parent node, and in the fourth step, an edge from the field server 6 to the parent node are created.

7 DATA GATHERING PROCEDURE

The data gathering procedure will be described. Data is gathered to the parent node according to the tree structure of the graph created in the transfer network graph creation flow. Here, the operation is explained using the example in Fig. 6.

In the first step, transmission corresponding to the edges belonging to the branch set E_1 representing the processing in the first step is performed. That is, transmissions from the field server 5 to the parent node, from the field server 11 to the field server 7, from the field server 3 to the field server 2, from the field server 10 to the field server 9, from the field server 8 to the field server 6, and from the field server 1 to the field server 4 are performed. The field server receiving the transmission merges the data held by itself and the transmitted data, creates a frame, and prepares for the next transmission. In the second step, transmission corresponding to the edge belonging to the edge set E_2representing the processing in the second step is performed. That is, transmissions from the field server 9 to the parent node, from the field server 7 to the field server 6, and from the field server 2 to the field server 4 are performed. The field server receiving the transmission merges the data held by itself and the transmitted data, and creates a frame. In the third step, transmission corresponding to the edge belonging to the branch set E_3 that is transmission from the field server 4 to the parent node is performed. In the fourth step, transmission corresponding to the edge belonging to the branch set E_4 that is transmission from the field server 6 to the parent node is performed.

Figure 7 shows a sequence diagram of data exchange between the parent node and field servers. The data of each field servers is merged step by step, and all data is sent to the parent node in the fourth step.

The number of field servers in operation at step T is defined as N (T). The number of steps required to gather data can be expressed by the following equation.

$$N(T+1)=\lceil (N(T)-1)/2 \rceil$$

Namely, T where N (T + 1) = 0 is a necessary step to transmit all data to the parent node. For example, in the example of Fig. 6,

$$N(0)=11$$
$$N(1)=\lceil (11-1)/2 \rceil=5$$
$$N(2)=\lceil (5-1)/2 \rceil=2$$
$$N(3)=\lceil (2-1)/2 \rceil=1$$
$$N(4)=\lceil (1-1)/2 \rceil=0$$

In this example, since the number of field servers is 11, N (0) = 11.Then, all data can be transmit to the parent node in 4 steps.

From the above consideration, the number of steps T required for data gathering can be expressed by the following equation.

$$T = \lceil log_2 n \rceil \quad (2)$$

If one time slot is set to 30 seconds and there are 200 field servers, a time of 30 * $\lceil log_2\ 200 \rceil$ = 228 seconds = 3.8 minutes is required, and a request of 5 minutes or less can be satisfied.

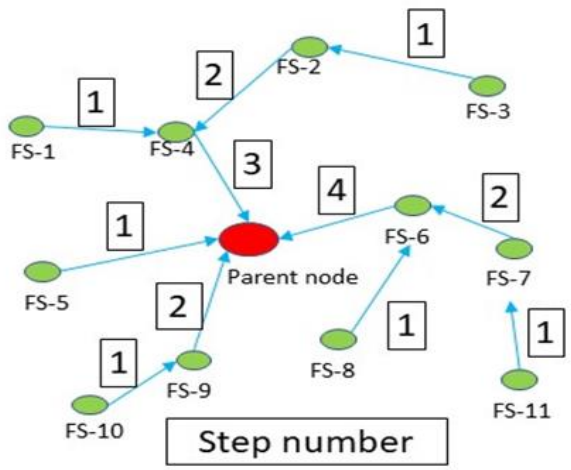


Figure 6: Operation example of one round

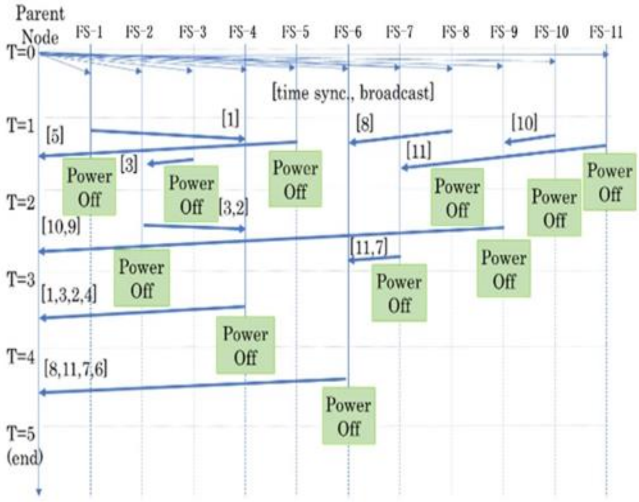


Figure 7: Sequence diagram of operation example

8 PERFORMANCE EVALUATION BY SIMULATION

A comparison was made by simulation to check the data gathering time. Here, we focus on the prompt confirmation of the condition of the rice field required for rice cultivation. Therefore, we verify the data gathering time and the power consumption required to collect all the field server data. Being able to verify the data within an allowable time is an important item even if the number of field servers is increased. The simulator was written in C language. First, the method of calculating the power consumption will be described. Owing to the fact that the parent node is always supplied with power, its power consumption is excluded from the scope of consideration and only field servers are calculated. Each field server has a battery and has a capacity of 75,000 mWh. The field server has six modes (Table 4) according to the operating condition. The current consumption was set from the actual power consumption measurement report [18] using LoRa. The transmission power depends on the distance. Operation voltage is set at 2.5 V. The time of data merge is not considered because it is 1/100 or less of the time required for transmission and reception. Also, since the data transfer network is created only once at first, the influence on the power consumption is not considered because it is small.

Table 4: Six mode of field server

State	State name	Power consumption (mA/s)	Processing time (s)
1	Power OFF	0	-
2	Data transmission mode	53, 62, 69, 78	3.4
3	Data reception mode	13.5	3.4
4	System startup mode	Parent 78 FS 13.5	3.4
5	Sensor data acquisition mode	60.0	60.0
6	Standby mode	2.7	3.4

Table 5: Details of simulation data

	#Parent	#FS	Radius(m)
DATA1	1	99	500
DATA2	1	200	500
DATA3	1	300	3000

The sensor data to be acquired is the data on temperature and humidity. In each round, when each field server reaches the activation time, the system is started and the data is acquired from the sensor. Next, the parent node transmits a data transmission request to all of the field servers. Upon receiving the data transmission request, the field server measures the data from the sensor, and transmits the data to the parent node. Each field server turns off the power source. In this simulation, it is assumed that the time synchronization is perfectly performed. It was assumed that there was no failure of data transmission / reception. In fact, the merging time of the data by the field server was ignored in this simulation. Given that the payload can be transmitted at a data rate of up to 60 bytes, the transmission time can be shortened by the data merging process.

Table 5 shows the three instances of test data created. The field servers have been randomly assigned within the circle of a specified radius, centered on the parent machine.

Comparisons were made between the following methods: direct [20], pegasis [21], epegasis [22], chiron [23] and the proposed method. In addition, the following three items are compared to enable the farmer to confirm the data of the rice field in real time, while reducing the data gathering time and operating with batteries from the time of rice planting to harvesting.

Power Efficient Gathering in Sensor Information Systems (pegasis method) is a chain-based protocol [21]. The field servers are connected only to the closest field server and are chained. Data is sent along the chain to the chain head (the last filed server of the chain) and from the chain head to the parent node. The chain head is randomly selected and has a duty to transmit data to the parent node. The time required to gather data for pegasis method is long.

Epegasis method [22] has been proposed as a method to solve the pegasis method problem. The field servers are broken down into clusters at a distance from the parent node, and a chain is created in each cluster in the same way as pegasis method. Each cluster sends data to the chain head along the chain. When data gathers in the chain head in each cluster, the chain head transfers data to the parent node. Therefore, the time required for data collection can be reduced compared to pegasis method.

Chiron method [23] has been proposed as a method for achieving further improvement. It is characterized by using Beam Star technology [24] to divide into smaller cluster areas than epegasis method. In each cluster, build a chain. Initially, the data for each cluster is collected by the chain head in the same way as pegasis method. Next, a chain between chain heads is created in order from the chain head farthest from the parent node, along which data is sent to the parent node. Chiron method can shorten the time required for data collection in each chain by shortening the chain length, but has the problem that the time required for gathering data in each area to the parent node increases.

i. Time required for data gathering in one round.
ii. Total power consumption of all field servers in one round.
iii. Maximum working days (the number of days during which all the field servers are in operation).

The simulation results of DATA 1 are shown in Table 6. Table 6 shows the operation time, which is the time required for the parent node to gather data, the total value of the power consumed by the field server, and the number of operable days. Figure 8 shows the size of data received by the parent node in the one round. The vertical axis in Fig. 8 represents the data size (byte) received by the parent node, and the horizontal axis represents operating time (seconds). Figure 9 shows the number of field server whose transmission processing has been completed and the power has been turned off within one round. The horizontal axis shows the operating time (seconds), and the vertical axis shows the number of field servers whose power has been turned off. It can be confirmed that the slope of pegasis method is the slowest.

Table 6: Simulation results of data1

	Data collection time / Round (s)	**Total power consumption of field servers /Round (mAS)**	**Number of working days (Day)**
Direct	400.0	74508.5	372
PEGASIS	1902.9	299237.6	58
EPEGASIS	226.6	55253.0	384
CHIRON	158.6	40131.1	448
Proposed	**114.4**	**34177.0**	**566**

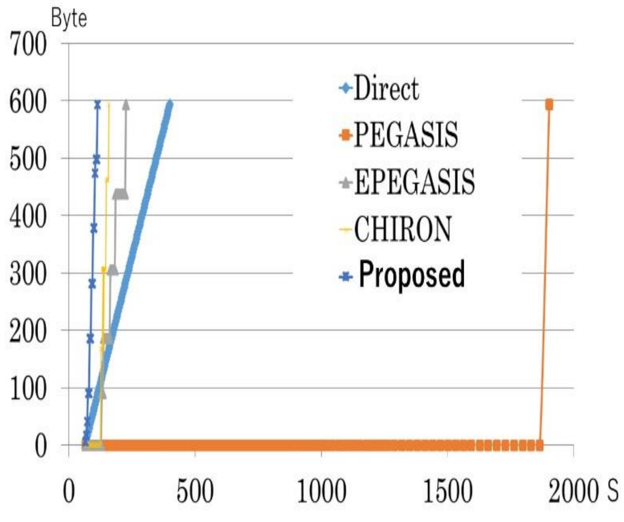


Figure 8: Number of data received by parent node of DATA1.

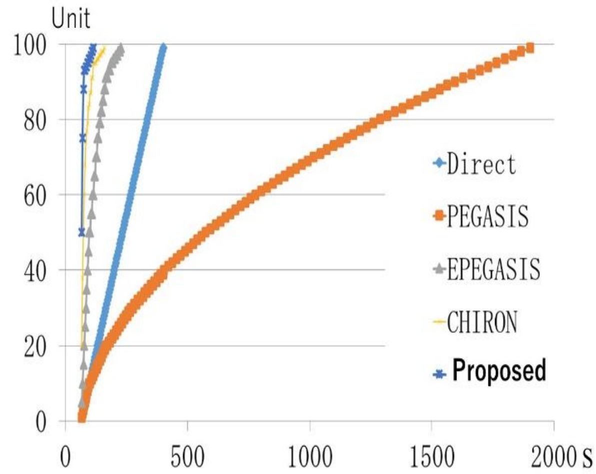


Figure 9: Number of power off field servers of DATA1.

Compared to other methods, the proposed method shows that the time to gather data to the parent node is the shortest. Moreover, the total power consumption is the smallest, based on the simulation, it can be concluded that it can operate for a period that extends from rice planting to harvesting.

The simulation results for DATA 2 are shown in Table 7. Compared to other methods, the proposed method shows that the time for the parent node to collect data is the shortest. In addition, it is understood that the total power consumption is the smallest, and the lifetime is the longest. In the simulation, we can see that it can operate for a period that extends from rice planting to harvesting. In addition, although the proposed method doubles the number of field servers as compared with DATA 1, it can be seen that the increase in time required for data gathering is as short as about 7 s.

The simulation results for DATA 3 are shown in Table 8. Compared to other methods, the proposed method shows that the time to for the parent node to collect data is the shortest. In addition, it is understood that the total power consumption is the smallest, and the lifetime is the longest. Moreover, in terms of the theoretical value, it can be understood that it can operate for a period extending from rice planting to harvesting. In addition, compared to DATA 1, the number of field servers is tripled, but the increase of time increase necessary for collection is short as about 10 s.

Table 7: Simulation results of data2

	Data collection time / Round (s)	**Total power consumption of field servers /Round (mAS)**	**Number of working days (Day)**
Direct	743.4	243604.6	210
PEGASIS	7135.4	1763010.2	18
EPEGASIS	721.0	201747.2	195
CHIRON	248.7	90308.8	209
Proposed	**121.2**	**70925.4**	**231**

Table 8: Simulation results of data3

	Data collection time / Round (s)	Total power consumption of field servers /Round (mAS)	Number of working days (Day)
Direct	1083.4	519160.0	67
PEGASIS	15771.0	5672543.4	8
EPEGASIS	695.8	487720.8	97
CHIRON	322.1	168756.0	151
Proposed	**124.6**	**117871.6**	**185**

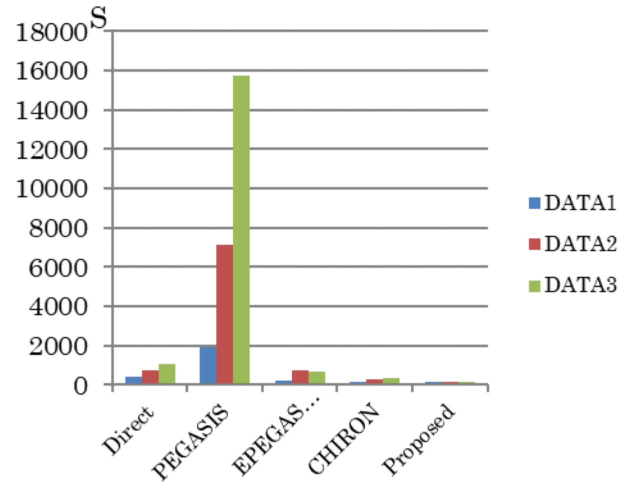


Figure 10: Comparison of time required to gather data in one round.

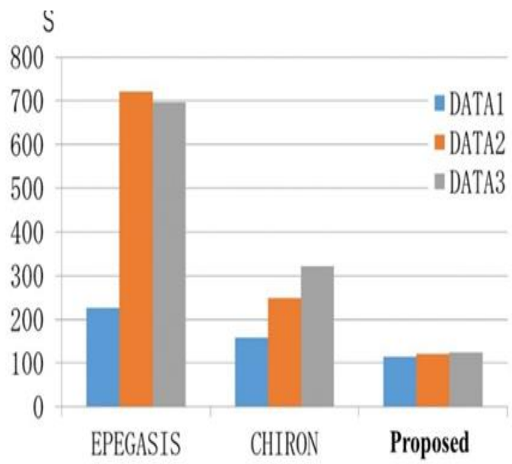


Figure 11: Excerpt from the Fig. 10

Figure 10 displays the time to gather data to the parent node in one round. Figure 11 is excerpt from the Fig. 10. The vertical axis represents the operation time (seconds). For the direct and pegasis methods, as the amount data increases, the time required to gather data to the parent node rapidly increases. For the proposed method, the rate of increase of the time required for data gathering according to an increase in the amount of data is the lowest, and data can be collected in a short time even if the number of field servers increases. In addition, the elongation rate is about a 5% growth with respect to an increase in the number of field servers. This result confirms the validity of Equation (2) and also indicates that the proposed method is superior to other methods in terms of the data gathering time.

The proposed method gather all field server data in a short time. Since all data can be gathered in a short time, the operation time is short and the power consumption can be reduced. In order to gather the data in a short time, three measures were taken to reduce the number of the data transmission to the parent node. The first is compression by the data merging to reduce the number of the transmissions. In order to reduce the number of the transmissions, the upper field server merges the data from multiple lower field servers. The power consumed in the merge process is negligible compared to the transmission. The second is the scheduling method in which the parent node keeps receiving data without idle time. The third is the transfer network graph configuration that minimizes the number of multihops. This has reduced the total amount of data that must be sent.

The proposed method is for IoT, and it is the method for a small data size such as the environmental data of temperature and humidity, or the consumption data of electricity and gas. It is also assumed that the size of the data to be handled is almost the same. Therefore, when the data size is large, the compression effect is reduced and the efficiency may be reduced. And when the data size is significantly different, latency may occur and efficiency may be reduced.

9 CONCLUSION

The telecommunications carrier adopts a direct method, but the direct method is not suitable for a usage method in which data are transmitted from IoT nodes at the same time. To solve this problem, a method that gathers data in the shortest time is proposed to attain the capacity of the network fully and provide data at the time desired by the users.

In this paper, we assume an application that acquires and transmits data at an interval of tens of minutes, particularly a system for rice cultivation. In addition, a wireless network that connects all rice fields locally is assumed. The wireless network is of the same type as a wired LAN + Wi-Fi network in a company. This method is effective in terms of security and cost. We proposed a communication protocol to realize real-time performance with this wireless network.

It is difficult to supply power to rice fields unlike other fields, so low power consumption is an important factor. In addition, the line usage fee required for communication must be reduced as much as possible from the standpoint of the entire cost required for agriculture. Therefore, we investigated interfiled communication using LoRa which is an IoT communication standard, low power consumption using the ISM band and a long transmission distance.

As a result of the comparison with other methods using the simulation, we confirmed that the time required to gather field server data to the parent node is the shortest. Moreover, it was confirmed that the power consumption is smaller than that of conventional systems. Furthermore, we confirmed that the elongation rate of time necessary for gathering data to the parent node due to an increase in the number of field servers is lower than that of the other methods. This result supports the theoretical formula.

Therefore, this proposal makes it possible to transmit the situation of a rice field in a timely manner and can be considered to be a useful method for reducing the burden of agricultural work. In addition, it is effective for other applications that require real-time performance.

REFERENCES

[1] Factory Management Editorial Department, “SMEs begin! Production site IoT,” Nikkan Kogyo Shimbun.

[2] “Use LPWA to understand LP gas usage. Large-scale demonstration of participation by thousands of households,” https://www.itmedia.co.jp/smartjapan/articles/1805/11/news029.html, (Access 2019/10/21).

[3] “Practical use of automatic meter reading with new technology "LPWA" -Metering solution for water supply using "SIGFOX" -,” http://news.kddi.com/kddi/corporate/newsrelease/2017/03/30/2379.html, (Access 2019/10/22).

[4] K.Tanaka, M.Sode, T.Ozaki, M.Nishigaki and T.Mizuno: “A Study on Time Synchronization Method for Field Servers for Rice Cultivation,” International Journal of Informatics Society (IJIS 2019), Vol. 11, No. 1, pp.33-pp.44, (2019/6/19).

[5] T. Hagino, “Rice farming in Japan,” https://www.fukoku-life.co.jp/economy/report /download/analyst-VOL215.pdf, (Access 2019/10/22).

[6] RiceWatch, https://field-server.jp/ricewatch/rental/index.html, (Access 2018/10/2).

[7] Sigfox, https://www.sigfox.com/en, (Access 2018/10/2).

[8] Cabinet office, “Changes in management scale per farmhouse,” https://www.cao.go.jp/sasshin/kisei-seido/meeting/2011/agri/120712/item2-2_2.pdf, (Access 2019/10/22).

[9] T. Watanabe, T. Morito, M. Minami, and H. Morikawa, “Design and Implementation of a Radio Controlled Clock based Battery-less Wireless Sensor Network,” IPSJ SIG Mobile Computing and Ubiquitous Communications (SIG-MBL), Vol. 2007, No.44, pp.113-118, (2007).

[10] Position Co., Ltd., “About GPS reception as a time reference,” http://www.posit.co.jp/mission/pdf/1998.6.pdf, (Access 2017/5/3) .

[11] J. Elson, L. Girod, and D. Estrin, “Fine-Grained Network Time Synchronization using Reference Broadcasts,” Proceedings of the 5th Symposium on Operating Systems Design and Implementation (OSDI’ 02), Boston, Massachusetts, (2002).

[12] S. Ganeriwal, R. Kumar, and M. B. Srivastava, “Timing-sync Protocol for Sensor Networks,” Proceedings of the 1st ACM Conference on Embedded Network Sensor Systems (SenSys'03), Los Angeles, California (2003).

[13] T. Furuta, H. Nakagawa, T. Kitamura, Y. Kawakami, K. Kurosawa, K. Kogami, and M. S. Tanaka, “Agricultural support system equipped with short-range wireless communication system,” 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) , (2016/07/15-16).

[14] Ministry of Agriculture “Statistics on agricultural labor force”, http://www.maff.go.jp/j/tokei/sihyo/DATA/08.html, (2016/11/02 reference).

[15] K. Kurosawa, I. Iizima, Y. Amemiya, S. Yamamichi, M. Toyota, and M. S. Tanaka, "Development of operational control system for rice cultivation," IEICE-CS 2016-12-IE-CS-AVM-BCT, (2016/12/8-9).

[16] Vegetalia, RiceWatch, http://field-server.jp/ricewatch/, (2016/03/02 reference).

[17] PS Solutions, e-kakashi, https://www.e-kakashi.com/, (2016/03/02 reference).

[18] K. Terada, M. Toyota, T. Hirata, Y. Takada, K. Matumoto, and M. S. Tanaka, “Proposal of communication protocol for field management using LoRa,” Information Processing Society DICOMO2017 Symposium, pp. 1671-pp. 1678, (2017/6/28-30).

[19] M. S. Tanaka, “Data gathering method of all sensor node in shortest time for LPWA system,” International Workshop on Informatics, (2019/9/8-11).

[20] S. Rani, and S. H. Ahmed, “Multi-hop Routing in Wireless Sensor networks,” Springer.

[21] S. Lindsey, and C. Raghavendra, “Pegasis: power-efficient gathering in sensor information systems,” Proceedings of the IEEE Aerospace conference proceedings 2002, pp.3-1125 –pp.3-1130, vol.3. IEEE. 10.1109/AERO.2002.1035242.

[22] S. Jung, Y. Han, and T. Chung, “The concentric clustering scheme for efficient energy consumption in the pegasis,” Proceedings of the 9th international conference on advanced communication technology, pp.260-pp.265, Gangwon-Do: IEEE. 10.1109/ICACT.2007.358351.

[23] K. Chen, J. Huang, and C. Hsiao, “Chiron: an energy-efficient chain-based hierarchical routing protocol in wireless sensor networks,” Proceedings of the Wireless Telecommunications Symposium (Wts'2009), pp.183-pp.187, USA: IEEE.

[24] S. Mao, and T. Hou, “Bamstar: an edge-based approach to routing in wireless sensor networks,” IEEE Transactions on Mobile Computing, Vol. 6, Issue 11, Nov. 2007, pp.1284-pp.1296.

(Received November 02, 2019)
(Revised March 08, 2020)

Mikiko Sode Tanaka received Dr. Eng. degrees from Waseda University in Fundamental Science and Engineering. She joined NEC Corporation, NEC Electronics Corporation, and Renesas Electronics Corporation. She is Associate Professor of International College of Technology, Kanazawa. Her research interests include wireless communications, AI chip, and personal authentication. She is a member of IPSJ (Information Processing Society of Japan), IEICE (Institute of Electronics, Information and Communication Engineers) and IEEE (Institute of Electrical and Electronics Engineers).

Regular Paper

CG Training Model Application Method Using Cycle-consistent Adversarial Network

Tsukasa Kudo[†]

[†]Faculty of Informatics, Shizuoka Institute of Science and Technology, Japan
kudo.tsukasa@sist.ac.jp

Abstract - Deep learning-based image recognitions have achieved high accuracy; however, its application to many actual business systems remains difficult. First, deep learning requires extensive training data, which are often difficult to obtain for particular businesses. Second, actual business environmental scenes change owing to various factors, such as the time of day, season, and weather. Consequently, typically, there is a domain gap between the training and target images, which often reduces recognition accuracy. To address these problems, a method to automatically generate training data that is not affected by scene changes using computer graphics (CG) has been proposed. However, domain gap problem between the CG generated images and actual images remains. Recently, Cycle-Consistent Adversarial Networks (Cycle-GAN) have been proposed to translate an image from one domain to a fake image of another domain. In this study, a method to use a model trained with CG images is proposed. In this method, actual images are translated into fake CG images using a Cycle-GAN. An experimental evaluation demonstrates improved accuracy; the proposed method is applied to an inventory estimation of parts in a bulk container using deep learning regression model.

Keywords: deep learning, GAN, Cycle-GAN, regression model, stock-taking, computer graphics, CG, image recognition

1 INTRODUCTION

Currently, deep learning has been applied effectively to various fields [6], [4]. Its effectiveness has been demonstrated in the ImageNet Large Scale Visual Recognition Challenge [12]. And, image recognition accuracy has improved rapidly [19]. With improved accuracy, deep learning has been used with a wider range of image processing applications, such as face recognition, medical image analysis, and plant disease detection [20], [8], [17]. Additionally, an increasing number of images have been stored on cloud servers; therefore, obtaining sufficient images for training data has become easier. However, obtaining sufficient images as training data for individual businesses remains difficult and time-consuming.

In previous studies, I have investigated stock-taking in a machine factory. Typically, in a machine factory, most parts are stored in bulk containers, and for stock-taking, counting the parts manually through a superficial visual examination is impossible. Thus, to count the parts, they must be removed from the container, which needs heavy work-load. To address this problem, I investigated applying deep learning to inventory estimation using image recognition and confirmed that practical accuracy could be achieved using a deep learning regression model [14]. This study was performed in a laboratory where images of lightweight marbles and nuts were used for the training and test data; 1,600 original images were increased to 8,000 by padding.

However, applying this method to an actual factory was difficult. First, to capture images for training data, heavy parts must be repositioned manually. Additionally, typically, there are thousands of bulk containers in a factory, and capturing images of the parts in each bulk container is not practical. Second, since the scene in an actual factory changes due to various factors, such as the time of day, season, and weather, there may be a domain gap between the training and target images. Such domain gaps can reduce recognition accuracy [21].

To address the first problem, we have previously developed a method to generate sufficient training data automatically using computer graphics (CG) [13]. In this study, I propose an inventory estimation system that uses CG generated training data and target data created from images of actual parts. To eliminate the domain gap between the actual and CG images, the actual images are translated to fake CG images before the inventory estimation. Thus, it is expected that inventory estimation accuracy can be maintained regardless of the change of scenes.

Here, this translation is performed using a Cycle-Consistent Adversarial Network (Cycle-GAN) [24], which is a type of generative adversarial network (GAN) [6]. A Cycle-GAN model is trained using different image groups, such as zebras and horses, where images of one group (zebras) are automatically translated into fake images of another group (horses). For some types of objects, it has been shown that images can be translated into highly accurate fake images using this trained model.

In this study, the objective is to verify the feasibility of using a Cycle-GAN to improve the estimation accuracy of the above-mentioned inventory estimation system [13]. It comprises a regression model trained with CG images, and the target data of actual images.

The remainder of this paper is organized as follows. Section 2 reviews related works, the author's previous studies, and states the primary goal of this study. In Section 3, a method to estimate inventory using a training data generator and Cycle-GAN is described. The effectiveness of using fake CG images translated by a Cycle-GAN is evaluated in

Figure 1: Parts inventory in bulk containers

Section 4. Evaluation results are discussed in Section 5, and conclusions and suggestions for future work are presented in Section 6.

2 RELATED WORK

In this section, the author's previous studies and related studies are described. Also, the objective of the current study is discussed. The parts inventory in a machine factory is managed based on a theoretical inventory planned using a production management system. However, actual and theoretical inventories often differ due to product defects, work errors, work delays and so on. Therefore, manual stock-taking must be undertaken. However, manual stock-taking needs heavy workload because, typically, parts are stored in bulk containers (Fig. 1) and to obtain an accurate count of individual parts, the parts must be removed from containers.

On the other hand, in recent years, with the progress of deep learning, the accuracy of image recognition has improved and image recognition has been applied to various fields [8], [12], [17], [19], [20]. Therefore, my colleague and I developed and evaluated a method to estimate inventories using images taken from the outside of the bulk containers. First, we evaluated results obtained using a deep learning multi-class classification model and found that it could achieve a certain estimation accuracy.

Also, we demonstrated how this method could be applied to actual inventory management [10]. The method produced some estimation errors; however, we found that the number of out-of-stock parts could be reduced by increasing the safety stock according to the error range. Also, accuracy could be improved by comparing the estimated inventory with the theoretical inventory. Note that in the multi-class classification model, inventory quantities are treated as discrete classes, and an inventory estimated from a bulk container image is considered a discrete class (quantities) [4].

However, preparing training data for that study was extremely time-consuming. Lightweight marbles and nuts were used to represent parts; 1,600 different original images were captured by repeatedly rearranging these items manually. The 1,600 images were padded to create 8,000 images; 6,000 were used as training data, and 2,000 were used as test data. However, in an actual factory, using this process to obtain training and test data is not practical because typical parts are extremely heavy and there are thousands of bulk containers.

Therefore, I developed a training data generator and determined that a model trained using CG generated data could achieve a certain accuracy [13]. It became evident that based on a machine drawing of the part, the shape of the CG model of the part could also be created relatively easily. In contrast, given various textures and illumination environments, a trial and error process was required to obtain the color tone of parts. Note that, compared to the types of parts, the number of textures was small because the parts were made using limited types of materials, such as iron and aluminum.

In the multi-class classification model, training data must be prepared for each inventory quantity. That is, when inventory quantities fluctuate over a wide range, it is necessary to prepare a large amount of training data. Therefore, accuracy was evaluated using a regression model [14], which estimates an inventory as a continuous quantity [4]. As a result, compared to the multi-class classification model, high estimation accuracy was obtained when actual images were used for both the training and test data. Also, similar accuracy could be obtained for plural parts when estimation was performed with CG images using models trained with CG images.

However, when estimation was performed with actual images using the regression model trained with CG images, estimation accuracy was reduced significantly due to differences in color tone between the actual and CG images, i.e., the domain gap. Other studies have also found that image recognition accuracy deteriorates due to the domain gap between training and target images [21]. This suggests that applying deep learning to stock-taking in an actual factory will be problematic because factory conditions are not stable. Consequently, the domain gap occurs in both the training and target images.

Besides, in 2015, the automatic generation of images using DeepDream, which is an image processing method that uses deep learning, was reported. Since then, studies on automatic image generation and image style conversion by applying deep learning have been actively conducted [4], [18]. It has been shown that a GAN method can be used to generate a fake image from a genuine image. This method comprises a generator network (hereinafter, 'generator') that generates a fake image and a discriminator network (hereinafter, 'discriminator') that classifies images as fake or genuine [7]. By training both networks together, it is possible to generate a fake image that is close to the genuine image. Various methods to generate fake images based on the GAN have been proposed [15], such as pix2pix, DiscoGAN and UNIT [9], [11], [16].

Among such methods, a Cycle-GAN can be applied to translate an image from one domain into an image of another domain. For example, an image of zebras is translated into a fake image of horses by training a Cycle-GAN model with the image group of zebras and horses [24]. Figure 2 shows the structure of a Cycle-GAN model, where X and Y show different groups of images such as zebras and horses.

In Fig. 2, image x of group X is translated to a fake image $\hat{y}$ of group Y by generator G, and discriminator D_Y determines whether $\hat{y}$ is a genuine or fake image of Y. Further-

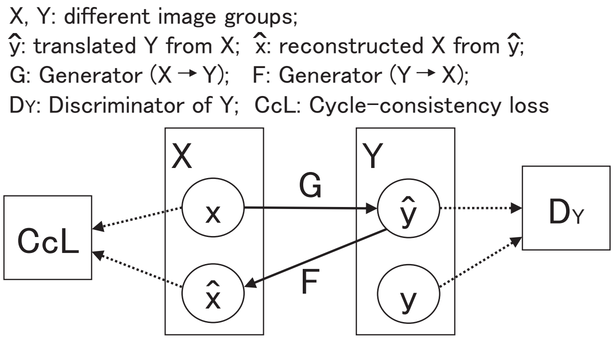


Figure 2: Structure of Cycle-GAN model

more, $\hat{x}$ is reconstructed from $\hat{y}$ by generator F, and Cycle-consistency loss, which is the loss between x and $\hat{x}$, is evaluated. In this study, the feasibility of applying Cycle-GAN to solve the domain gap problem between training and target data in regression models is evaluated.

I assume a case where the Cycle-GAN model is used for quantity estimation from images, such as stock-taking of parts in bulk containers. In other words, I attempt to show that estimation accuracy can be improved by translating actual images to fake CG images using Cycle-GAN. Here, Fig. 2 shows the case of translating X to a fake image of Y. Similarly, images of Y are also translated to fake images of X, and the discriminator and cycle-consistency loss evaluation are performed. In this way, similar to a GAN, by training the generator and discriminator together, images of one group can be translated into high-precision fake images of another group.

Also, given the structure of the Cycle-GAN model (Fig. 2), image translation can be performed automatically between unrelated image groups X and Y. Therefore, by applying the Cycle-GAN model to the above-mentioned inventory estimation, actual images of bulk containers taken under various scene conditions can be automatically translated into fake CG images generated by the training data generator. Most importantly, since CG images can be generated as in a fixed environment, it is expected that the domain gap problem can be solved.

Currently, Cycle-GAN has been applied to data translation in various fields such as videos, voices, and human images [2], [5], [21]. Also, various image translation models have been proposed using the concept of Cycle-GAN, such as Combogan, Sem-GAN, CYC-DGH, and CinCGAN [1], [3], [22], [23]. However, to the best of my knowledge, no study has applied Cycle-GAN to image recognition and evaluated accuracy improvement by eliminating the domain gap between the training and target images.

In this study, the feasibility of applying Cycle-GAN to solve the domain gap problem between training and target data in regression models is evaluated. I assume a case where a regression model trained with CG images is used for quantity estimation from actual images, such as stock-taking of parts in bulk containers. In other words, I attempt to show that estimation accuracy can be improved by translating actual images to fake CG images using Cycle-GAN.

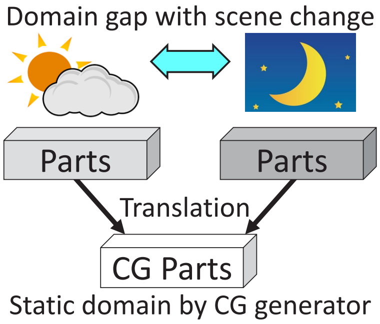


Figure 3: Domain gap in factory and generation of a static domain images

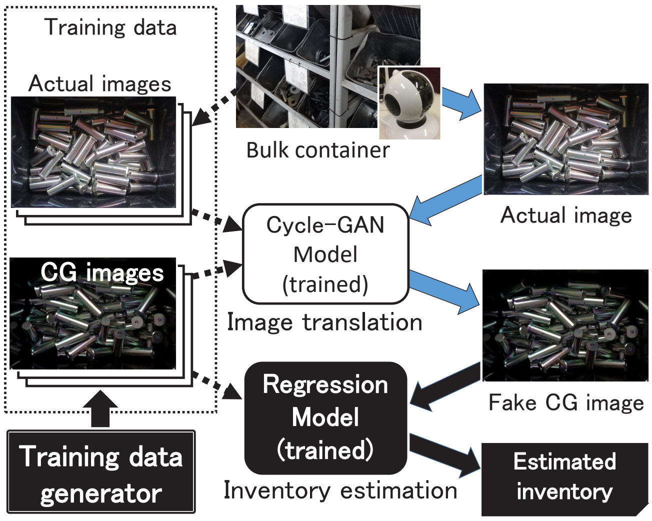


Figure 4: Inventory estimation system dataflow

3 INVENTORY ESTIMATION PROCESS USING TRAINING DATA GENERATOR

Section 2 discussed the two problems when applying deep learning to parts inventory estimation in an actual factory, i.e., preparation of a large amount of training data and domain gap among images due to scene changes. The method shown in Fig. 3 is proposed to address these issues. First, a large amount of training data ('CG parts' in Fig. 3) is generated by the training data generator using CG mentioned in Section 2 under specified conditions. Second, the target parts images ('Parts' in Fig. 3) in the factory are translated into fake images of CG parts (fake CG images). As a result, the domain gap between the training and target images is eliminated.

Figure 4 shows the data flow of the inventory estimation system, in which a Cycle-GAN is used to translate actual images into fake CG images. As shown in the dashed box in Fig. 4, training data for the Cycle-GAN model are prepared as follows. Actual images are created from pictures of factory bulk containers, and CG images are generated using the training data generator. This system uses the regression model trained with CG images, as shown in the black hatched round box shown in Fig. 4. This regression model comprises convolutional layers with pooling layers and fully connected layers, as shown in Fig. 5. Note that the mean square error (MSE) is

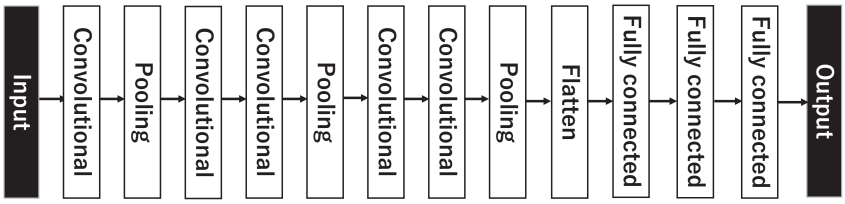


Figure 5: Structure of regression model

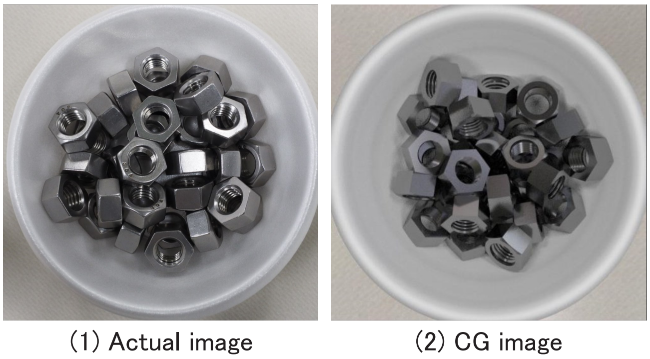


Figure 6: Target nut images

used for its loss function.

As shown in Fig. 4, the actual image of each bulk container is captured by a camera installed around the containers, and the inventory is estimated from this image. To suppress the increase in the inventory estimation error due to the domain gap between the actual and CG image, the captured image is translated into a fake CG image using the Cycle-GAN model. Then, the regression model estimates inventory quantity using this fake CG image.

For the regression model's training data, the shape of the part is important relative to maintaining estimation accuracy. Also, a single camera monitors many bulk containers; thus it is necessary to create training data images not only from just the above direction but also from the direction of the camera for each bulk container. Using the training data generator, the shape of the CG part model can be created easily based on a drawing of the machining of the target part, and the camera position of the CG can be designated at rendering based on the actual camera position. In other words, the training data generator can generate a large amount of training data automatically for the parts in each bulk container.

Therefore, the Cycle-GAN model is only used to eliminate the domain gap due to the color tone of the actual and CG images. It primarily depends on the scene change shown in Fig. 3 and part texture due to its material. Here, there are relatively few types of part materials as mentioned in Section 2, thus Cycle-GAN model training is performed for only each of these materials. In other words, for the training data, CG images are generated by the training data generator, and the actual images are created by collecting images of parts made of the same material. As discussed in Section 2, relative to the Cycle-GAN training data, it is not necessary to associate an individual actual image with an individual CG image. Thus, it is possible to accumulate actual images without investigating

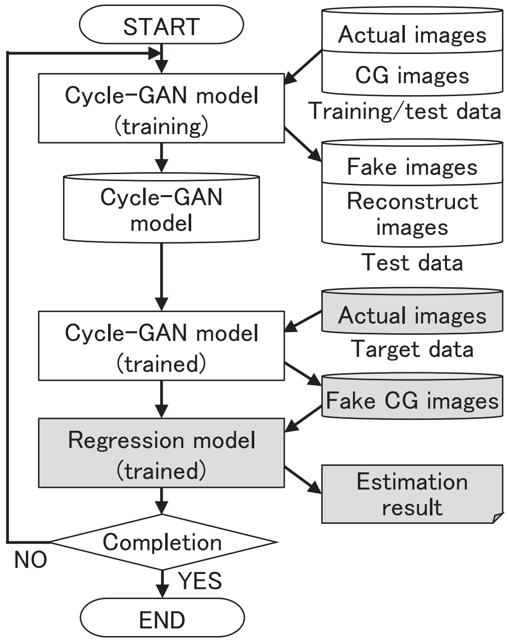


Figure 7: Structure of experimental system

part quantities.

4 EXPERIMENTS AND EVALUATIONS

4.1 Experimental Environment

Figure 4 shows the experimental environment used to evaluate the effect of translating actual images to fake CG images using the Cycle-GAN in inventory estimation. First, images of nuts placed in a bowl were used as the target parts, which were captured from above as shown in Fig. 6. For example, Fig. 6 (1) shows an actual image created from the picture of the nuts, and Fig. 6 (2) shows a CG image generated by the training data generator. Here, 100 images were prepared for each nut quantity from five to 80 for every five, for each of these (1) and (2). In total, 1,600 images were prepared and divided into training (1,200 images) and test (400 images) data.

The structure of the experimental system is shown in Fig. 7. In this experiment, the Cycle-GAN model was trained using training data comprised of actual and CG images. The training was performed as batch processing, and the number of batch per epoch was 1000. Here, discriminator loss (d_loss) and cycle-consistency loss (g_loss) were monitored every 200 batches. Simultaneously, the test data (i.e., actual and CG images) were translated into fake images. Then, each original image was reconstructed from each fake image. Figure 8

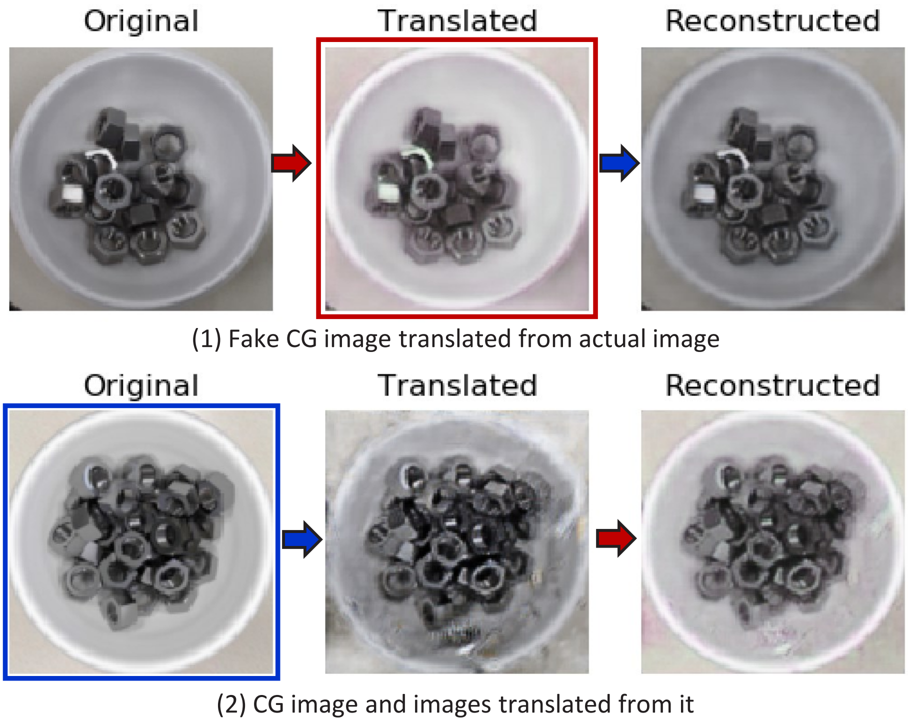


Figure 8: Images generated by Cycle-GAN model

shows examples of these data.

The Cycle-GAN model was saved, and the target data for inventory estimation, which were actual images prepared separately, were translated into fake CG images using this saved model. Here, the actual images comprised 16 types of quantity groups, similar to the training and test data for the Cycle-GAN model, i.e. five to 80 for every five. Also, 50 images were prepared for each group, i.e., 800 images in total. To evaluate the estimation error automatically, the number of nuts in each image was added as a correct label.

The experimental system was implemented on a PC with Windows 10 using Python and Keras. Note that TensorFlow was used for the Keras backend, and OpenCV was used for image conversion. The images were converted to 128×128 pixels and used for training and estimation. For the training data generator, a nut and bowl were modeled using the Blender (a 3DCG modeling tool). Here, the nuts were placed into the bowl using Blender's physical simulation function. Note, Blender's physical simulation and rendering processes were automated using Python.

To training the Cycle-GAN model, its hyper-parameters were set as follows. λ was set to 10.0, which is the strength of the cycle-consistency loss against discriminator D_Y (Fig. 2) and D_X (discriminator for generator F). The argument of the Adam optimizer was set to $Adam(0.0003, 0.5)$. Similarly, for the regression model, the reduction rate of the learning rate was 0.1, the minimum learning rate was set to 10^{-10}, dropout was not used, the output dimension of the fully connected layers was 128 (except for the last layer), and the best model in the training transition was saved. Note that the output dimension of the last layer of a regression model is one.

4.2 Evaluations

Figure 9 shows the transition of d_loss and g_loss of the Cycle-GAN model and the MSE of the inventory estimated by the regression model according to the training progress of the Cycle-GAN model. As can be seen, the MSE decreased as training progressed and became the smallest at batch number 1000 of epoch 12. Then the MSE increased. Note that d_loss fluctuated greatly before the MSE became the smallest and became a very small value several times at batch number 1000 of epoch 8, batch number 600 of epoch 5 and so on. However, there was no tendency for the MSE to improve at these times. Similarly, no clear correlation between g_loss and MSE was observed.

Figure 8 shows an example of the input and output images of the Cycle-GAN model when the MSE became the smallest value. Figure 8 (1) shows (from the left) the actual image, the fake CG image translated from the actual image, and the reconstructed image obtained using the fake CG image. Each corresponds to x, $\hat{y}$, and $\hat{x}$ in Fig. 2. Similarly, Fig. 8 (2) shows a CG image, its fake actual image, and the reconstructed CG image. Inventory estimation was performed by the regression model using the fake CG image at the center in Fig. 8 (1). The original image at the left in Fig. 8 (2) is a CG image and the translated fake CG image in (1) was generated as a fake image of this CG image.

Next, to evaluate the effect of images translated using the Cycle-GAN on inventory estimation, i.e., the effect of fake CG images, we performed a comparative evaluation using CG images, fake CG images, and actual images. Figure 10 shows the MSE results for these image types. As can be seen, using fake CG images, MSE improved by approximately 2.8 times

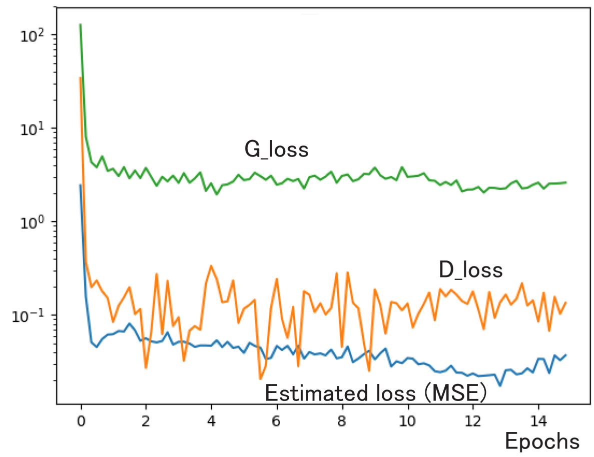


Figure 9: Transition of MSE in inventory estimation with Cycle-GAN training

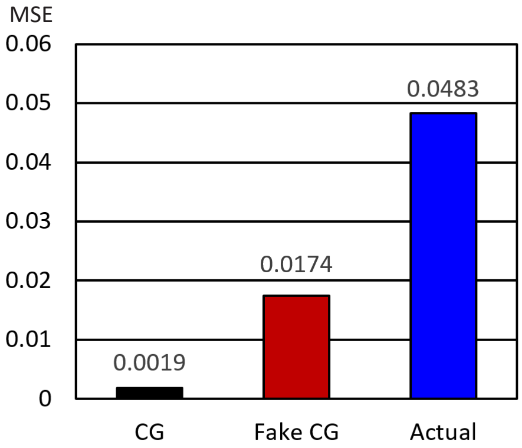


Figure 10: MSE of estimated inventory with different image types

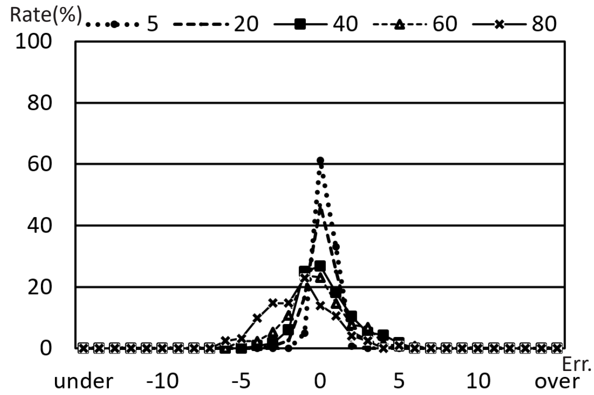


Figure 11: Histogram of error distribution with CG images

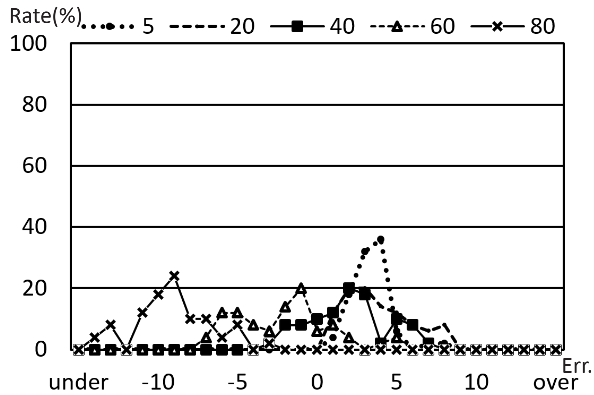


Figure 12: Histogram of error distribution with fake CG images

compared to using actual images. However, MSE deteriorated by approximately 9.2 times compared to using CG images.

Figures 11, 12 and 13 show histograms of the estimation errors for CG images, fake CG images, and actual images. In each figure, the horizontal axis shows the error; thus the position of 0 is the correct estimation quantity. Note that 'under' gives the total number of images for which estimation error was -15 or less, and 'over' represents the case for 15 or greater. Besides, the quantities were estimated by the regression model; thus estimation errors were also decimal values. Therefore, the graph in Figs. 11, 12 and 13 was created after rounding off estimation errors to integers. The vertical axis shows the rate of the number of occurrences of each error, and each figure shows cases of five, 20, 40, 60, and 80 nuts as shown in the legend.

As shown in Fig. 11, in the case where inventory was estimated using CG images using the regression model trained with CG images, the distribution of estimation error was approximately within ± 5. Conversely, as shown in Fig. 13, when using actual images for the same model, the distribution was spread across a wider range, i.e., between -10 and over. Note that errors became over in the 40 and 60 nuts cases. As shown in Fig. 12, when using the fake CG images, the error distribution improved (except the case of 80 nuts) compared to using the actual images, i.e., the distribution was within ± 9. However, deviation to the negative direction increased with 80 nuts.

Furthermore, for inventory estimation using fake CG images, the magnitude and deviation of the errors were evaluated for each quantity of nuts using the mean absolute error (MAE) and the average error. Figure 14 shows the transition of MAE and average error with increasing nut quantity. Here, the vertical axis shows the errors, and the horizontal axis shows nut quantity. In the range of five to 20 nuts, both the MAE and the average error were approximately 4. At 25 nuts, deviation in the positive direction increased. When nut quantity was over 25, the deviation in the negative direction increased linearly as nut quantity increased.

5 DISCUSSION

In a previous study, the estimation accuracy of parts inventories in bulk containers using the regression model was evaluated using the following data: CG data generated by the data generator were used for training the model, and actual images were used to estimate inventory using the trained model. As a result, although training data generation efficiency improved, the problem caused by the domain gap of images occurred and estimation accuracy deteriorated. Furthermore, it

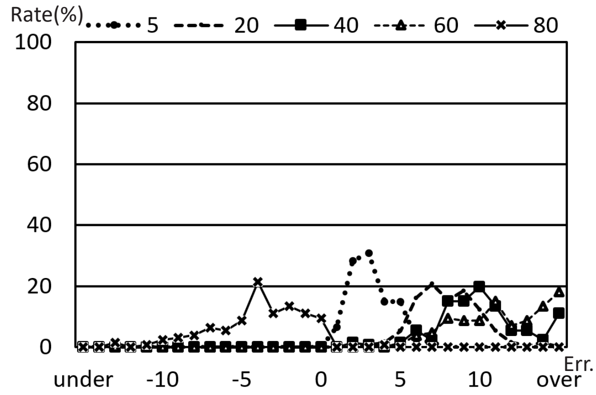


Figure 13: Histogram of error distribution with actual images

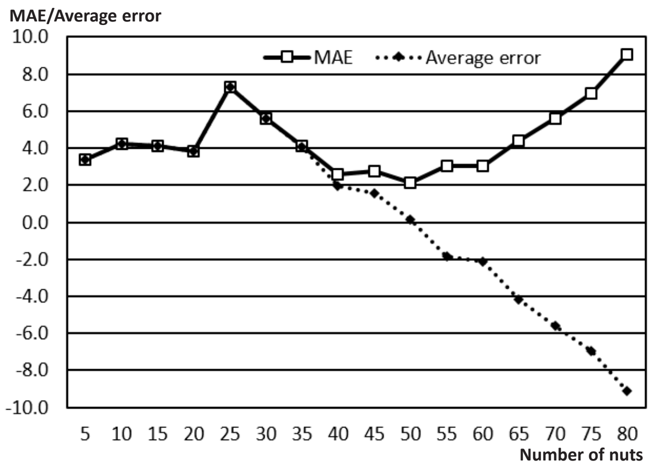


Figure 14: Transition of MAE and average error

was predicted that the same problem would occur due to scene changes in an actual factory. Therefore, to address this problem, the feasibility of translating actual images to fake CG images using a Cycle-GAN was investigated in the current study.

First, I performed comparative evaluations of estimation accuracy with and without image translation using the Cycle-GAN model. As shown in Fig. 10, this image translation process was effective relative to improving estimation accuracy. In particular, as shown in Fig. 14, when the quantity of nut was 20 or fewer, the MAE of the estimation was approximately 4. In actual factories, one of the most important purposes of stock-taking is to prevent running out-of-stock. For this purpose, bulk containers with small inventory quantities are targeted. As discussed in Section 2, deep learning techniques can be applied to actual factory environments by increasing the safety stock and collating with the theoretical inventory. For example, as shown in Fig. 12, when the inventory quantity was 20 or fewer, the error was 10 or less. Therefore, out of stock can be suppressed by increasing the safety stock by 10. Incidentally, it has been confirmed that the types of materials of parts were limited; thus, it is considered that the estimation method using a Cycle-GAN can be applied effectively to multiple parts. Therefore, it is expected that this method can be applied practically in some field even with its current accuracy.

And, two issues related to applying the Cycle-GAN model have been identified. The first issue is related to the method by which the optimal Cycle-GAN model is detected. As shown in Fig. 9, in the range of epochs I have experimented with, strong correlations between the estimation accuracy (MSE) of the regression model and the loss (d_loss, g_loss) of the Cycle-GAN model could not be observed. In other words, from the viewpoint of generating optimal fake CG images for the regression model, the optimal model could not be detected by monitoring only the loss transition of the Cycle-GAN models. As a result, inventory estimation accuracy had to be examined for all fake CG images translated by the models in each training stage, as shown in Fig 9.

The second issue is related to inventory estimation accuracy. As shown in Fig. 10, the MSE obtained using fake CG images was approximately 9.2 times that of using CG images. Furthermore, the transition of MAE varied with increasing nut quantity as shown in Fig. 14. In particular, the MAE increased at 25, 30, and 70 or greater nuts. However, to use this method for stock-taking, it will be necessary to maintain a certain inventory estimation accuracy regardless of the part quantity.

To address these issues, it will be necessary to make the loss function of the Cycle-GAN reflect the loss function of the regression model. Using this model, it is expected that the Cycle-GAN model can be trained to optimal for the regression model to estimate inventory. This will be the focus of future work.

6 CONCLUSION

Inventory estimation of bulk containers using the regression model of deep learning has achieved certain accuracy. However, to apply this method to an actual factory, there were two issues that needed to be addressed, i.e., a large amount of training data must be prepared, and the deterioration of estimation accuracy caused by the domain gap with scene changes must be prevented.

For these problems, in this study, training a regression model using CG images and estimating inventory using fake CG images translated from actual images by Cycle-GAN were investigated. Comparative evaluations of the model's estimation accuracy were performed using CG images, fake CG images, and actual images. As a result, inventory estimation accuracy could be improved using fake images rather than the original images. On the other hand, the estimation accuracy obtained using fake CG images was less than when using CG images.

To improve accuracy when using fake CG images, it will be necessary to reflect the loss of the regression model into the loss function of the Cycle-GAN model, which will be the focus of future work.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 19K11985, and a research grant from Toukai Foundation for Technology.

REFERENCES

[1] A. Anoosheh, et al., “Combogan: Unrestrained scalability for image domain translation,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops, pp. 783–790 (2018).

[2] A. Bansal, et al., “Recycle-gan: Unsupervised video retargeting,” Proc. European Conference on Computer Vision (ECCV), pp. 119–135 (2018).

[3] A. Cherian, and A. Sullivan, “Sem-GAN: Semantically-consistent image-to-image translation, ’ ’ Proc. 2019 IEEE Winter Conf. on Applications of Computer Vision, pp. 1797–1806 (2019).

[4] F. Chollet, “Deep learning with python,” Manning Publications Co. (2017).

[5] F. Fang, et al., “High-quality nonparallel voice conversion based on cycle-consistent adversarial network,” Proc. 2018 IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 5279–5283 (2018).

[6] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, “Deep learning,” MIT press (2016).

[7] I. Goodfellow, “Nips 2016 tutorial: Generative adversarial networks,” arXiv preprint arXiv:1701.00160 (2016).

[8] H. Greenspan, B. van Ginneken, and R. M. Summers, “Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique,” IEEE Transactions on Medical Imaging, Vol. 35, No. 5, pp. 1153–1159 (2016).

[9] P. Isola, et al., “Image-to-image translation with conditional adversarial networks,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1125–1134 (2017).

[10] T. Kawanaka, and T. Kudo, “Inventory Satisfaction Discrimination Method Utilizing Images and Deep Learning,” Procedia Computer Science, Vol. 126, 937-946 (2018).

[11] T. Kim, et al., “Learning to discover cross-domain relations with generative adversarial networks,” Proc. 34th Int. Conf. on Machine Learning-Volume 70, pp. 1857–1865 (2017).

[12] A. Krizhevsky, I. Sutskever, G.E. Hinton, “Imagenet classification with deep convolutional neural networks,” Advances in neural information processing systems, pp. 1097–1105 (2012).

[13] T. Kudo, and R. Takimoto, “A proposal of deep learning training data generator for image recognition,” Proc. Int. Wokshop on Informatics (IWIN2018), pp. 119–126 (2018), http://www.sist.ac.jp/~kudo/research/IWIN2018_119_126.pdf (reffered Oct. 19, 2019).

[14] T. Kudo, “CG Utilization for Creation of Regression Model Training Data in Deep Learning,” Procedia Computer Science, Vol. 159, pp. 832–841 (2019).

[15] E. Linder-Norén, “Keras implementations of Generative Adversarial Networks”, https://github.com/eriklindernoren/Keras-GAN (reffered Oct. 19, 2019).

[16] M. Y. Liu, T. Breuel, and J. Kautz, “Unsupervised image-to-image translation networks,” Advances in neural information processing systems, pp. 700–708 (2017).

[17] S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” Frontiers in plant science, Vol. 7, Art. 1419 (2016).

[18] A. Mordvintsev, C. Olah, M. Tyka, “DeepDream - a code example for visualizing Neural Networks,” Google AI Blog, https://mng.bz/xXlM (reffered June 5, 2019).

[19] M. Verhelst, and B. Moons, “Embedded Deep Neural Network Processing: Algorithmic and Processor Techniques Bring Deep Learning to IoT and Edge Devices,” IEEE Solid-State Circuits Magazine, Vol. 9, No. 4, pp. 55–65 (2017).

[20] Y. Wen, et al., “A discriminative feature learning approach for deep face recognition,” European conference on computer vision, Springer, pp. 499–515 (2016).

[21] L. Wei, S. Zhang, W. Gao, and Q. Tian, “Person transfer gan to bridge domain gap for person re-identification,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 79–88 (2018).

[22] L. Wu, Y. Wang, and L. Shao, “Cycle-consistent deep generative hashing for cross-modal retrieval,” IEEE Transactions on Image Processing, Vol. 28m No. 4, 1602–1612, (2018).

[23] Y. Yuan, et al., “Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops, pp. 701–710 (2018).

[24] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” Proc. IEEE Int. Conf. on Computer Vision, pp. 2223–2232 (2017).

(Received October 29, 2019)

Tsukasa Kudo received the BSc and ME from Hokkaido University in 1978 and 1980, and the Dr. Eng. from Shizuoka University in 2008. In 1980, he joined Mitsubishi Electric Corp. He was a researcher of parallel computer architecture and engineer of business information systems. Since 2010, he is a professor of Shizuoka Institute of Science and Technology. Now, his research interests include deep learning and database application. He is a member of IEIEC and IPSJ.

International Journal of Informatics Society, VOL.12, NO.1 (2020) 49-58

Regular Paper

Implementation and Evaluation of Prompting Changeover System from Repetitive Behavior

Yuya Miyawaki† and Katsuhiko Kaji†

†Aichi Institute of Technology
1247 Yachigusa, Yakusa, Toyota, Aichi, 470-0392 Japan
b18726bb@aitech.ac.jp

Abstract - Children with developmental disorder have more difficulty engaging in changeover behavior than children without such disorders. Since this characteristic will be more critical as they grow and become adults, prompting changeover behavior is very important. When supporting prompting changeover, reducing the load on them and their supporters is also required. We developed support equipment and verified its effectiveness to provide assistance that prompts changeover by voice and vibrations. The supporters operate our support device when a changeover is required. The changeover support is done based on the target object of the repetitive behavior. Therefore, since children with developmental disorders feel that the target object of the repetitive behavior is itself providing the support, they will respond to the prompt of the changeover support. We analyzed the effect of support in a case where changeover support was implemented continuously and verified it by continuing to prompt changeover support at the children's home with them and recording their progress. We video-recorded their emotional state with the supporter who also evaluated our device. The prompting changeover success rate improved with repeated support. However, it might decrease due to changes in their physical condition. We conclude that our proposed support provides effective changeover support for repetitive behavior.

Keywords: Children with developmental disorder, Repetitive behavior, Changeover support, Support device

1 INTRODUCTION

Children with developmental disorders have more difficulty doing changeover behavior than children without them [1]. This is a major issue for children with developmental disabilities as they become adults and join society [2]. Such children must learn changeover behaviors to live productive lives. The load on supporters during changeover support is another major issue [3]. Many supporters want to reduce their load and seek mental support. Based on this background, we reduced the changeover support load for both the benefit of children with developmental disabilities and their supporters.

This research proposes a device and support method that provides changeover support for developmental disabled children by voice and vibrations (Fig. 1). We implemented our proposed support and verified its effects. In our proposed support, a supporter operates a device when the target child must make a changeover. The device is installed on the child's target object. The device supports speech, vibrations, etc. based on its particular operation. The target child experiences the support from the target object. In other words, we provide support from within the repetitive behavior space[1].

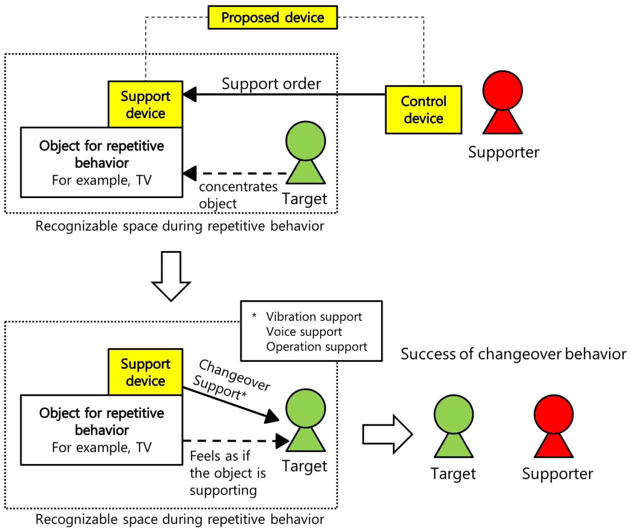


Figure 1: Support system outline

We designed a support device and implemented a prototype to achieve our proposed support. The prototype provided three kinds of support: vibration, speech, and operation. First, vibration support gives a moving stimulation through the target object that is held or worn by the target child. Thus, the target child knows when the changeover support will begin. Second, speech support is generated from the target object and provides two types of changeover voice support: preparatory and parental voices. Third, operation support is provided so that the target child can understand the changeover support mechanism. To that end, the supporter directly manipulates the target object to overwrite the existing operation on the target object. This enables the target child to understand the support.

We verified our system to determine whether the prototype provides support. A family with a special needs child cooperated in our study and we video-recorded the child's performance with the supporter who also evaluated our system. We analyzed the difficulties of the changeover behaviors, which were conducted continuously and evaluated the effectiveness

[1]Repetitive behavior space is the target child's recognition space during repetitive behavior.

of our system using the analysis data.

2 RELATED WORK

In recent years, research on developmental disorders has increased [4]–[6]. Although changeover support away from obsessive behavior is one basic type of support [7], unfortunately, few studies have addressed it. Direct support using voice and gestures is common.

However, if the changeover support does not work very well, the burden falls on the supporter [8], [9]. In addition, smooth behavior in a group might be hindered. Therefore, in the research on changeover support, effective methods that provide general direct support and ways of creating ideal environments for changeover support have been discussed [10]: in other words, making successful prompts. Little research has focused on ”what to do.”

In addition, prompting tools have been developed, for example, a support device that offers an auxiliary role for intentional transmission. Support devices that use visual information, which are typically represented by picture cards, and support devices that use such aural information as voice, are starting to be used in some fields.

2.1 Related Support Method

Support exists that matches the environment to the target child [11]. For example, the child's supporter waits until the environment and the target child's mindset can be switched. In this way, the supporter acts in coordination with the target child who can live based on his or her own ideal situation. Therefore, such support benefits the target child. On the contrary, forced changeover behavior increases the stress on the target child and strengthens the attachment to the obsessive behavior.

Other support prepares the target child's engagement in the activity [12]. In this support, the child must act base on a schedule. A schedule table or timetable combined with picture cards should be set in front of the target child to reduce the anxiety of the changeover behavior. The schedule also prompts self-directed changeover behavior. However, due to the support for changeover based on the situation, it is impossible to act based on the surrounding situation. Therefore, the target child cannot accomplish collective action.

Voice support establishes changeover behavior by a repeated voice, for example, providing preliminary preparation for an imminent changeover. Voice support does not just denote that such sound is required. Timing and voice strength are also important.

2.2 Related Support Device

Support devices for communication using voice also exist [13]. Speech recognition can be achieved using such a support device for children who can't speak very well [14].

Such recognition enables the voice-based indication of the intention of the target children who want to communicate but can't speak well. The target child operates a device in which the supporter's voice was recorded in advance and emits audio that transmits the target child's intention to the supporter by voice. The supporter can also communicate through the hand of the target child who holds the device. In other words, the support device can actively communicate the supporter's intention [2].

Table 1: Hearing and observation survey

Location	Repetitive behavior
Home	Watching on DVD player Watching on a mobile DVD player Reading a magazine
Outing	Watching on mobile DVD player Supporter sings
Support facility A	Listen to CD player Reading a magazine
Support facility A Outing	Supporter sings
Support facility B	Reading magazine
Support facility C	Listening to CD player

For digital devices, systems have been proposed that extend conventional support using tablets and PCs [15]. Using digital devices can significantly reduce the cost of producing individual tools, which is a problem with conventional support. This step will also reduce the burden on supporters. Unfortunately, many target children can't use digital devices because they have difficulty correctly inputting commands or they don't know which icons to touch. Supporters might also not be familiar with advanced technology. They might also have trouble making settings.

3 REPETITIVE BEHAVIOR FOR CHILDLEN WITH DEVELOMENTAL DISORDER

The repetitive behaviors of children with developmental disorders are significantly different from those of children without them, especially in the time, period, and attachment force that occur. Children with developmental disorders show a marked attachment to certain items or situations. They never want to relinquish them, which causes repetitive behavior.

Repetitive behavior is associated with place. However, the target child only changes the repetitive behavior based on the place. In other words, the essence of repetitive behavior does not change. Table 1 shows the results of interviews with collaborators in this study. From interview surveys, we identified behaviors where children repeatedly looked at a particular scene in a video or listened to a particular piece of the music. Repetitive behavior changes depending on the place. However, the behavior of playing videos, listening to music, and reading does not change. In other words, some acoustic or visual stimuli are repetitive behaviors.

Since the changeover from repetitive behavior is stressful for children with developmental disorders, they have difficulty changing. One major obstacle has been identified in the performance of children with developmental disorders and those without them. In addition, the group life of children

[2]Current support includes a type that allows children to choose cards and intentional communication by pointing.

with developmental disabilities is extremely difficult due to their repetitive behavior. Therefore, since they can't cope with their environment because it fails to match their feelings, they can't digest sudden schedule changes. Compulsory changeover to adjust to the surrounding situation causes a great deal of stress on the target child. As a result, he may engage in self-harm or hurt others. Forced changeover supports counterproductive effects. To reduce such stress, it is important to spontaneously prompt the child to make the changeover.

The following are supports that promote conventional changeover action: using picture cards; using timers and watches; using voice overs. Such support is provided at homes and schools for children with developmental disabilities and afterschool services. In the support by picture cards, a picture card shows the next action. By focusing on the target child before the action or the changeover, we can make image the action after the changeover and promote changeover behavior. A timer or a clock can provide support by announcing the changeover time. The target child can imagine how much of their own behavior they can do later to promote changeover behavior. Before a switch is made, the voice support announces: "It will soon be time to make a changeover." We can immediately create an image that changes the child's target and encourage the changeover behavior. Critically, these types of support should be used to gradually increase actions that the target child can do but dislikes. In other words, our method avoids compelling responses from target children and making them feel bad. Such behaviors must only be increased with activities the target child wants to do but can't.

However, children with developmental disorders only concentrate on objects of interest. Therefore, often support from the surrounding people is not transmitted. The following support is provided, but it is stressful for children with developmental disorders. One example is support between the objects of interest and children with developmental disorders: physically separate the objects and children with developmental disorders. Stop the operation. In addition, attachment to a supporter may occur. Such attachment to a supporter might turn him/her into a favorite, and they decide they prefer support from that supporter and not from the others. Hence, they only accept support from their favorite supporter.

4 CHANGEOVER PROMOTION SYSTEM FROM REPETITIVE BEHAVIOR

Our proposed system provides the following support, as shown in Fig. 1. The target child concentrates on his target object. Perhaps the assistance from the surrounding supporters will not be accepted. In this research, the target child's recognition range is defined as the Repetitive Behavior Space. This space is conceptual and varies in scope and size depending on the target child. A support device is attached the child's target device to provide support to the objects of choice. The supporters benefit the target child using remote control equipment outside of the Repetitive Behavior Space. Our changeover promotion system generates voice and vibrations to support the target child who can recognize whether the target object itself is vibrating or speaking. This enables support from inside the Repetitive Behavior Space that is expected to encourage changeover behavior. Even if the support person changes, the voice support from the object of attention does not change. This resolves the issue of attachment to a particular supporter.

The following is the support order of our proposed system that promotes changeover behavior. The vibration support device is activated, and it notifies the target child at the beginning of the changeover behavior. The voice support device is activated to prompt the child to prepare for a changeover. The operation support device is activated to orally state the action. Actions can be associated with states. To realize this kind of support, we designed a device as shown in the Fig. 2.

4.1 Vibration Support Device

The vibration support device generates vibrations on the target object of interest to inform the target child of the support. Since this support uses vibrations, the target child must be touching the target object to receive support. When not making contact, an object must be used that is in contact with an item that is related to repetitive behaviors.

The supporter starts the device before beginning the support. After activation, the vibration motor attached to the vibration support device is turned on, and the object of interest or the related object to which the device is attached vibrates. This vibration stimulates the target child's skin because we expect an action that simplifies responding to the stimuli due to poor hearing.

4.2 Voice Support Device

The speech support system provides support using speech from the target object. The audio attached to the object of interest plays back a sound that can facilitate the changeover. We chose the same voice used at homes and educational settings based on our assumptions that it would be effective.

When the target child is notified by vibration support, support is provided using this device. The assistant starts the voice support device, which reproduces a voice registered in advance. Since the sound is reproduced from the target object, the target child recognizes that target object itself is speaking. Support is provided from the target object enjoyed by the target child. We believe that this support will have the same effect as support provided by a favorite supporter.

4.3 Operation Support Device

The operation support device enhances the effect of our proposed system by controlling the target object's power so that operational control cannot be relinquished to the target child. This device is not intended to be used continuously. In other words, we only assumed its use at the initial stage of support. This device is only effective when the object of interest runs on electricity.

It controls the actual operation with the voice support device. Since the changes in the target might simplify the target child's understanding, changeover actions are only promoted by voice support.

4.4 Implementations for Home Applications

Three devices are controlled by a remote-control device, which wirelessly controls the device. Its operation must be simple so that the supporter who uses the support device can do so easily. For practical operation of our proposed system, the devices are linked by the implementation and home environments.

We conducted interview surveys and observed the home environments of the families from whom we obtained prior consent for cooperation with our research. Based on the information obtained from the household environment, the interview surveys, and home observations, we tailored our equipment to the living environment of the target child. Since target of the basic commitment at home was watching DVDs, we installed our equipment on a DVD and attached our equipment to it.

Based on a preliminary survey, the child's attention behavior was performed for the following period: 1) from waking up to going to school and 2) from returning home to taking a bath.

At home, in addition to watching a DVD, viewing on a mobile DVD player was also possible. Changeover behavior is classified as not continuable or continuable. Not continuable changeover is defined as changeover behavior that can't continue the repetitive behavior after the changeover (Fig. 3). Such changeover behavior is completely separated from the target object. Continuable changeover is defined as changeover behavior where the repetitive behavior can continue after changeover (Fig. 4).

One item that the target child must have during the selective actions is the DVD remote control. We attached a vibration support device to the remote control so that it could always be used during playback. We attached a small device to the remote control to reduce the risk of inconvenience when using it. Figure5 shows the remote control to which we attached our mounted device. The vibration support device was attached to the back of the remote control for the target object. Since the device is controlled by an infrared signal, we designed it so that the light receiver can't be hidden when the target child carries it.

The target child's repetitive behavior is watching DVDs. In this experiment, we defined the obsessed object as the object of repetitive behaviors. Therefore, the target object in this example is the TV screen. The voice support device needs to be installed around it so that the sound can be heard. The home appliance control support device needs to be installed around the output cable of the DVD player. We integrated the voice support device and the home appliance control support device next to the TV display. The target object's power is drawn to the device, and control is done using an infrared signal. The voice support device also incorporates an independent control mechanism using a commercially available system and controls from a smartphone application. Figure6 shows the attached device, and Figure7 shows its installation.

The supporter operates each device at each changeover behavior. Basic operations are done using the infrared remote control (Fig. 8) that was implemented in an easy-to-understand manner using large operation instructions and colored buttons. By pressing the button, the infrared signal is emitted to control the target device. The voice support device can record, register new assistance, and so on. Each operation is performed using a smartphone application. Figure9 shows the positional relationship introduced into an actual environment. The vibration support device is held by the target child.

Table 2: Classification of changeover behavior in child

	Not continuable	Continuable
Favorable		Message
Unfriendly	Meal Going out Doing homework Taking a bath	Brushing teeth
Voluntary	Going to the bathroom	

5 VERIFICATION OF EFFECT BY SUPPORT FROM THE TARGET OBJECT

Our proposed system provided continuous support with the cooperation of a female junior high school special support school, a first grader, and her family. This child does not talk with others. However, she can understand the voices of her supporters to some extent. Communication from her is based on gestures or pointing. When the child is engaged in repetitive behavior, she tends to prioritize the repetitive behavior over the voices of the surrounding supporters. Therefore, in many situations she cannot transition to the changeover behavior.

We recorded from March to September 2018. Spring and summer breaks were included. The recording was mainly done by the mother who was also a supporter at home, and we did video recording on our visitation days. We observed the changeover behaviors that occurred at home and classified them with the supporters. The classifications are shown in Table 2. We predicted and classified the situations we observed and determined whether the target child liked or disliked the changeover behaviors itself. Since the changeover behavior for going to the bathroom is a voluntary changeover, we did not analyze it.

As shown in Table 2, we made two major classifications based on the content of the changeover behavior. Not continuable changeover behavior is defined as changeover behavior that cannot be continued after the changeover. Since repetitive behavior can't be continued after a changeover, we can't changeover a behavior even with such support such as the direct voice of supporters and gestures. Continuable changeover behavior is defined as repetitive behavior that can continue after a changeover. In many cases, changeover behavior can also be done by summoning the supporters. In addition, the target child performs a changeover behavior based on the information of the sound of the voice support. However, there was also a case where the changeover behavior differed from the content of the support. Therefore, the target child probably performed the changeover behavior by observing the surrounding situation based on the assistance of the voice support.

We propose support based on the timing of the occurrence

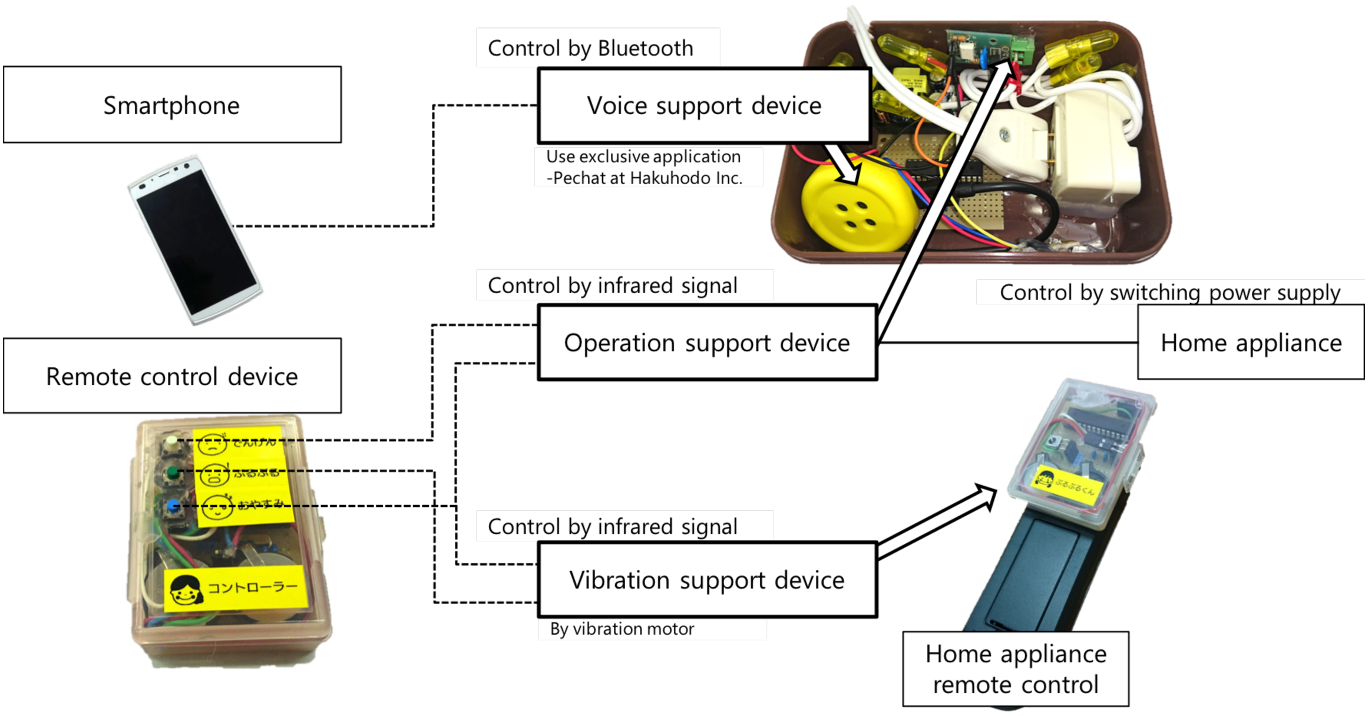


Figure 2: Support system detailed overview

Figure 3: Example of not continuable behavior (Dinner)

Figure 4: Example of continuable behavior (brushing teeth)

of the daily changeover behaviors. The supporter operates the support device based on the procedure. The supporter observes and records the target child's behavior. In the procedure, the voice support for the preparation of the target child's changeover is implemented just before the occurrence timing of the changeover behavior. Here is an example of a voice support: "Please finish soon." The voice's content was decided in consultation with the support person based on the voice usually heard at home.

At the early stage, the content was based on the usual practices at home. As the talking time increased, the child's cooperation became more effective. In other words, by increasing the number of sentences per utterance, we more effectively prompted changeover behaviors. In the second half, the same word is repeated twice or the name of the target child is spoken.

Commitment behavior in the home is performed intermittently from the time of getting up for school on weekdays, going out (weekends and holidays), and returning home for a bath. The target child watches his favorite DVD (Fig. 10). In some cases, he watched on a mobile DVD player (Fig. 11).

5.1 Experiment Setting

In this experiment, we calculated the success rate of the prompting of the changeover behavior using our proposed support. Then we analyzed its effectiveness. The success rate was based on situations where the target child did the changeover behavior just based on reminders from the sup-

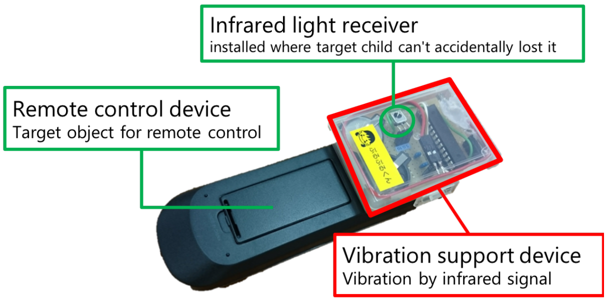


Figure 5: Vibration support device

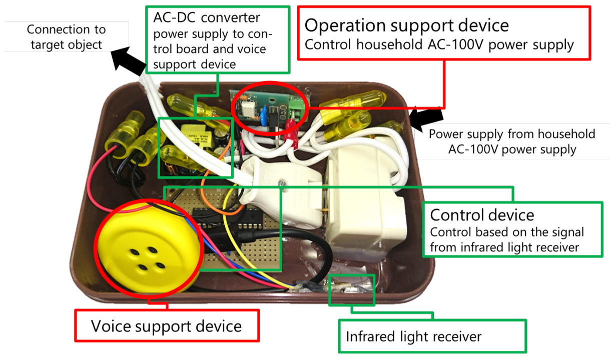


Figure 6: Integrated voice and operation support devices

Figure 7: Support device installation diagram

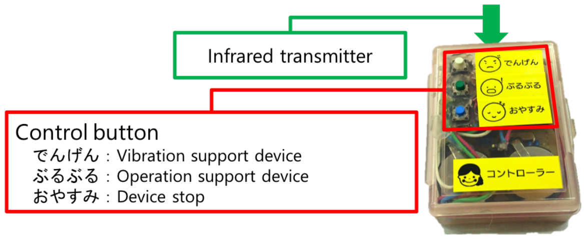


Figure 8: Infrared remote control for basic operation

port device. First, the support person operated the support device for the changeover. Next the support person called out to promote the changeover behavior. At this stage, the supporter was aware of and pretended to ignore the target child so that we can analyze the effect of the summoning support from the support device. If no changeover behavior is seen in this state, the strength of the support from the support person is sequentially increased to promote changeover behaviors. We increased the strength of the support in the following order: direct support from the supporter, gestures, stopping the object of interest, and taking arms and guiding.

We recorded the extent of the increase of the support intensity in a five-step evaluation. Table 3 shows the success criteria for the changeover behaviors. In the not continuable behavior, a case where the behavior shown in the table 3 is performed after the intention to end or end the sticking action by itself is regarded as success. In the continuable behavior, the continuable behavior is performed while continuing the repetitive behavior. Therefore, even if the repetitive behavior is ongoing, if the behavior shown in the table 3 is performed, it is successful. In addition to the changeover behavior results, the target child's daily situation was recorded and analyzed. Figure12 shows an actual paper record of a target child. Our analysis used these aggregated weekly data for Fig. 12.

5.2 Success Rate of Changeover Behavior by Proposed Support

The graph obtained by aggregating the weekly changeover behavior success rate based on our proposed support is shown in Fig. 13. The success rate varies greatly in a short time due to factors such as physical condition. Therefore, analysis is performed using the average of the success rates for one week. Each element represents a not continuable behavior and an continuable behavior, respectively. Each of the changeover behaviors was also grouped and averaged.

The not continuable changeover behaviors can't continue the repetitive behavior after the changeovers. Therefore, the changeover success rate is low. On the other hand, Continuable changeover behavior can continue the repetitive behavior after the changeover. So the target child easily made the changeover. The data from March 25 and April 8 are missing because the subject was sick and the proposed support was not implemented. The success rate of the changeover behaviors between the not continuable and continuable changeover behaviors fluctuated. For example, since the target child's physical condition did not improve in the week of May 27th, the tabulation end date where the changeover behavior success rate was significantly reduced for both the not continuable and continuable changeover behaviors.

We focused on such changes in the physical condition and the environment and analyzed their occurrence rates, the changes in the environment, and the changes in the continuation of support. Section 5.3 describes the transition of the changeover behavior success rate, Section 5.4 describes the changes in the environment, and Section 5.5 describes the changes due to continuous support.

5.3 Transition of Changeover Behavior Success Rate

The success rates of the changeover behavior between the not continuable and continuable changeover behaviors changed depending on the physical condition of the target child. In the week where the changeover behavior success rate decreased, both the not continuable and the continuable changeover be-

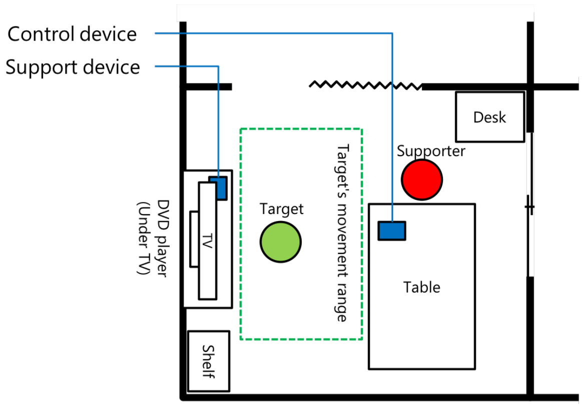


Figure 9: Positional relationship among support system, supporters, and target child

Figure 10: Target child who watched DVDs

haviors decreased, and the rate increased in subsequent weeks. The decrease factor is sensitive not only to the child's physical condition but also to her environmental changes.

In the continuable changeover behavior, almost 100% was maintained in the week when the child had a good physical condition and a good environment. Therefore, for the continuable changeover behavior, we continue to propose support and establish the changeover behavior. However, if the support person fails to show the toothbrush at the changeover behavior for brushing the teeth [3], the target child may exhibit this behavior (the changeover behavior to the massage) and direct her foot to the supporter (Fig. 14).

5.4 Change in Changeover Behavior Success Rate due to Environmental Changes

A major change in the environment revolved around whether to go to school. The experimental period included spring vacation from March 18 to April 8 and summer vacation from July 22 to September 2. We confirmed a decrease in the changeover behavior success rate during both holidays.

The child's waking-up time and bedtime changed because the routine of going to school was interrupted. In addition,

[3]The supporter has a toothbrush in his hand but does not show it to the target child.

Figure 11: Target child who watch DVDs by mobile DVD player

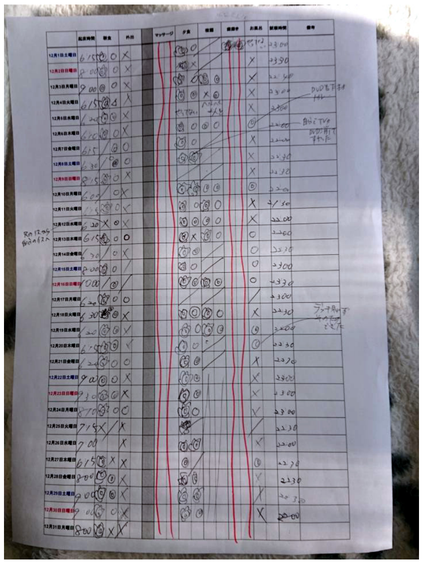


Figure 12: Support situation record sheet

the decrease in the changeover behavior success rate that occurred early in the first semester was probably affected by the change in the support method at school due to a new teacher. The changeover behavior success rate decreased based on such environmental changes, and the changeover behavior success rate and changes in the living environment have a large relationship.

5.5 Change in Changeover Success Rate due to Continuation of Support

Although the changeover behavior success rate fluctuates based on environmental changes and physical conditions, the appearance tendency tends to improve as a whole. This improvement trend shows better use of our continuous proposed support. The decrease in the success rate due to environmental and physical condition factors can be judged by a comparison with the continuable changeover behavior, which is the success rate at a normal 100%. In other words, except for

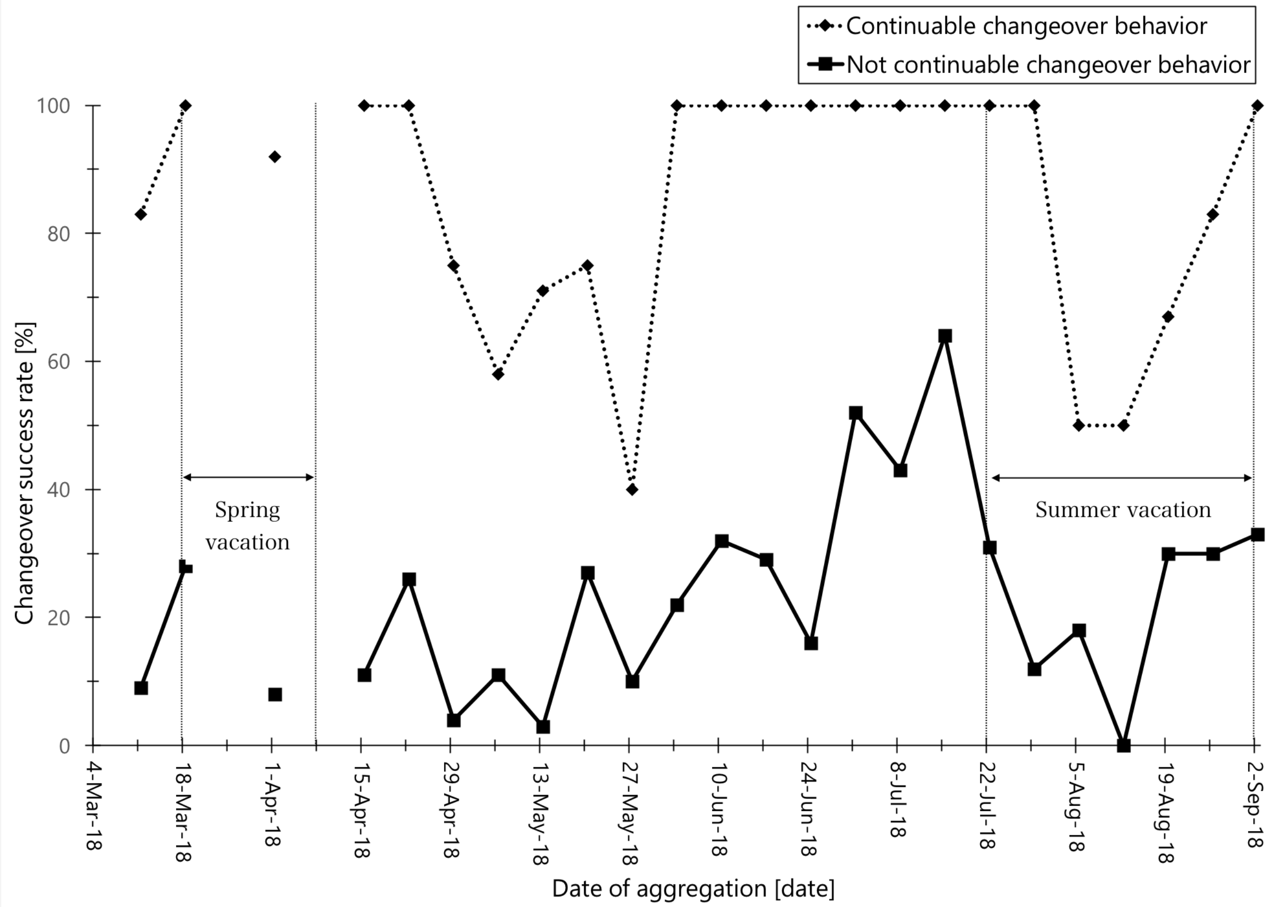


Figure 13: Graph of changeover behavior success rate

Figure 14: Failure to brush teeth changeover behavior

the week when the changeover success rate of the continuable changeover behavior decreased, the general tendency showed an increase. The assistance with our continuous proposed support is effective for successful changeover behaviors.

Before summer vacation, the changeover behaviors were successful with a 60% probability. The incidence rate improved by about 40% in three months. Due to the environmental change of summer vacation, the emergence rate decreased. If we can incorporate such environmental changes, we expect a more stable changeover behavior success rate. If our proposed support is used continuously and assistance is done as such, the target child can expect better growth. We believe that our support device is effective. The supporters themselves also commented that the target children's reactions improved.

Table 3: Achievement standard of goals

Changeover behavior	Achievement standard
Breakfast	Moving to the table and eating
Dinner	Moving to the table and eating
Go out	Moving to the door
Homework	Doing homework
Taking a bath	Moving to the bathroom
Massage	Putting feet on the supporters
Toothpaste	Turning head to supporter and lie down

5.6 Change in Supporter Burden

After the experiment, we investigated the burden on supporter. For this reason, supporter were interviewed. From the results of interviews with supporters, the following are the factors that reduce and increase the burden.

The burden reduction factors are as follows: I can help when I am away from my child. Even at a position where my voice could not normally be heard by the target child, support was provided through the device. (However, it took about a year to be fully responsive.) For the same reason, the same effect was felt even when the supporter could not speak due to a cold or the like. The device changed the supporter's feelings for support. As a result, we were able to provide continuous support. Changeover is now successful even if the supporter changes. (Can change over with the support of her brother.) At first, they did not respond without assistance from the subject. With continuous support, I became responsive when I heard the sound.

The factors that increase the burden are as follows: Immediately after I started support, she responded with a rejection. I bit myself and me.(About 2 to 3 months) The operation support device has reduced the number of DVD viewings. She began to watch mobile DVDs that could not be controlled by the device. She started watching DVDs in her favorite room, and I could no longer monitor her. (However, watching mobile DVDs has increased her mobility. We think that her stress was reduced by this.) I make a mistake in operation the device. There is a bias in the devices that can be controlled. Therefore, the target child can escape to another repetitive behavior. For example, she changed from a DVD player to a mobile DVD player.

6 CONCLUSION

We reduced the burden on both the target child and the supporters during changeover behavior from obsessive behavior and obtained spontaneous changeover behavior. The early stage not continuable changeover behavior is not practical because the changeover behavior success rate is low. However, continuous support was maintained, and we eventually achieved an emergence rate of 60%, proving that this support is effective. On the other hand, a 100% success rate can be maintained in continuable changeover behaviors. We believe that our process has advanced to a practical stage.

In the future, we will provide more continuous support and investigate how long it can improve the success rate in accordance with environmental changes and physical conditions. We must also analyze the problems of our proposed support based on the situation and changes of the target child and investigate whether a significant adverse effect exists.

We haven't considered the burden on supporters yet. Therefore, we must implement a new support device that can automate operations and support with a low load. However, we received the following comments from a supporter. "Using the device as a trigger for changeover behavior allowed smooth changeover behavior." "Although at times using the device was burdensome, it switched my feelings." "The child asked to use the device." In other words, using the device itself is a burden. However, the support environment was improved by the device. Currently, control is done by a smartphone. We can achieve operation with less burden by implementing control with physical buttons that can be operated immediately, automatic voice call support based on the situation of the target child, and semi-automatic trigger call support, etc. We must also devise equivalent support in multiple places to cope with the target child's movement in the room and a smaller device that can cope with changes in the target object.

REFERENCES

[1] F. R. Volkmar, R. Paul, A. Klin, and D. J. Cohen. "*Handbook of autism and pervasive developmental disorders, diagnosis, development, neurobiology, and behavior*". John Wiley & Sons, (2005).

[2] C. Johnston and E. J. Mash. "families of children with attention-deficit/hyperactivity disorder: review and recommendations for future research". *Clinical child and family psychology review*, Vol. 4, No. 3, pp. 183–207, (2001).

[3] A. Yamada, M. Suzuki, M. Kato, M. Suzuki, S. Tanaka, T. Shindo, K. Taketani, T. Akechi, and T. A. Furukawa. "emotional distress and its correlates among parents of children with pervasive developmental disorders". *Psychiatry and clinical neurosciences*, Vol. 61, No. 6, pp. 651–657, (2007).

[4] T. Smith, A. D. Groen, and J. W. Wynn. "randomized trial of intensive early intervention for children with pervasive developmental disorder". *American Journal on Mental Retardation*, Vol. 105, No. 4, pp. 269–285, (2000).

[5] S. Herring, K. Gray, J. Taffe, B. Tonge, D. Sweeney, and S. Einfeld. "behaviour and emotional problems in toddlers with pervasive developmental disorders and developmental delay: associations with parental mental health and family functioning". *Journal of Intellectual Disability Research*, Vol. 50, No. 12, pp. 874–882, (2006).

[6] E. G. Carr, G. Dunlap, R. H. Horner, R. L. Koegel, A. P. Turnbull, W. Sailor, J. L. Anderson, R. W. Albin, L. K. Koegel, and L. Fox. "positive behavior support: Evolution of an applied science". *Journal of positive behavior interventions*, Vol. 4, No. 1, pp. 4–16, (2002).

[7] O. I. Lovaas. "*Teaching individuals with developmental delays: Basic intervention techniques*". Pro-ed, (2003).

[8] J. Bromley, D. J. Hare, K. Davison, and E. Emerson. "mothers supporting children with autistic spectrum disorders: Social support, mental health status and satisfaction with services". *Autism*, Vol. 8, No. 4, pp. 409–423, (2004).

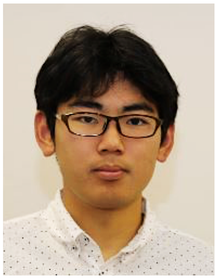
[9] B. A. Boyd. "examining the relationship between stress and lack of social support in mothers of children with autism". *Focus on autism and other developmental disabilities*, Vol. 17, No. 4, pp. 208–215, (2002).

[10] D. Erbas. "a collaborative approach to implement positive behavior support plans for children with problem behaviors: A comparison of consultation versus consultation and feedback approach". Vol. 45, No. 1, pp. 94–106, (2010).

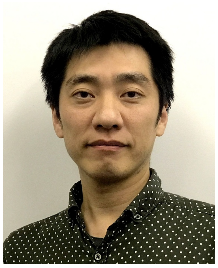
[11] R. Iovannone, G. Dunlap, H. Huber, and D. Kincaid. "effective educational practices for students with autism spectrum disorders". *Focus on autism and other developmental disabilities*, Vol. 18, No. 3, pp. 150–165, (2003).

[12] S. Dettmer, R. L. Simpson, B. S. Myles, and J. B. Ganz. “the use of visual supports to facilitate transitions of students with autism”. *Focus on Autism and Other Developmental Disabilities*, Vol. 15, No. 3, pp. 163–169, (2000).
[13] J. B. Ganz, M. B. Boles, F. D. Goodwyn, and M. M. Flores. “efficacy of handheld electronic visual supports to enhance vocabulary in children with asd”. *Focus on Autism and Other Developmental Disabilities*, Vol. 29, No. 1, pp. 3–12, (2014).
[14] J.Richer. “the partial noncommunication of culture to autistic children—an application of human ethology”. In *Autism*, pp. 47–61. Springer, (1978).
[15] M. F. Kamaruzaman, N. M. Rani, H. M. Nor, and M. H. H. Azahari. “developing user interface design application for children with autism”. *Procedia-Social and Behavioral Sciences*, Vol. 217, pp. 887–894, (2016).

(Received October 31, 2019)
(Revised January 05, 2020)



Yuya Miyawaki is a graduate student at the Aichi Institute of Technology. His current research focuses on support methods for children with developmental disabilities as well as the supporters of such children. He is also interested in support methods for challenged children.



Katsuhiko Kaji received a Ph. D. in information science from Nagoya University in 2007. He became a RA at the NTT Communication Science Laboratories in 2007 and an assistant professor at Nagoya University in 2010. He has been an associate professor of the Faculty of Information Science, Aichi Institute of Technology since 2015. His research interests include indoor positioning and remote interaction. He is a member of IPSJ and JSSST.

Industrial Paper

Development of a Life Watching Service for Elderly People by using Interactive Home Robot

Keiichi Abe*, Hiroki Fuse**, Makiko Okumura*, and Masao Isshiki *

* Kanagawa Institute of Technology, Japan
*** Graduate School of Engineering, Kanagawa Institute of Technology, Japan
abe@he.kanagawa-it.ac.jp

Abstract - Since the life watching system using the existing voice interactive robot uses a camera, there is a problem that some users feel mental stress from the constant monitoring by the camera. Therefore, in this research, we propose a service to watch the life of a target person without using any camera by using a voice interactive home robot for the life watching target including the elderly living away from their families. Specifically, by recording and analyzing information such as conversations between the watching target and the home robot as a log, we examined whether it could be applied as a life watching service equivalent to a camera. We developed a trial system of the life watching service using an interactive home robot proposed in this paper and evaluated by comparing it with a camera.

Keywords: Communication Robot, Life Watching Service, ECHONET-Lite.

1 INTRODUCTION

In Japan, the number of elderly people living alone is increasing with the declining birthrate and aging population. Under such circumstances, life watching services for the elderly [1]-[2] are attracting public attention. There are many services [3]-[4] that use robots for the watching. The basic functions of the existing watch robots are communication, and various functions such as telephone functions and schedule functions are implemented. Many of these robots use cameras for the watching. A robot equipped with a camera [3]-[4] can watch the life of the target person with video, but it cannot watch the target person when he/she is in a room where the robot is not installed or within a blind spot of the camera. There is also a possibility that the image may be diffused on the net by being hacked. Also, the subjects may have multiple problems such as the mental stress due to the constant monitoring with a camera [5].

Therefore, in this study, we propose a service to watch the life of a target person without using any camera by using a voice interactive home robot for the life watching target including the elderly living away from their families. Specifically, we examined whether it could be applied as a new life watching service by analyzing a log recording the talks between the watching target and the home robot. What is essential for such a system is how to increase the chances of conversation with the subjects of the life watching including the elderly. Therefore, the dialogue type home robot for life watching that we developed this time starts talking when someone is detected by the human sensor etc. The robot also produces facial expressions and movements such as "nodding" and "swinging" when talking. We will describe how such a method helps to increase the chances of conversation in an easy-to-talk manner. The prototype development and verification evaluation of the interactive home robot proposed by this research is carried out, so the details will be described. In this paper, Chapter 2 describes related works and issues, Chapter 3 is an outline of the life watching service using the interactive home robot proposed by this research, Chapter 4 relates to prototype development, Chapter 5 is evaluation by demonstration experiment of prototype, Chapter 6 is questionnaire evaluation on comparing prototype and AI speaker, Chapter 7 presents the conclusion, and Chapter 8 is concerned with future works.

2 RELATED WORKS AND ISSUES

Existing technologies for home robots can be divided into two types: interactive and non-interactive. Examples of interactive home robots include Pepper [6] and RoBoHoN [7]. In addition, there is LAVOT [8] developed by GROOVE X as a non-interactive home robot. LAVOT is a robot that cannot communicate with the user, but can express a warm feel and cute gestures like a pet. On the other hand, in the case of interactive home robots such as Pepper and RoBoHoN, their appearance is designed to be familiar to general users. In addition, by adding a dedicated application and developing an original application, it is possible to enter a loving nod or gesture, recognize a person with a camera, etc., and have a voice conversation with a specific person. These home robots such as Pepper [6], RoBoHoN [7], and LAVOT [8] are all equipped with cameras, so they can identify the target person or monitor the situation in the house with the cameras. It is possible to perform advanced monitoring, such as sending the captured images by e-mail to the users. However, when using a camera-equipped robot, some target persons feel mental stress due to the constant monitoring by a camera. Also, if the system is hacked, there is a possibility that the life situation of the target person will be completely viewable, and there are security and privacy issues with the camera-mounted type. In this study, we propose a service that uses an interactive home robot to monitor the user's life only by analyzing log data without using any camera.

3 LIFE WATCHING SERVICE USING INTERACTIVE HOME ROBOT

Chapter 3 gives an overview of a life watching service

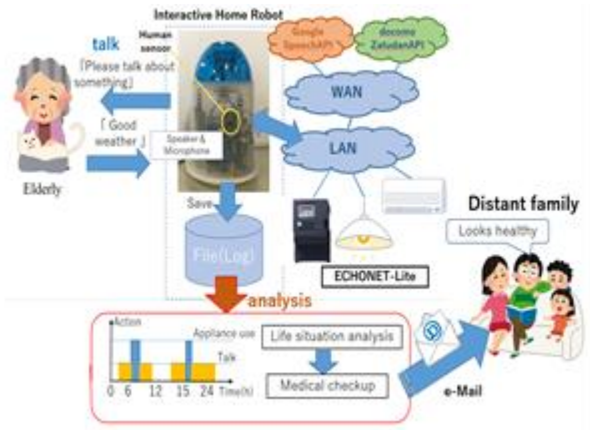


Figure 1: Overview of life watching service using IHR

using the interactive home robot (hereinafter referred to as “IHR”) proposed by this paper and details of the prototype IHR developed this time.

3.1 Outline of life watching service using IHR

Fig. 1 shows the outline of the life watching service using IHR proposed by this paper. This service has an implemented function to record the contents of the conversations between IHR and target person as a log. Log data is saved as a file in CSV format and analyzed by a control program of the interactive home robot and used for the life watching. IHR's natural conversation is realized by combining Google Speech API of speech recognition service and chat conversation API of NTT docomo [9]. Classification numbers are provided according to the contents of the log data acquired by IHR (conversation, home appliance operation, etc.), and IHR estimates the life behavior of elderly people living alone by graphing the data in time series etc. Using those functions, we devised a system service to monitor life by comparing the living condition data with those acquired in the past. The final analysis results of life surveillance are transmitted to the families and relatives living in remote areas by e-mail. The IHR is designed for use in a smart house [10] where home appliances and sensors compatible with ECHONET-Lite are connected to the home network.

4 PROTOTYPE DEVELOPMENT OF IHR

4.1 Hardware of IHR

In developing the IHR, we referred to the Magbot [11] devised by Professor Koike and modified it to enable natural conversation with the user freely. Fig. 2 shows the system configuration of the IHR developed this time. As shown in Fig. 2, RaspberryPi3 B + (hereinafter referred to as RPI) is used as the main MCU (Micro Control Unit) of IHR, and distributed processing is performed using the Arduino microcomputer as the sub MCU. RPI is in charge of the IHR main processing, performs communication processing with ECHONET-Lite compatible home appliances, etc., performs speech synthesis processing and communicates with various

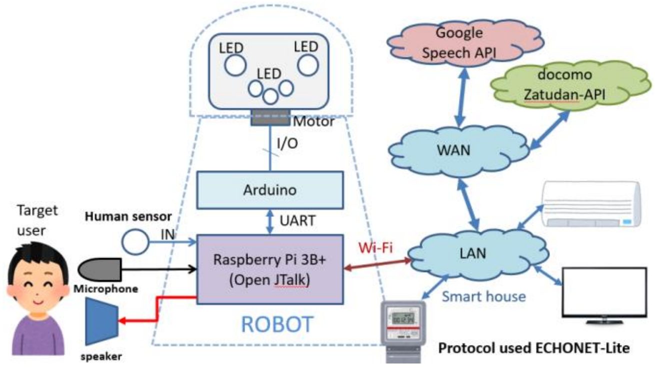


Figure 2: System of IHR

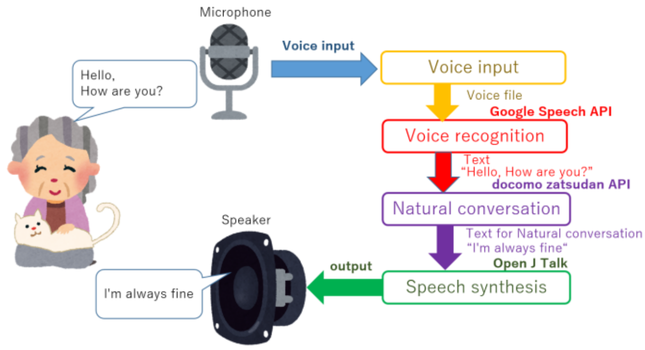


Figure 3: Flowchart of natural conversation processing of IHR

network services (Google Speech API and docomo’s chat conversation API), etc.The Arduino microcontroller performs sub-controls such as the light emission of five LEDs that appear to be the face of the IHR and the swing motion of the robot by a servomotor. Therefore, the Ardunio microcontroller controls the part of the IHR facial expression and movement. In addition, a human sensor (Model Type EKMB1301111K / Panasonic) is installed on the body of the IHR to enable the IHR to actively talk to a user near the IHR. The sensor circuit is designed to react to a person within 300 mm of detection distance. Therefore, the proposed IHR operates only when a human movement is detected, and a program that can record the result as a log in the IHR file is implemented in the RPI. The main program of this IHR RPI was developed in Python 2.7 language, and the processing program on the Arudino microcomputer side was developed in C language.

4.2 Mechanism of natural conversation

In order to allow the target person to talk naturally with the IHR, we have used Google Speech API for speech recognition service and docomo’s chat conversation API for natural conversation service for our IHR. Fig. 3 shows the flow of the natural conversation processing implemented in the IHR [9]. The IHR acquires the user's voice information with the microphone connected to the RPI, and first creates a voice input file. It performs speech recognition processing that converts speech input files into text information using the Google Speech API. The converted speech text information is sent to docomo's Zatudan-API, which creates text information containing natural speech information. This text information is speech-synthesized using Open JTalk prepared in the IHR, and the obtained voice conversation is outputted from the speaker connected to RPI. In this manner, the natural conversation between human and the IHR has been realized.

4.3 IHR facial expression

Fig. 4 shows the face of the IHR developed this time. The robot's face was made of a breadboard, five LEDs, and five resistors. We arranged two large-size LEDs side by side, protruding as the eyes, and three LEDs located below the center as the mouth. Various expressions were created by lighting and blinking each LED as shown in Fig. 4 as needed in the conversation. In practice, a translucent plastic mug was placed over the breadboard to prevent the circuit from being exposed. When talking with people, IHR lights its eyes and mouth, and produces a swinging motion with the DC servomotor so that it looks friendly and is easy to talk to.

4.4 Conversation log recording function

IHR has a function to keep a log of the talks of the IHR with the target person of the life watching. In order to distinguish between daily conversations and household appliance operation requests as the conversation content, classification numbers are provided to make it easy to extract necessary data using computer programs. The classification number is 1 for daily conversation, 2 for household appliance operation request, and 10 for no conversation. This "no conversation" is written as a log to a file inside the IHR when nobody is detected by the human sensor and there is no conversation.

4.5 Life watching method by conversation log analysis

By analyzing the log data acquired by the IHR using a program or application, the life rhythm and behavior of the watched person can be estimated to some extent. For example, a set of log collected over a long period can be treated as reference data of the living behavior of the target person. Comparing the latest data with this reference data makes it possible to detect a continuation of an abnormal situation and notify relatives and families living in remote areas of the situation by e-mail etc. For example, if the state without conversation has continued for half a day or more during the time when such conversations are normally expected, it is possible to determine that the life behavior is different from usual, so the IHR notifies the family and relatives of the situation by e-mail etc. with a message such as "Please call Mr. OO/Mrs. OO," to prompt them to contact the watching target. The e-mail transmission is performed by the main program of RPI, and realized by the standard library of python.

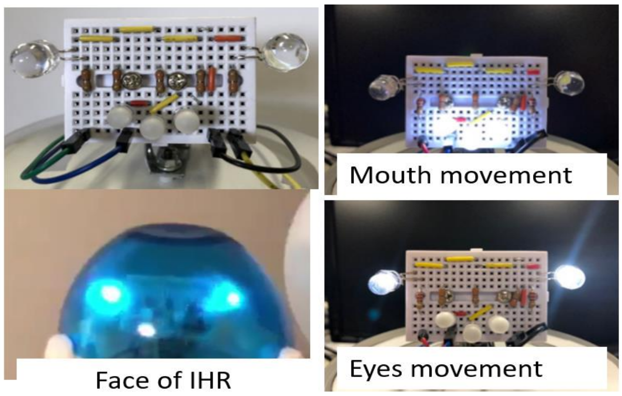


Figure 4: Facial expression of IHR

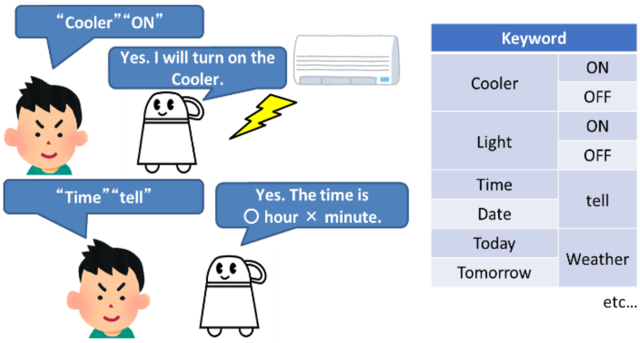


Figure 5: Appliance operation and information service by keyword of 2 words

4.6 Acquisition of living information and appliance operation in the smart house

The IHR developed this time uses ECHONET-Lite [12] as a communication protocol for acquiring living information and operating home appliances in the smart house. By communicating with smart meters and home appliances in the smart house, the IHR is able to acquire the power consumption of the entire home, acquire the ON / OFF status of various home appliances, and remotely operate home appliances. Also, when the user asks time, weather, etc., the IHR acquires necessary data in JSON format from the time [13] or weather web service site [14], and outputs it as voice from IHR. The IHR developed this time was designed to remotely turn on / off a home appliance with the words corresponding to the keyword or to provide the weather and time information when the corresponding keywords are included in the conversation. Fig. 5 shows an example of the remote control of the home appliances and information provision of the weather and time by keywords. For example, if you ask the IHR to turn on the Cooler in the room, there is a possibility that the word “Cooler” may be misrecognized.

Therefore, the IHR is programmed to remotely turn on the “Cooler” only when the two words "Cooler" and "ON" have been recognized found. If you want to hear the weather information, the IHR provides today's weather information by voice only when the two words "Today" and "Weather" have been recognized.

4.7 Software of IHR

Fig. 6 shows the flow chart of the main software implemented in the prototype IHR. A pyroelectric infrared sensor is used as the human sensor. When the motion of a person is detected by the human sensor, the robot starts talking, and when the person is absent, the IHR does not speak. Upon receiving an input of human voice through the microphone, the IHR creates a voice file and sends it to Google Speech API to convert the audio information into text information. If there is no keyword related to home appliance operations in this text information, the IHR checks whether there is a keyword related to time or weather. If there are no keywords related to time or weather, the text information is passed to docomo's chat service and converted to text information for natural conversation. The text information is converted to speech information using the speech synthesis software Open JTalk, and speech is output from the speaker to realize natural daily conversation between a human and a robot [9]. Also, if the 2-word keyword of the home appliance has been recognized in the acquired voice information, the target home appliance compatible with ECHONET-Lite connected to the network is remotely operated and the result of the operation is presented to the user. Also, if there is a keyword related to time or weather, the IHR acquires information from the dedicated site and provides it to the user by voice.

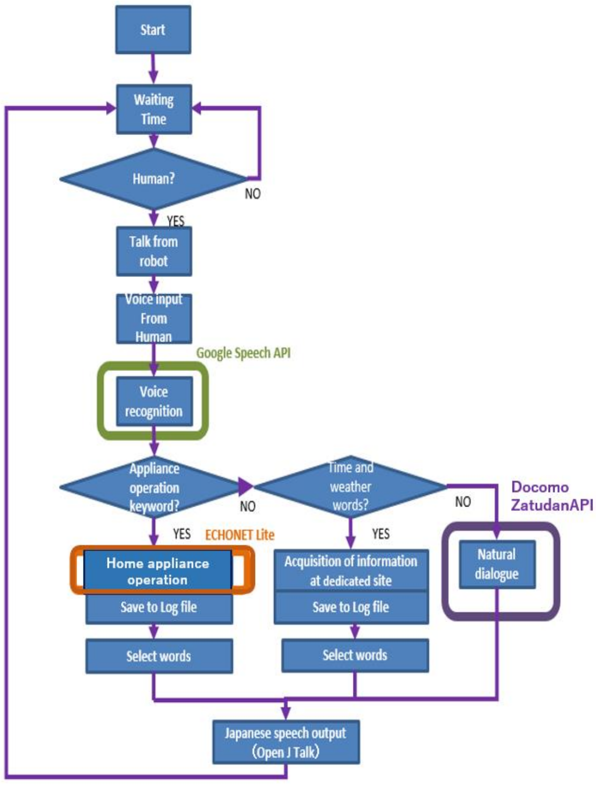


Figure 6: Main software flow chart of IHR

5 EXPERIMENT

An experiment using the IHR prototype developed this time was conducted to demonstrate how accurately the living behavior of a subject can be analyzed from the log data acquired by the IHR.

5.1 Experimental method

This demonstration experiment was conducted at E602 on the 6th floor of Kanagawa Institute of Technology C2. The subject was a 20-year-old male, and the experiment period was from May 14, 2019 to May 20 except Saturday and Sunday. Fig. 7 shows the system of the demonstration experiment, and Fig. 8 shows the layout of the demonstration experiment.

Using the IHR prototype developed this time, we asked the subject to behave as usual, and recorded information such as the subject talking with the IHR or remotely operating home appliances via the robot as log data. The recorded information was analyzed to evaluate the accuracy of human behavior estimation. A camera was installed to acquire correct data in the demonstration experiment. Fig. 9 shows a picture of the IHR and the subject during the demonstration experiment taken with the camera. Table 1 shows the equipment used in the demonstration experiment. As shown in Fig. 8, the two appliances used this time were an ECHONET-Lite compatible Cooler and lighting. The subject entered the room and had a conversation with the IHR at any time. During the demonstration experiment, the camera captured moving images of the IHR and the user's conversation situation along the time display. The correct data was created manually while recognizing the content and time of the action of the subject from the moving image of this camera. After the demonstration experiment, the reliability of the IHR log data was evaluated by comparing the log data recorded by the IHR with the correct data obtained from the camera's moving image. Also, as shown in Fig. 9, the subject sat at his desk and talked to the IHR only when necessary. The subject was asked to take action with a natural feeling, such as requesting the IHR to turn on the Cooler when it is hot, or to turn on the lighting when the room is dark. As shown in Fig. 8, the distance between the IHR and

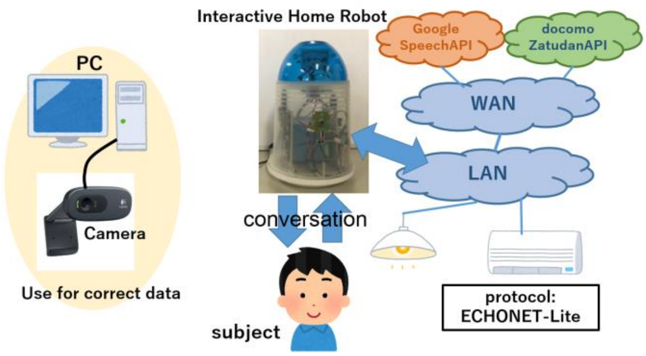


Figure 7: Experimental system

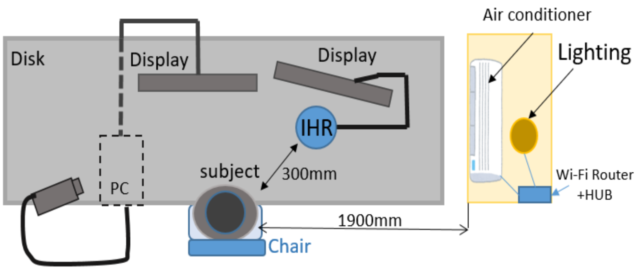


Figure 8: Layout of Experimental (Top View)

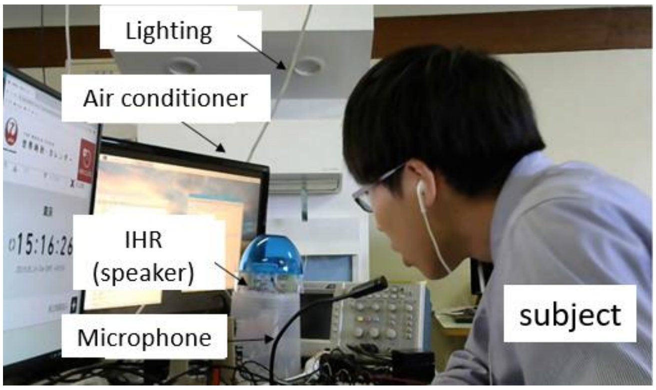


Figure 9: Demonstration experiment state from camera

Tabel.1 Equipment used

Name	Type	Manufacturer
Microphone	MM-MCUSB25	SANWA-SUPPLY
Camera	C270	Logicool
Air conditioner	LDF7N-GX53/D2	TOSHIBA
Lighting	LDF7N-GX53/D2	TOSHIBA
speaker	SP-05	ALPEX

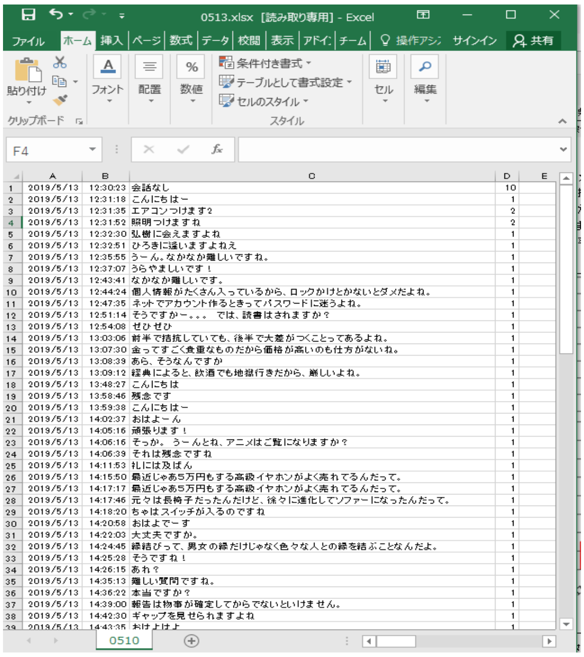


Figure10: Log data acquired by IHR (CSV format data)

the subject was 300 mm, while the distance between the Cooler or lighting and the subject was 1900 mm.

Fig. 10 shows log data acquired by the IHR. The log data was saved in CSV format in order of date, time, conversation content, and classification number. This classification number is an optional item provided so that information can be easily extracted later by a database or Excel.

The classification number is 1 for conversation and 2 for home appliance operation and 10 for no conversation. In addition, the number of conversations of the subjects and the number of times the home appliances were operated were obtained from the log data obtained by the IHR every 30 minutes. In addition, the number of conversations and the number of home appliance operations performed by the subject in 30-minute units from the images taken by the camera were used as the correct answer values. The error and error rate were obtained by equations (4.1) and (4.2) from the log data obtained by the IHR and the correct answer values obtained by the camera.

$$\text{Error} = NC_{log} - NC_{cam} \quad (4.1)$$

$$\text{Relative error} = \frac{NC_{log} - NC_{cam}}{NC_{cam}} \times 100[\%] \quad (4.2)$$

NC_{log} is the number of conversations captured in the log or number of appliance operations captured in the log. **NC_{cam}** is the number of conversations acquired by the camera or number of appliance operations acquired by the camera.

5.2 Results and discussion

Fig. 11 is a graph in which the log data obtained in the demonstration experiment from 14 May to 20 May, 2019 excluding Saturday and Sunday and the correct answer data acquired by the camera are summarized as the number of appliance operations and the number of conversations in 30 minutes.

The horizontal axis is the time axis, and the vertical axis is the number of conversations and home appliance operations. The upper graph is the graph acquired by the IHR, and the lower graph is the correct data obtained by the camera. In addition, Table 2 shows the correct answer rate of the number of conversations and the correct answer rate of the number of operations of the home appliance. As shown in Fig. 11, the number of conversations in the log acquired by the IHR is larger than in the correct answer data acquired by the camera. Next, Table 2 shows the error and error rate of the number of conversations with the user and the number of times the home appliances were operated. The number of conversations acquired by the IHR has an error of 2 to 25 times compared with the camera, and it can be seen that when converted to the error rate, it is up to 82%. A possible reason for this was that the IHR unnecessarily responded to the voices of individuals who were there other than the subject during the demonstration experiment, which led to an increase in the number of conversations. Therefore, the error in the number of conversations in Table 2 can be considered as the number of conversations other than the subject. On the other hand, the error rate of the number of home appli-

Figure 11: Number of conversations and appliance operations from 14-May-2019 to 20-May-2019 (The upper graph shows IHR log data. The lower graph shows camera data)

ance operations was zero except for the data on May 15th, indicating that log information was obtained accurately.

Next, referring to the 5-day data in Fig. 11 in chronological order, the time zone for conversation, the type of conversation, etc. are almost synchronized with the correct data taken with the camera. Therefore, IHR's log data can be used to estimate the subject's behavior as well as the camera's correct data. The working hours of the subject this time are 11:00 to 16:00, but it is possible to estimate the early departure and overtime hours etc. depending on the day of the week by combining with the subjects' action memos in Fig. 11. It will be also possible to accurately estimate the time when the subject left for lunch. With 14-May and 16-May data, it is possible to clearly read the time spent leaving for lunch because there is no conversation with the IHR or the operation of home appliances, but with 15-May data it cannot be judged whether the subjected was absent, because there is a conversation recorded. As we summarized the data in units of 30 minutes this time, an error of ± 30 minutes may occur depending on the time zone as compared with the correct data as shown in the graph of 17-May in Fig. 11. For this reason, it is difficult to estimate the exact leaving time etc. when the subject leaves for a short time at lunch etc. in the analysis in units of 30 minutes. Table 3 compares the actual working hours with the working hours estimated from the log data acquired by the IHR. In this result, it is within about ± 15 minutes, but it can be expected that the error will deviate by about 30 minutes if the room entry time is before 11:00, for example. Therefore, the accuracy of the living activity estimation using the proposed IHR depends on the aggregation interval of logs. In other words, if the log data is aggregated every 30 minutes, it is thought that the daily activities such as entering, leaving and working hours will have an error of about ± 30 minutes. Also, since it is possible to determine whether the conversation content with the IHR is a daily conversation or an operation of a home appliance, it is considered that it is possible to extract a subject-specific life behavior pattern from these kinds of information. Next, in order to analyze the subject's specific behavior, Fig. 12 shows the number of conversations with the IHR and the number of home appliance operations in each time zone from 5/14 to 5/20. From this graph, it seems to be possible to find out the time zone where there are many conversations with the IHR and the time zone where there are many times to operate home appliances. However, it is actually difficult to locate such time zones at a glance in those graphs. Therefore, as shown in Fig. 13, we tried the average value of the number of conversations with the IHR and the number of operations of home appliances for each time slot from 5/14 to 5/20. The red line in Fig. 13 shows the average value of the number of conversations and the number of operations of home appliances from 11:00 to 16:00. The time zone where the bar exceeded the red line indicating the average value in Fig. 13 was considered to be the time zone where the subject talked with the IHR and operated the home appliance frequently.

As shown in Fig. 13, when the time zone exceeding the red line is extracted, the time zone in which the subject talks frequently with the IHR is 11:30, 12:30, 14:00, 15:00, 15:30. By comparing this extraction result with that in Fig. 12, it is possible to extract a high frequency time zone as a whole, although it includes a day without conversation and home appliance operation. In this way, subject-specific behavior can be analyzed from log data analysis acquired by the IHR. Moreover, it can be used as a judgment standard of the life watching for determining whether the life is ongoing as usual by comparing the data obtained through this behavior analysis and the log data of the latest value.

Table.2 Comparison of Error of log data vs camera (correct data)

Date	Number of conversation				Number of appliance operation			
	Camera	Log data	Error	Relative error	Camera	Log data	Error	Relative error (%)
14-May-19	51	62	11	22%	18	18	0	0%
15-May-19	58	80	2	51%	14	13	-1	-7%
16-May-19	58	76	18	31%	12	12	0	0%
17-May-19	11	20	9	82%	12	12	0	0%
20-May-19	41	66	25	61%	10	10	0	0%

Table.3 Comparison of actual working time vs working time obtained from log data

Date	Actual working time	Working time obtained from log data
14-May-19	11:12～15:38	11:00～15:59
15-May-19	11:12～16:16	11:00～16:29
16-May-19	11:06～15:53	11:00～15:59
17-May-19	11:10～15:20	11:00～15:29
20-May-19	11:08～15:41	11:00～15:59

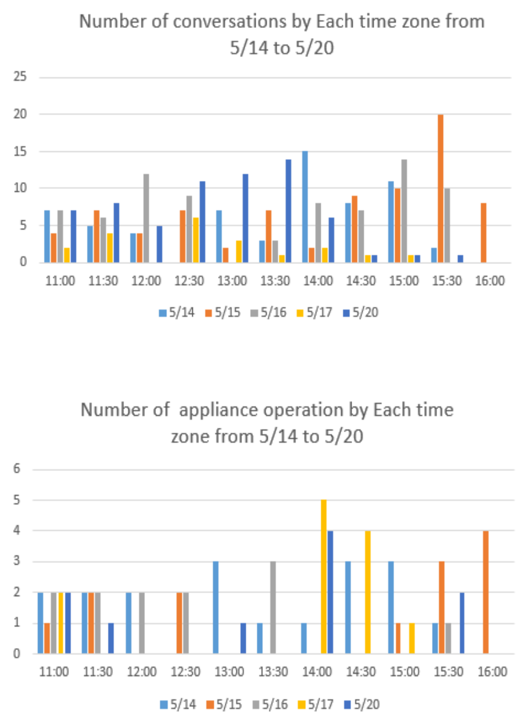


Figure 12: Number of conversations and appliance operations by each time zone from 14-May-2019 to 20-May-2019

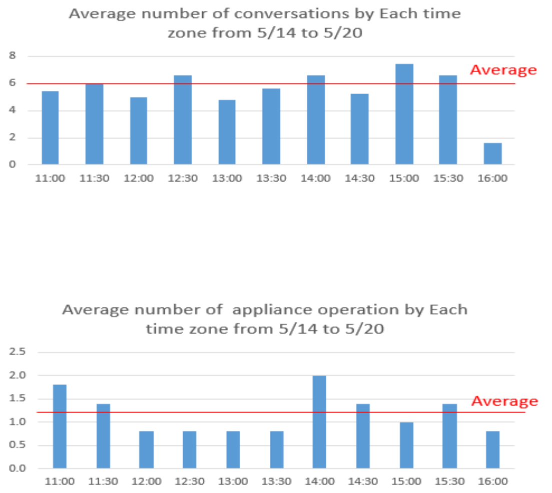


Figure 13: Average number of conversations and appliance operations by each time zone from 14-May-2019 to 20-May-2019

6 QUESTIONNAIRE EVALUATION

The IHR proposed by this study is a technology that does not have a camera at all, but performs natural conversation based on voice information from users, voice control of home appliances, and watching life by collecting log data. There is also a technique using AI speakers [15]-[16] as a related technology for voice interactive system without cameras. Most of the commercially available AI speakers have an inorganic design, so it seems that there are a certain number of people who find it difficult to talk to. Therefore, we aimed to develop an IHR without cameras that is not a simple AI speaker but can also express non-verbal information such as facial expressions and movements [17] like human beings in a rich manner and that can perform dialogue by voice.

We thought that IHR would be useful in a life watching service if it could increase the opportunity of conversation with the user. Therefore, in order to evaluate whether the design of the IHR proposed in this study makes conversation easier than AI speakers, we conducted a comparative evaluation questionnaire using the IHR vs AI speaker. As for the implementation method, we asked the public to see a demonstration of the IHR prototype we developed and an AI speaker. A questionnaire was conducted on 29 men and women in their teens and 60s. Fig. 14 shows the age and gender of the questionnaire respondents. By age group, 20s and less than 20s occupy 60% of the whole. Fig. 15 shows the answer to the question, “Which of the IHR or AI speakers did you think is easier to talk to?” As a result, about 80% of the respondents answered that the IHR we developed was easier to talk to than AI speakers.

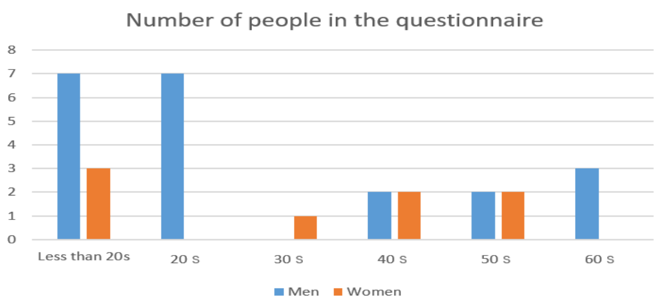


Figure 14: Age and gender of questionnaire respondents

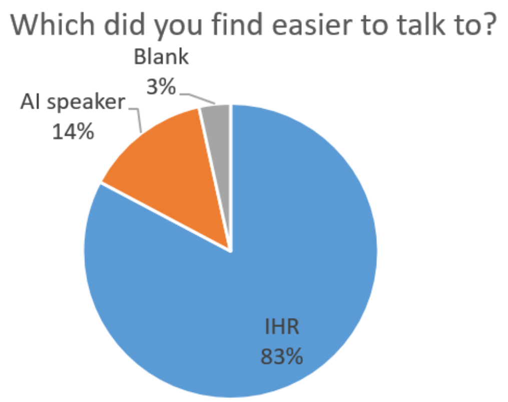


Figure 15: Questionnaire on comparing IHR and AI speaker

7 CONCLUSION

In this paper, we proposed a service that uses an IHR to monitor the life of a target person without using a camera at all for those who want to monitor the lives of their target persons including the elderly, who live away from their families. Specifically, by increasing the chances of conversation between the IHR and the target person, it was possible to collect a large amount of log information, so that the accuracy of the target person's behavior estimation was improved, and as a result, we thought that it could be used as a life watching service. The prototype of the IHR was developed, and the reliability of the data acquired by the log file was evaluated by comparing the log file acquired by the IHR with the result of the moving image acquired by the camera. As a result, it became clear that when there were multiple people other than the subjects, the voices of those other people were picked up, and the number of conversations counted in the log data of the IHR was larger than the actual number. However, it was found that it was possible to estimate the behavior at which time the subject talks with the IHR well and operates home appliances. In addition, it was found that the estimation accuracy of the living activities such as entering, leaving, and working hours depends on the aggregation interval of IHR log data. In addition to this, it is possible to determine whether the contents of the log information acquired by the IHR is a daily conversation or an operation of a home appliance, so it is possible to extract a subject-specific life pattern from these kinds of information.

By the way, there is AI speakers as a related technology of the voice interactive system without a camera. Most of the commercially available AI speakers have an inorganic design, so it seems that a certain number of people will find it difficult to talk to. Therefore, as a means for increasing opportunities for conversations with users, we developed an IHR that can express richly non-verbal information such as facial expressions and movements, but also human speech. This time, we conducted a questionnaire on comparative evaluation of the IHR and the AI speaker to evaluate whether the design of the IHR we developed was easier to talk to than an AI speaker. As a result, according to the answer to the question, “Which of the IHR and AI speakers did you think is easier to talk to?”, approximately 80% of the respondents answered that the IHR was easier to talk to. Therefore, the proposed IHR is more effective as a life watching service for acquiring conversation log data than commercially available AI speakers because the IHR can more easily induce users to talk to.

8 FUTURE WORKS

It is thought that human conversations can be made to go smoothly and useful information can be easily extracted from users by a robot that makes appropriate gestures such as agreeable responses and nodding in the conversation [18]. In the future, we would like to study how to control the timing of the swinging, nodding and other motions of the IHR when talking with users, and to conduct smooth conversations with the users. In addition, we would like to consider the use of displays etc. in order to increase variations in the IHR expressions.

REFERENCES

[1] “About the present conditions and future way of watching of elderly people spreading-Mizuho Information & Research Institute (In Japanese)”, <https://www.-mizuhoir.co.jp/publication/report/2017/pdf/mhir14 kourei.pdf>, [referred April.2019]

[2]“Elderly person watching business | Akiruno city (In Japanese)”, <http://www.city.akiruno.tokyo.jp/ 000000-4255.html>, [referred April.2019].

[3] “Tapia ROBOT”, <https://mjirobotics.co.jp>, (referred April.2019).

[4] “Kibiro ROBOT (In Japanese)”, <https://www.softbank-selec-tion.jp/cart/ProductDetail.aspx?sku=4589723160100>, [referred April.2019].

[5] "Soft Bank Pepper View (In Japanese)", <https://www.-soft-bank.jp/robot/consumer/products/ap**plic**ation/ncp/view/>, [referred Dec.2019].

[6] "RoBoHoN (In Japanese)", <https://robohon.com/apps-/app31.php>, [referred Dec.2019].

[7] LOVOT, <https://lovot.life/features/>, [referred Dec.-2019].

[8] T. Sugihara, T. Fujinami, and K. Nakagawa, ”Effects of Cameras and Monitors on Caregivers’ Work Stress in the Group Home (In Japanese)”, The Institute of Electronics, Information and Communication Engineers, IEICE-107, No. 555, WIT 2007-100, pp.57-62, (2008).

[9] Tai Ritu, “Let's make a natural conversation robot! Artificial intelligence made with RaspberryPi and Arduino(In Japanese)”, SYUWA SYSTEM, (2016).

[10] S. Mori, S. Yoshida, " ECHONET device adapter (Special feature / net home appliance: home network, terminal) (In Japanese)", Matsushita technical journal 47(1), pp.79-83, (2001).

[11] S. Koike, “Talking Robot [Mug bot]-electronic work with Raspberry PI and Arduino-(In Japanese)”, RIC TELECOM. (2016).

[12] “ECHONET Consortium”, <http://www.echonet.gr.jp-/>, [referred April.2019].

[13] “Sample to get time from NTP in JSON format(In Japanese)”, <http://etc9.hatenablog.com/entry/20110325/ 1301073502>, [referred April.2019].

[14] “ weather hacks(In Japanese)”, < http://weather. Livedoor.com /weather_hacks/webservice>, [referred April-.2019].

[15] ” Google Home-Smart Speaker & Home Assistant-Google Store”, <https://store.google.com/jp/product /google_home>, [referred April.2019].

[16] LINE Clova suite, <https://clova.line.me/>, [referred April.2019].

[17] M. Yokoyama *et al.*, “Controlling Nonverbal Information in Speaker- changing for Spoken Dialogue Interface of Humanoid Robot”, Journal of Information Pro-

cessing Society, Vol.40, No.2, (1999).

[18] T. Misu *et al.*, "Toward Construction of Spoken Dialogue System that Evokes Users' Spontaneous Backchannels (In Japanese)", The transactions of the Institute of Electronics, Information and Communication Engineers. A, Vol.J95-A, No.1, pp.16-26, (2012).

(Received November 6, 2019)
(Revised February 5, 2020)

Keiichi Abe received Ph. D. in informatics from Shizuoka University in 2012. He has been an associate professor of the Faculty of Creative Engineeri-ng, Kanagawa Institute of Technology since 2016. His research interests include smart system using IOT technology, Interactive system using Robot and CG, and Victims Information Management System at Large Scale Disaster, HEMS. He is a member of Information Processing Society of Japan, the institute of Electronics, Information and Communication Engineers of Japan and a member of IEEE.

Hiroki Fuse received his B.I. degree from Kanagawa Institute of Technology and entered the Graduate School of Kanagawa Institute of Technology in 2019. His research subjects are HEMS and AI robots.

Makiko Okumura received her B.E. degree from Keio University in 1982, and Ph.D. degree in electrical engineering from Tohoku University in 1994. She joined Toshiba R & D Center, Kawasaki, Japan, and transferred to Kanagawa Insti-tute of Technology in 2000. She has been working on the development of analog circuit simulators and now studying consumer electronics. She is a member of the institute of Electronics, Information and Communication Engineers of Japan and a member of IEEE.

Masao Isshiki Ph.D Prof. Professor of Kanagawa Institute of Technology since 2011. Director of Smart House Rese-arch Center since 2012. JSCA HEMS-Task Force chairman since 2011 and also working for many chairmen of METI, MIC and MLIT. Project professor of Keio University since 2009 to 2015 responsible for W3C/keio site manager. Fellow of ECHONET Consortium. Masao holds a Ph.D. degree in heat transfer engineering from the Tokyo University of Agriculture and Technology, and a master's degree in mechanical engineering from the University of Tokyo Institute of Technology. He currently lives in Tokyo, Japan.

Industrial Paper

Optimum Method of Spindle control for Multi-tasking Machine

Kazuhiko Tsutsui[*†], Jun Sawaki[*], Naoya Chujo[††], and Tadanori Mizuno[††]

[*]Mitsubishi Electric Corporation
[†]Graduate School, Aichi Institute of technology, Japan
[††]Aichi Institute of technology, Japan
Tsutsui.Kazuhiko@ay.MitsubishiElectric.co.jp

Abstract - Recent machine tools control quite an increased number of axes for multiple cutting processes. Furthermore, more advanced functions have become necessary for machine tools: sharing data with the host system or using AI to improve machining accuracy, supporting collaboration with robots, and so forth. Therefore, one of the most critical issues for CNC controllers of machine tools is to enhance the CPU performance.

To address this issue, we have established distributed control of servo/spindle position/speed feedback loops in the power amplifier independently from the CNC controller.

In the past, only speed control was necessary for spindles. However, switching between position control and speed control has become necessary to allow for various machining modes for the increasing number of systems with advanced spindle functions. The time and complexity required for switching control loops are bottlenecks for improvement in productivity of machine tools.

In this research, we propose an approach to perform consecutive position control for spindles.

Keywords: Spindle motor control, Multi-tasking machine, Consecutive position control, Tracking delay compensation

1 INTRODUCTION

CNC equipment consist of the CNC controller, actuators, amplifiers, and detectors. To run the user's machining program, the CNC controller generates position/speed commands sent to the feed axes and spindles (tools) that constitute the machine coordinate system. Actuators such as servo motors and spindle motors are used to drive the feed axes and spindles. Servo amplifiers and spindle amplifiers are used to supply variable power to the actuators. Detectors such as encoders and linear scales are used to feed back the position/speed of operating parts of machines or motors.

The CNC controller, i.e. computerized numerical control controller, is the brain that controls the entire operation of machine tools. The CNC controller sequentially analyzes a G-code program that include description of the machining path, feed speed, and number of rotations of the tool, and generates commands of the travel distance per unit time for each feed axis. The CNC controller also generates speed commands for spindle motors according to the cutting conditions. The CNC controller is the open-loop control in which the control action is not dependent on feedback from the path or position of the actual machine, and the control to ensure correct positions is not performed.

Meanwhile, machine tools are required to have performance to minimize the reaction forces (disturbances) during heavy cutting and the effects of friction and stiffness in machines to achieve micrometer or nanometer level cutting accuracy in as short time as possible. High-speed, high-accuracy tracking of position/speed commands generated by the controller is critical for the actual machines and motors. Hence, high-response feedback loop control is one of the most essential factors in developing CNC equipment.

There are two ways to establish the feedback loop: one is to establish the loop of the system as a whole using the CNC controller, servo/spindle amplifiers, motors, and detectors, and the other is to establish the loop without the CNC controller. The design philosophy differs between the two models because the performance required for each component and the interface between components are different for each model.

While servo/spindle amplifiers are connected one-on-one to motors with power lines, serial communication through one network interface is used between the CNC controller and multiple amplifiers. When the feedback loop control is processed by the CNC controller, the communication cycle time becomes longer in the control loop with increasing difficulty in seeking high-level response performance due to the massive amount of exchanged data. For that reason, we developed CNC equipment in which servo amplifiers perform position/speed feedback loop control for servo motors that move the feed axes.

On the other hand, the spindle motor control has been more complex to achieve high productivity. The servo axes don't have to change the control mode. It means that the servo axes are controlled by position loop consecutively. On the other hand, spindle motor control is required to use properly the speed loop control and position loop control. Although spindle amplifier is only needed to control the spindle motor speed and output power on the process of turning and milling, it is needed to control the angle position of cutting tool and to synchronize with other servo axes on the process of tapping C-axis mode for multi-tasking machine.

In the former systems, the mechanical angle of spindles had to be adjusted every time when the machining mode was changed. As it was necessary to recalculate a designated position, it was not possible to reduce the machining time.

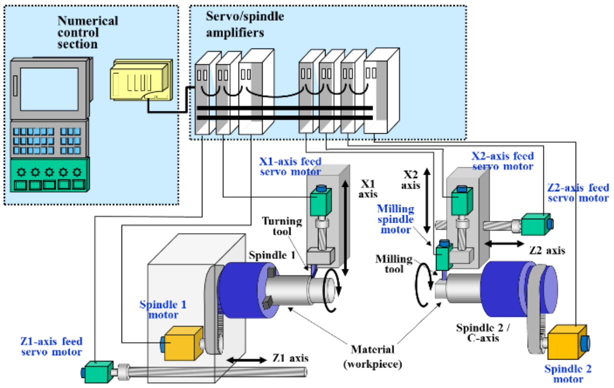


Figure 1: System configuration example of multi-tasking machine

Besides, speed control operations were more susceptible to disturbances during cutting than position control operations.

Hence, we propose an approach to perform consecutive position control in spindle amplifiers for spindle motors regardless of the machining mode.

This approach reduces extra sequential control to change the machining mode, ensures robustness against disturbances, and enables establishment of the spindle control system for high-accuracy machining.

2 RELATED TECHNOLOGIES

2.1 Basic Configuration of CNC Equipment

Multi-tasking machines such as the one shown in Fig. 1 are increasing in recent years.

Each axes are driven by the servo motors connected to the ball screw in the machine. The tool used for cutting is attached to the spindle head and driven by the spindle motor. The CNC controller and the amplifiers which is used to control the speed and position in addition to power supply for the servo and the spindle motors are all installed in an electrical enclosure.

The path of the cutting tool affects accuracy of the workpiece directly. Therefore, it is significant for servo motors to ensure robust command tracking for position and feed speed control commands with minimal errors by suppressing load disturbances such as cutting reaction force or friction in machines. It is also important to keep the synchronous accuracy and the same response level for X, Y, and Z axes. Otherwise, the tool may follow the wrong path not intended by the CNC controller, resulting in unsatisfactory machining accuracy.

2.2 Basic Servo/Spindle Control Architecture (Distributed Control)

In some cases, positions and speeds of servo/spindle motors are controlled by the CNC controller showed by Fig. 2. By using the CNC controller, it becomes easier to develop equipment with synchronous control of multiple axes or to perform compensation for multiple axes. However, this architecture has a drawback: the controller is connected to the amplifiers

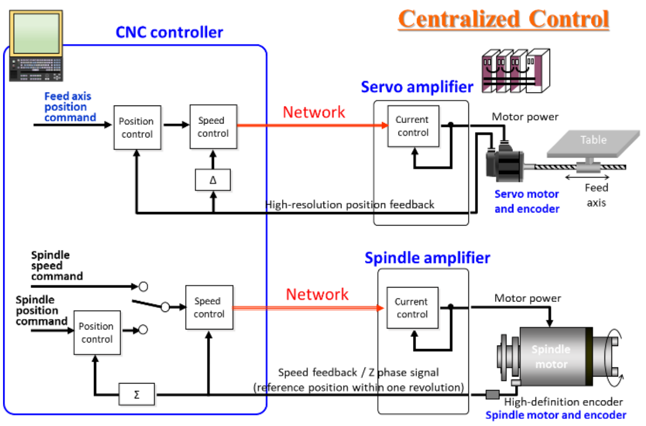


Figure 2: Centralized control of Servo/Spindle.

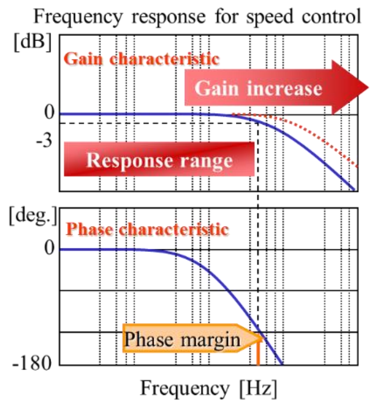


Figure 3: Response level of the speed control loop.

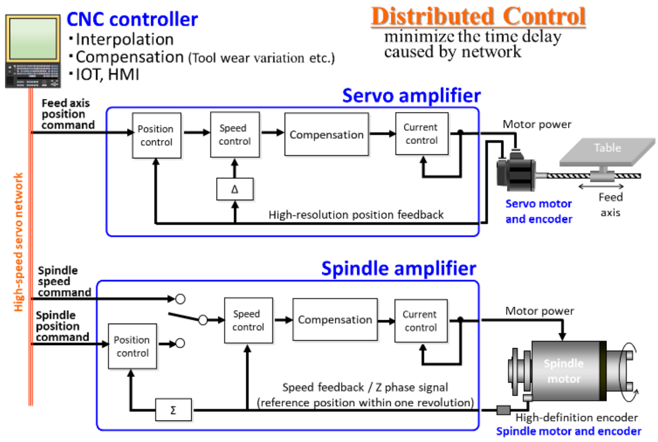


Figure 4: Distributed control of Servo/Spindle

over network, and it is not possible to allow high-speed control loops due to the dead time.

Figure 3 shows the frequency response characteristics that affect disturbance suppression and command tracking for servo and spindle speed control. Generally speaking, it is necessary to increase the loop gain and broaden the response range to improve disturbance suppression and command tracking. However, the phase margin is reduced as the dead time increases, and the feedback loop becomes unstable and

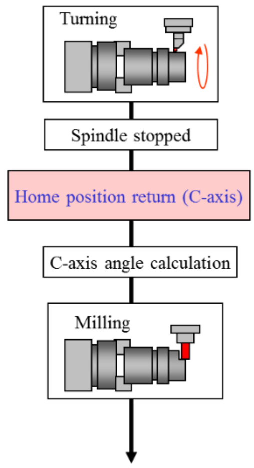


Figure 5: Switching from turning to milling (C-axis).

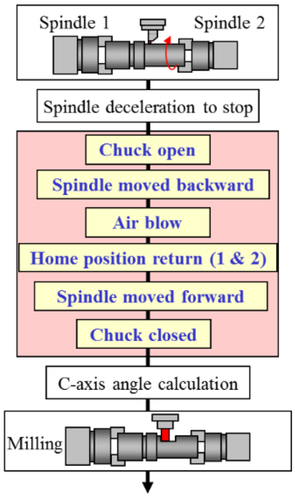


Figure 6: Switching from spindle-linked turning to C-axis-linked milling

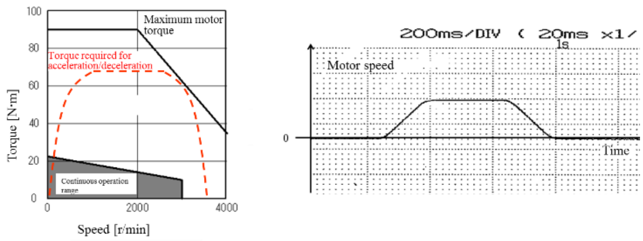


Figure 7: Normal acceleration/deceleration waveform of position loop control.

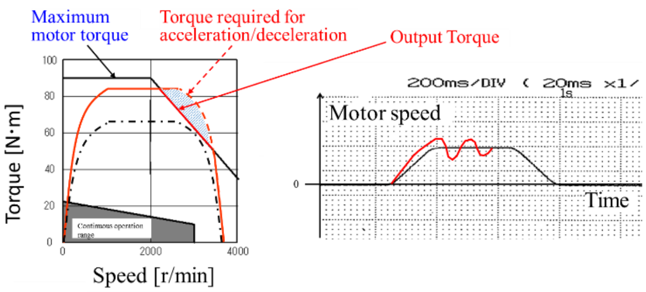


Figure 8: Abnormal acceleration/deceleration waveform of position loop control.

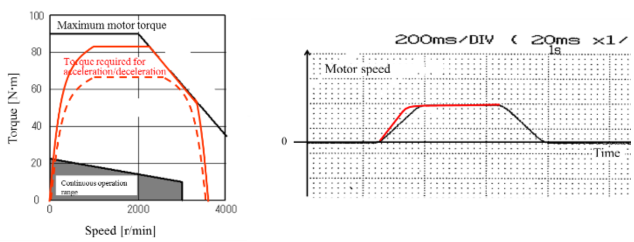


Figure 9: Acceleration/deceleration waveform of speed loop control.

causes oscillation. Furthermore, the load applied to the controller increases along with increase in the total number of control axes in recent machine tools.

In order to address the above-mentioned issues, we adopt the architecture shown in Fig. 4 for servo and spindle control loops. The position feedback and speed loops are implemented inside the amplifiers (distributed control) and the network communication that involves a substantial amount of dead time is established outside of the control loops.

2.3 Issues Identified with Spindle Control

Multi-tasking machines that perform multiple operations prevail in recent years. Position control and speed control for the spindle motors are switched in multi-tasking machines according to the machining program (refer to the spindle amplifier section in Fig. 4). In addition to operations such as turning, milling, or drilling by spindle rotation control, it is necessary for multi-tasking machines to perform operations such as C-axis control or synchronous tapping (C-axis control: synchronous position control with the servo axes).

2.3.1 Time required for switching the control methods

When the spindle requires only speed control in the machining mode such as turning, neither the CNC controller nor the spindle amplifier handles position data such as tool tip angles.

In this case, to change the control mode from turning to milling (C-axis control), it is necessary to stop the spindle once, perform home position return, and detect the tool tip angle as shown in Fig. 5.

Another example is shown in Fig. 6. To change the machining mode from turning one workpiece chucked by two spindles to milling (C-axis control), home position return is required for both spindles. The chuck of each spindle is opened to release the workpiece and home position return is performed one after another for two spindles. Then, the chuck is closed and the workpiece is held by the spindles again, the C-axis angle is calculated, and milling is started. It takes several to tens of seconds each time to switch between two modes. Therefore, repeating switching has significant impact on productivity.

Moreover, the switching process sequence control involves complex programming.

2.3.2 Consecutive position control in spindle amplifier

The following section describes the issues involved with consecutive position control in spindle amplifiers in the same way as in servo amplifiers.

(1) Acceleration and deceleration time

The tool attached to the spindle is frequently changed according to the machining conditions. The spindle motor must be decelerated and stopped once every time the tool is changed, and accelerated back to the maximum speed after the tool change. The number of spindle rotations per minute ranges from several thousands to tens of thousands and it takes several to tens of seconds each time to accelerate or decelerate the spindle. One of the requirements demanded for spindle motor control is to reduce the acceleration and deceleration time as much as possible.

Figure 7 is a diagram to show the maximum motor torque. The horizontal axis of the graph represents the motor speed. In general, the maximum torque generated by motors is reduced due to saturation of the induced voltage in the high-speed region even though the maximum current value of the amplifier is not changed.

Under these circumstances, when acceleration/deceleration is performed with the unchanged position loop, it is necessary to build a program with a time constant that would not allow the required torque to exceed the maximum motor torque as shown by the dashed line in Fig. 7.

If a program is built with a time constant that would require the torque exceeding the maximum motor torque as shown in Fig. 8, position control to follow the command is not achieved. Then, position errors are accumulated to cause overshooting, resulting in vibration or mechanical impact.

On the other hand, acceleration to the maximum speed or deceleration to a stop in a shortest possible time is desirable in the machining mode without position control as mentioned above. Therefore, the utmost motor torque must be applied for acceleration/deceleration.

As shown by the solid line in Fig. 9, in the case of speed loop control, the utmost motor torque can be applied for acceleration/deceleration without overshooting since position deviations are not accumulated even though a program is built with a time constant that would allow the torque require for acceleration/deceleration to exceed the maximum motor torque. However, it is not possible to control the cutting tool's angle position on the spindle since position control is not performed in this case. For that reason, position control and speed control are switched according to the machining mode for spindle control in preceding systems.

(2) Spindle encoder

Spindle motors require speeds exceeding 30,000 r/min (revolutions per minute). Besides, cutting fluid or other liquids may enter from the spindle end. For the reason of durability and speed, optical encoders for servo motors cannot be used for spindle motors. Therefore, magnetic detection type encoders are used in general.

In the preceding systems, it is difficult for magnetic encoders to keep tracking the spindle end angle accurately without deviation in the range over several tens of thousands revolutions per minutes. To switch to position control for C-axis control or the like, installation of an encoder for position control is required in addition to the encoder for speed detection.

However, we modified the spindle encoder which is able to use for both speed control and C-axis control without additional encoder.

We could achieve this new encoder by adopting mainly two attempt.

One is detection method. We adopted high accuracy cutting method of gear and also high resolution Magneto Resistive Sensor inside the encoder.

Another one is data transmission method. We had made the position data in amplifier by interpolating AD convertor data which was converted from analog output signal through the cable from encoder before. But it was limited by the analog

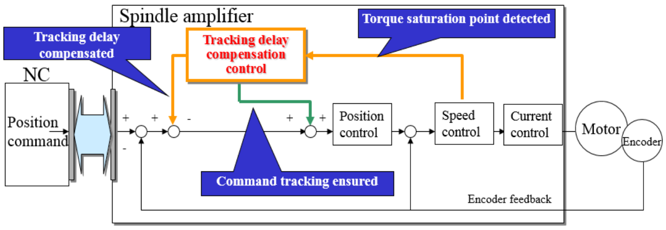


Figure 10: Tracking delay compensation control for consecutive position control.

output frequency and easily effected by noise. If we wanted to obtain high resolution data in the condition of high speed, analog frequency was far beyond the frequency of AD convertor performance. So we developed the serial communication interface which is able to send the high resolution data without any effect of noise.

By using these technologies, we could create the spindle end position data accurately without deviation even if the number of revolutions exceeds several tens of thousands per minute. The next section describes the control method to achieve consecutive position control with this spindle encoder.

3 PROPOSED METHOD

3.1 Outline of Tracking Delay Compensation Control

As mentioned above, a major issue for consecutive position control is overshooting. Overshooting occurs when position control to follow the command is not achieved, resulting in position deviation, during acceleration/deceleration with the utmost motor torque. Thus, we propose the use of spindle amplifier with the tracking delay compensation control as shown in Fig. 10.

Thus, consecutive position control can be performed in the spindle amplifier regardless of the machining mode. Consequently, the machining mode switching time can be reduced and disturbances during cutting can be suppressed further more.

3.2 Tracking Delay Compensation Control

Figure 11 is the block diagram to show the details of the control in the spindle amplifier which performs the tracking delay compensation control we propose.

(1) CNC controller

A CNC controller 1 gives speed command signals as output to the spindle during ordinary turning, machining, or milling operations. For tracking delay compensation control, speed command signals are integrated to generate position commands. On the other hand, position command signals are out-

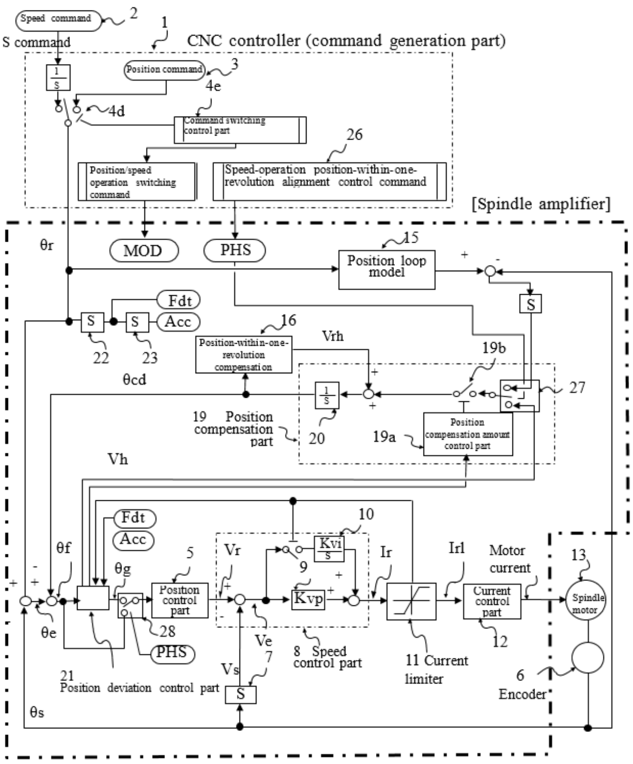


Figure 11: Tracking delay compensation.

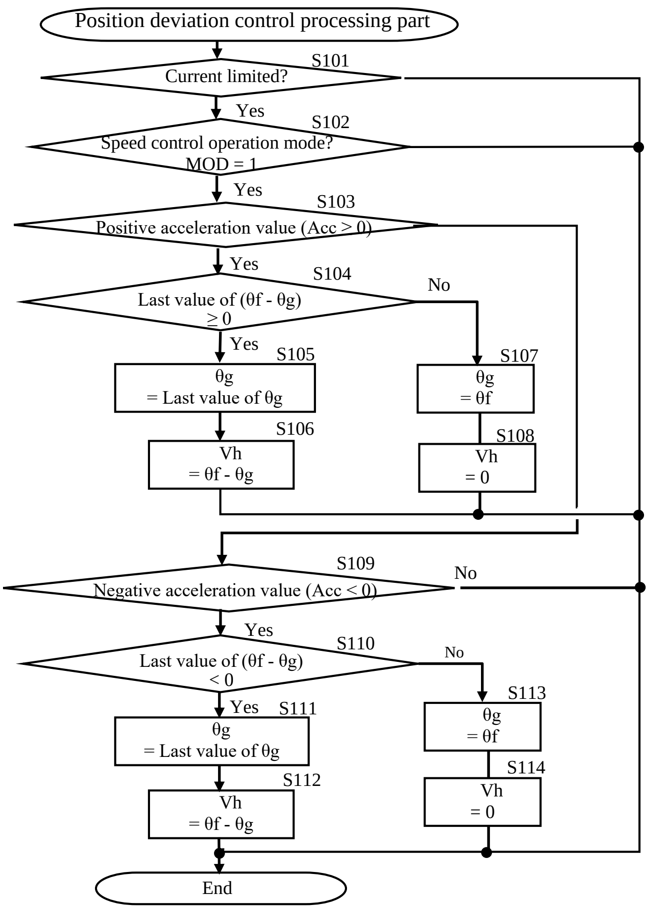


Figure 12: Operation flowchart in the position.

put according to the spindle-end rotation angle during synchronous tapping or C-axis control. In the CNC controller 1, the position of a switch 4d is changed according to the change in the machining mode, and a position command signal θr for the spindle amplifier is generated. Meanwhile, a command switching control part 4e outputs a position/speed operation switching command MOD, which contains data to show whether position control operation or speed control operation is performed at the time of switching.

(2) Position/Speed/Current control part

Position, speed and current control parts are used typical method. However, current limiter rolls as important functions.

The current control part 12 controls a current of a motor 13 based on a current limit value output by the current limiter 11.

While a current is limited by the current limiter 11, the current limiter 11 outputs a current limit command Il to the integral speed controller 10 to stop the integral action. The integral speed controller 10 stops the integral action to avoid unnecessary integral action of the speed deviation signal Ve while the current is limited, and to prevent overshooting from occurring for the speed command value after the current limit is released. The current limiter 11 outputs the current limit command Il also to the position deviation control part 21.

(3) Position deviation control part

A position deviation signal θe represents the difference between the position command signal θr output by the CNC controller and a position signal θs generated by a motor-end or spindle-end encoder. In typical position loop control, θe is input to position control part directly. However, this proposed method uses position deviation control part 21.

This position deviation control part 21 uses below input signals.

- θf: Deviation input signal created by θe and θcd which is generated by command compensation part 19.
- θr: Which is converted to a command speed signal Fdt by a differentiator 22.
- Acc: Command acceleration signal.
- MOD: Position/speed operation switching command.
- Fdt: Command speed signal.

After the predetermined operation is completed, the position deviation control part 21 outputs a deviation limiting output value θg to a position control part 5.

Figure 12 shows the flowchart of the operation carried out by the position deviation control part 21. In the position deviation control part 21, while the current limit command Il is given (S101) and the operation mode is speed control during which the position/speed operation switching command (MOD) does not require absolute position tracking (S102), when the command acceleration signal Acc has a positive value (Acc ≥ 0) (S103) and the deviation input signal θf increases in the positive direction (S104), it is assumed that the position deviation control section output value θg equals to the last value of θg (S105), and that the input/output deviation signal Vh has a (θf - θg) value (S106). Meanwhile, in order to send a command to the position compensation amount control part 19a in the position compensation part 19 to turn on the switch 19b (Fig. 11), the input/output deviation signal Vh output from the position deviation control part 21 is accumulated by the compensation position deviation amount θcd. The compensation position deviation amount θcd is subtracted from the position deviation signal θe to generate the deviation input signal θf.

When the command acceleration signal Acc has a negative value (Acc < 0) (S109) and the deviation input signal θf increases in the negative direction (S110), it is assumed that the output value θg equals to the last value of θg (S111), and that the input/output deviation signal Vh has a (θf - θg) value (S112).

When the input/output deviation signal Vh output from the position deviation control part 21 has a (θf - θg) value, the position deviation amount control part 19a of the position compensation part 19 shown in Fig. 11 turns on the switch 19b. Then, the input/output deviation signal Vh of the position deviation control part 21 is accumulated by the integrator 20 to output the compensation position deviation amount θcd. The compensation position deviation amount θcd is subtracted from the position deviation signal θe. The input/output deviation signal Vh (base of θcd) represents the difference between θcd and the signal θe that represents the difference between the position command signal θr and the motor 13 position signal θs. Therefore, when the current command value reaches the limit in the motor controller due to saturation of the motor output voltage during acceleration/deceleration of the motor, insufficient torque for the commanded acceleration, or other reason, it is possible to prevent the gap between the speed command calculation signal Vr (output from the position control part 5) and the actual motor speed Vs from increasing even if the position command signal θr value (converted from the speed command signal) is too large while the speed command 2 is selected with the switch 4d. Consequently, it is possible to reduce the delay in returning to position control when the current limit is removed.

(4) Position compensation part

A position-within-one-revolution compensation control part 16 is present on the output side of the position compensation part 19. The position-within-one-revolution compensation control part 16 normalizes the compensation position deviation amount θcd output by the position compensation part 19, and calculates a position-within-one-motor-revolution deviation signal Vrh (the data for two or more revolutions are discarded and the motor deviation amount within one revolution (difference between the position command and the actual motor position) is calculated). When it is confirmed that the current command is within the current limit after the current limit is removed, a position-within-one-revolution compensation amount Vrh is calculated so that the position within one motor revolution deviation becomes zero. Vrh is added to the integrator 20 in the position compensation part 19.

Figure 13 (b) shows the acceleration/deceleration slope profiles of the position-within-one-revolution deviation signal Vrh. The horizontal axis represents the speed, the vertical axis the slope (acceleration), the solid line the case 1 profile, the chain line the case 2 profile, and the dash-dotted line the motor torque characteristics.

As shown in Fig. 13 (a), the maximum speed in the Vrh compensation pattern is determined at a certain ratio γ (e.g. 10%) of the speed feedback value at the start point of actual compensation. The slope profile (acceleration) to the maximum speed in the Vrh compensation pattern is determined according to the motor output torque characteristics as shown by the dash-dotted line in Fig. 13 (b). The slope profile line may be straight as shown by the solid line (case 1) in Fig. 13 (b) by leaving a margin for the motor output torque characteristics if the controller's processing time and the memory capacity allow. Otherwise, the profile may be set in stages as shown by the chain line (case 2). Thus, the compensation is performed in a stable and fast way for the position within one revolution after the current limit is removed.

(5) In case of speed control mode

In the CNC controller 1 in Fig. 11, a command PHS, which shows that compensation is not required for the position within one motor revolution, is input to the motor controller from a position-within-one-revolution alignment control part 26 in speed operation mode that does not require absolute position tracking. The command PHS is transmitted to a switch 27 in the position compensation part 19 and a switch 28 between the position deviation control part 21 and the position control part 5. The switch 27 usually connects to the input/output deviation signal Vh input from the position deviation control part 21. However, when the command PHS is input, the switch 27 is switched to connect to the opposite side of the Vh signal. On the opposite side, the switch connects to the differential value input of the deviation of the actual motor 13 position measured by the encoder 6 from the ideal motor 13 position calculated from an equivalent control system model that has the characteristics of the controlled object by a position loop model 15 based on the position command signal θr. Ideally, motor 13 position is controlled to follow the characteristic of position control part 5. So position loop model 15 is designed based on parameters of position control part 5.

When the position/speed operation switching command MOD instructs speed control operation mode, the ideal motor 13 position is calculated and the deviation of the actual motor 13 position measured by the encoder 6 from the ideal position is used for generation. When the command PHS is input, the switch 28 is also switched and the deviation input signal θf is directly input to the position control part 5 by a route passing through the position deviation control part 21.

Thus, when compensation is not required for the position within one motor revolution, acceleration/deceleration time can be minimized according to the motor output torque.

(6) The signal waveform in the proposed control

Figure 13 also shows waveforms of the operation signals when the proposed approach is taken. In the topmost graph, the horizontal axis represents the time, the vertical axis the speed, the chain line the speed command signal Fdt, the dash-dotted line the speed command calculation signal Vr, and the solid line the motor speed Vs. In the second graph from the top, the horizontal axis represents the time, the vertical axis the current, and the solid line the current command value. The

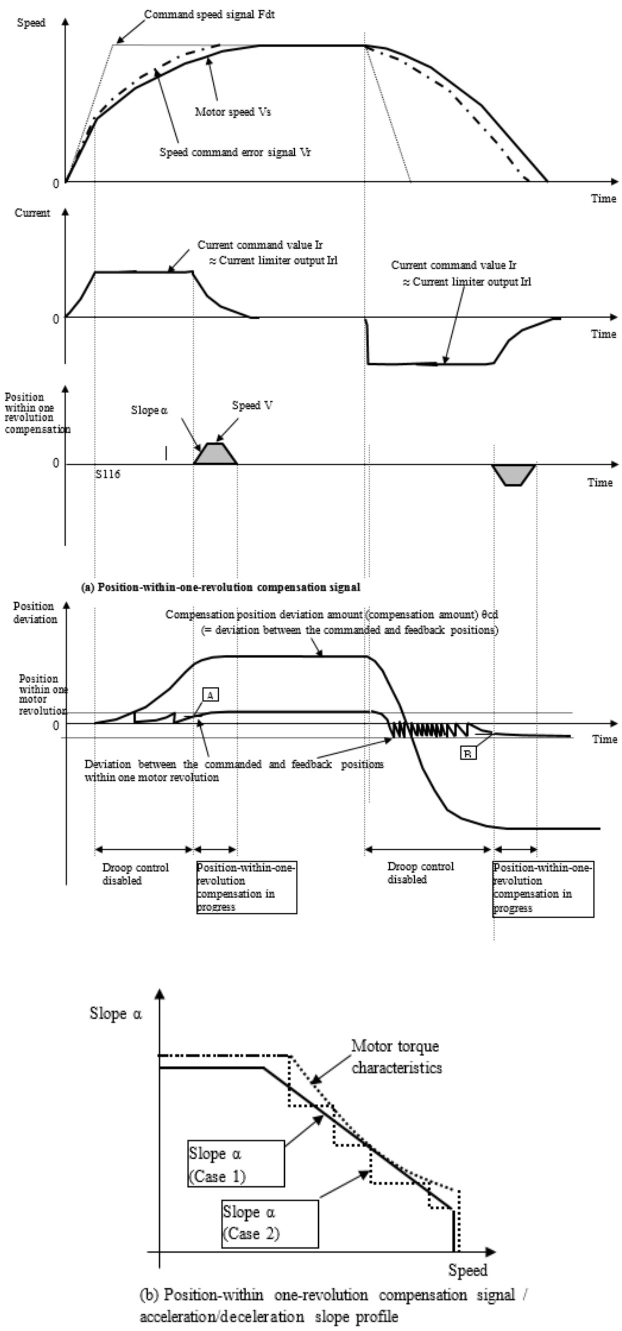


Figure 13: Tracking delay compensation control signal waveform.

third graph shows the position-within-one-revolution compensation signal Vrh. The horizontal axis represents the time, and the vertical axis the compensation amount. In the fourth graph, the horizontal axis represents the time, the vertical axis the position deviation, and the solid lines represent the compensation position deviation amount θcd and the deviation of the position within one motor revolution. In our approach, even if the motor cannot be fully accelerated due to the current limit and the motor speed Vs deviates significantly from a command speed signal Vrv output by the command generation part 1, the position deviation control part 21 limits the position deviation control part output value θg under certain conditions to prevent the speed deviation signal Ve, which represents the gap between the speed command calculation signal Vr output by the position control part 5 and the actual motor speed Vs, from increasing beyond the predetermined level. Thus, transition to position compensation is performed faster when the motor output torque characteristics are restored and the current limit is removed, preventing overshooting from occurring for speed or position control.

Furthermore, after the current limit is removed, the position-within-one-revolution compensation amount Vrh is added to the integrator 20 in the position compensation part 19 to enable consecutive control of the spindle position within one revolution to follow the command. The compensation position deviation amount θcd is reduced during acceleration and increased during deceleration to make the position within one revolution zero. Therefore, it takes less time to determine the compensation amount and perform the compensation for the position within one revolution.

4 VERIFICATION OF THE PROPOSED APPROACH BY SIMULATION

The following shows the result of verification by simulation of the effect of the proposed tracking delay compensation control.

Figure 14 shows the acceleration/deceleration waveform for an ordinary position loop without the tracking delay compensation control we propose. When a program is built with a time constant that would allow the torque to exceed the maximum motor torque, the intended acceleration/ deceleration torque becomes unachievable. Consequently, deviations from position commands are accumulated, and control loop works for recovery, resulting in overshooting.

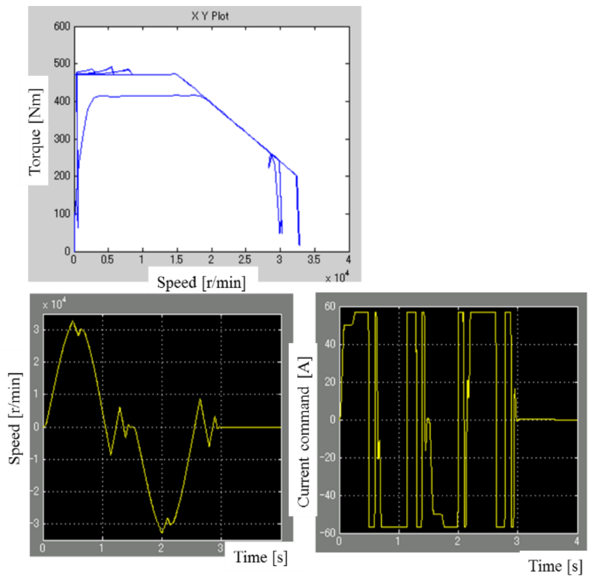


Figure 14: Acceleration/deceleration waveform without tracking delay compensation control.

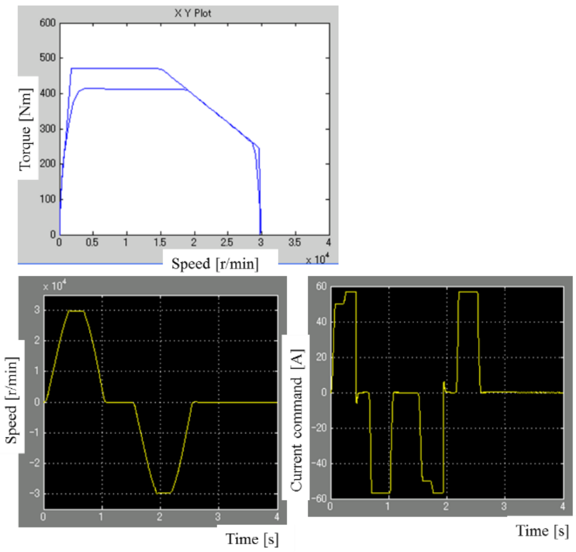


Figure 15: Acceleration/deceleration waveform with tracking delay compensation control enabled.

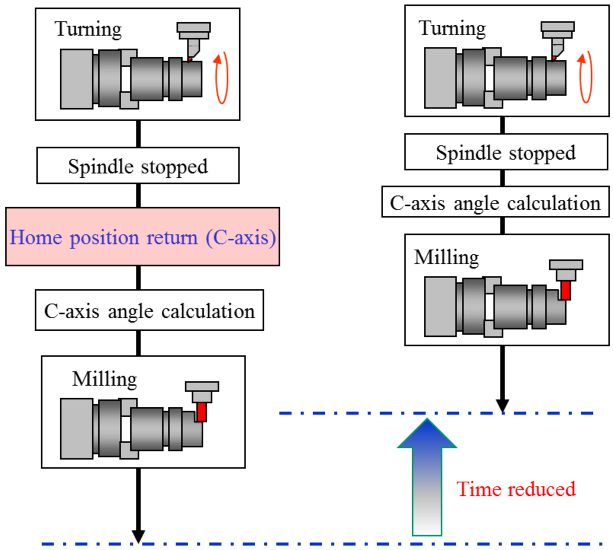


Figure 16: Time reduction for switching from turning to milling (C-axis control).

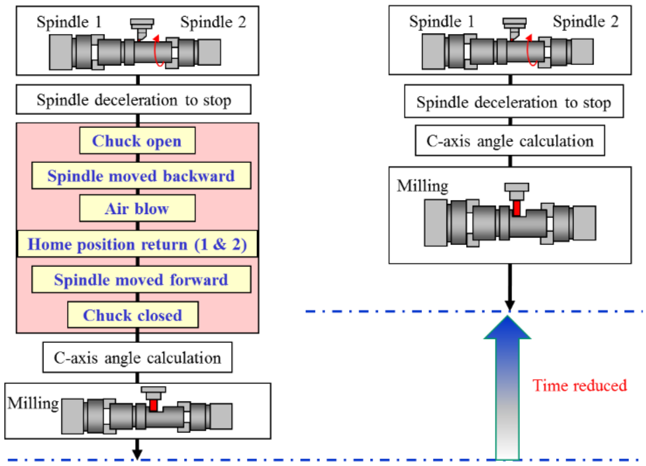


Figure 17: Time reduction for switching from spindle-linked turning to C-axis-linked milling (spindle synchronous control)

Moreover, when the torque reaches the maximum motor torque region, the current is limited in the amplifier and the control system enters nonlinear regions. Consequently, the current command behavior becomes unstable and causes reciprocation between upper and lower current limits. This phenomenon causes increase in mechanical impact and significant damage on the machine.

Figure 15 shows the acceleration/deceleration waveform with the tracking delay compensation control we propose. In order to apply the utmost motor torque, the program is built with an acceleration/deceleration time constant to exceed it, enabling acceleration/deceleration along the diagram of the maximum motor torque. Since the motor torque is restricted against the command, the current is limited at the maximum current in the amplifier. Overshooting does not occur and the control loop remains stable.

5 EVALUATION

5.1 Reduction in the Time Required for Control Mode Switching by Performing Consecutive Position Control

The following shows the result of verification of the effect of implementing the proposed approach with an actual spindle amplifier.

The system for verification is a multi-tasking machine with two main spindles for turning (spindle 1 and spindle 2) and a tool spindle for milling. Each main spindle has feed axes X1 and Z1, and X2 and Z2 respectively. Each main spindle can be used for machining of different workpieces, or a workpiece can be pre-cut by spindle 1, re-chucked by spindle 2 for machining without set-up change operation by an operator. Spindle 1 and spindle 2 can be also used to chuck one workpiece for heavy cutting.

[Reduction in the C-axis switching time]

Since consecutive position control in the system we propose enables position control with main spindles, it is possible to eliminate processes such as stopping the operation once, home position return of the C-axis, and C-axis angle calculation required for transition from speed control for turning to position control for milling with the tool spindle as shown in Fig. 16.

[Reduction in the spindle synchronization time]

As shown in Fig. 17, basically the same processes as shown in Fig. 16 are followed when the main spindle 1 and main spindle 2 chuck a workpiece for heavy cutting (turning) and C-axis control is required in the next process in former systems. In this case, position control is not performed for the spindle 1 and spindle 2 during turning. To perform home position return, each spindle must release the workpiece. Therefore, the chuck of each spindle is opened and the workpiece is held by the other spindle during home position return. The

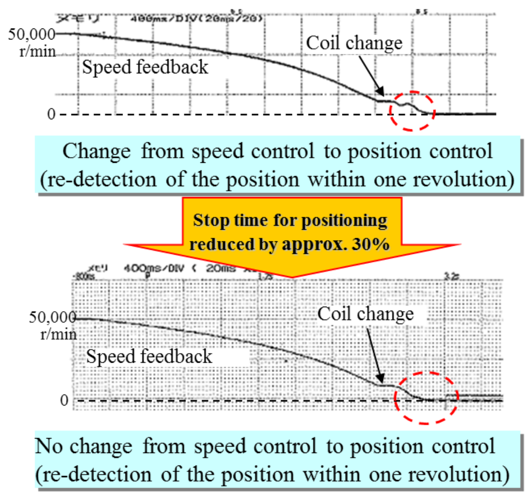


Figure 18: Result of spindle orientation time reduction.

Table 1: Effect of time reduction for operation of the multi-task machining program

New function	Description	Effect		
Spindle consecutive position control	Seamless transition of the operation pattern for switching to C-axis or spindle synchronous control		Previous amplifier	New amplifier
		Switching to C-axis control	0.8 to 1 s	0 s
		Switching to spindle synchronous control	11.2 s	0 s
	Reduction in the orientation positioning time	(1) Example of orientation from 5000 r/min for the main spindle		
			Previous amplifier	New amplifier
		Stop time for positioning	0.32 s	0.23 s

proposed approach reduces the no-cutting time required in existing systems.

[Stopping time reduction for spindle orientation]

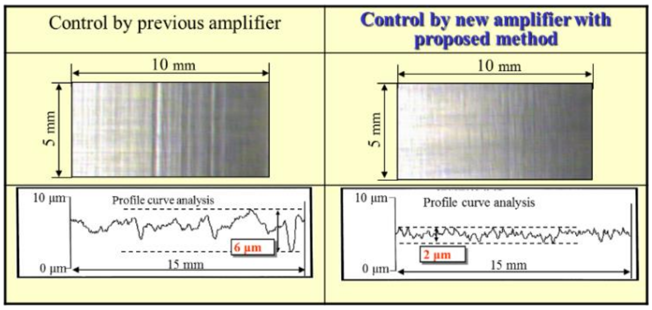
The tool to be used must be changed according to the machining process. To change the tool automatically by a program using an Auto Tool Changer (ATC) without tool change operation by an operator, the spindle must be stopped at a predetermined angle.

In existing systems, the speed of revolutions is changed from a high speed to a certain low speed once for a while for switching to position control, the spindle (tool) angle is detected, a distance command is generated in the amplifier to stop the spindle at a predetermined angle, and then the spindle is stopped.

These processes can be eliminated when the proposed approach is used since consecutive position control is established to keep tracking the spindle (tool) angle in the amplifier, resulting in reduction in the spindle orientation time. Figure 18 shows the test result of the proposed approach.

Table 1 also shows reduction in time required for above-mentioned switching of the machining mode.

5.2 Improvement of Cutting Accuracy by Integrating Servo Control and Spindle Control



Machining conditions
Milling spindle: spindle rotation speed : 9000 r/min,
Main spindle: machining surface speed : 900 mm/min
· Tool: C2MAD160 (Mitsubishi Materials), material: A7075 ALQUEEN
- Cutting fluid: Finecut 1000 (NEOS)

Figure 19: Improvement of cutting performance by integrating servo control and spindle control [reference]

Along with adoption of consecutive position control, we standardized the processor and algorithm between spindle control and servo control, and improved the accuracy of the spindle encoder.

As a result, the feedback control performance level is raised to the same level required for servo axis control that demands high-precision path control.

Figure 19 shows comparison of machining results for C-axis control between the spindle amplifier mainly used for speed control operation and the spindle amplifier with the control performance equivalent to servo control, which is adopted along with adoption of consecutive position control.

6 CONCLUSION

In the past, a focus is put on high-speed revolutions of the tool for spindle motors and spindle amplifiers in machine tools. Therefore, the critical issues are to generate the maximum torque to reduce the acceleration time to reach high speed revolutions, and to generate high power for heavy cutting.

However, in the recent progress in pursuing high productivity for cutting machining, multi-tasking machines for both turning and milling have become more and more dominant. As a consequence, position control such as C-axis control has become essential for spindles (motors and amplifiers) in increasing number of cases and how to reduce the switching time has become a controversial issue.

However, the other ways except for switching had not existed before our proposal. Because no other way but our proposal couldn't solve the overshooting issue of spindle's position loop during acceleration/deceleration with the utmost motor torque.

In this research, we proposed consecutive position control in spindle amplifiers in the same way as in servo amplifiers to address the issue, and verified its validity. This control method has been adopted in our products, which differentiates our products from previous spindle amplifiers and controllers. Furthermore, we are in the process of promoting standardization of amplifiers for motors driving machine tools to remove the barriers between different types of amplifiers.

REFERENCES

[1] S. Sajiki, “Perspective of numerical control for machine tool”, Journal of the Japan Society for Precision Engineering, Vol.24, No.286, pp. 663-667(1985).

[2] M. Sogabe, K. Nakamura, “Latest servo control equipment of NC machine tools”, Journal of the Japan Society for Precision Engineering, Vol.59, No.9, pp. 1427-1430(1993).

[3] D. Kim, C. H. Yim, “All digital high performance controller for spindle motor in CNC machine tool”, IEEE International Electric Machines and Drives Conference Record, pp. MC2-2.1-MC2-2.3(1997).

[4] T. Muraki, H. Yamamoto, “Current State and Outlook of the Multi-tasking Machine”, Journal of the Japan Society for Precision Engineering, Vol.78, No.9, pp. 740-743(2012).

[5] T. Tanaka, K. Tsutsui, “Driving system MDS-D/DH series for M700”, Mitsubishi Electric Advance, Vol.116, pp. 23-25(2006.12).

(Received Octorber 29, 2019)

Kazuhiko Tsutsui received the B.E. and M.E. degrees in Mechanical Engineering from Kumamoto University, Japan in 1990 and 1992 respectively. In 1992, he joined Mitsubishi Electric Corp. Since when he joined Mitsubishi Electric Corp, he has been engaged in development of Servo and Spindle drive system.

Jun Sawaki received the B.E. and M.E. degrees in Electrical Engineering from Meiji University, Japan in 1995 and 1997 respectively. In 1997, he joined Mitsubishi Electric Corp. Since when he joined Mitsubishi Electric Corp, he has been engaged in development of Servo and Spindle drive system.

Naoya Chujo received his B.E. degree in applied physics and his M.S. degree in information science and his Ph.D. degree in electrical engineering from Nagoya University in 1980, 1982 and 2004. He joined Toyota Central R&D Labs. in 1982. He has been a professor at Aichi Institute of Technology since 2010. His research interests are in the area of embedded system and automotive electronics. He is a member of IEEE, IPSJ, IEICE, IEEJ, and Informatics Society.

Tadanori Mizuno received the B.E. degree in Industrial Engineering from the Nagoya Institute of Technology in 1968 and received the Ph.D. degree in Engineering from Kyushu University, Japan, in 1987. In 1968, he joined Mitsubishi Electric Corp. From 1993 to 2011, he had been a Professor at Shizuoka University, Japan. From 2011 to 2016, he had been a Professor at the Aichi Institute of Technology, Japan. Since 2016, he is an Affiliate Professor at the Aichi Institute of Technology, Japan. His research interests include mobile computing, distributed computing, computer networks, broadcast communication and computing, and protocol engineering. He is a member of Information Processing Society of Japan, the Institute of Electronics, Information and Communication Engineers, the IEEE Computer Society and Consumer Electronics, and Informatics Society.

Submission Guidance

About IJIS

International Journal of Informatics Society (ISSN 1883-4566) is published in one volume of three issues a year. One should be a member of Informatics Society for the submission of the article at least. A submission article is reviewed at least two reviewer. The online version of the journal is available at the following site: http://www.infsoc.org.

Aims and Scope of Informatics Society

The evolution of informatics heralds a new information society. It provides more convenience to our life. Informatics and technologies have been integrated by various fields. For example, mathematics, linguistics, logics, engineering, and new fields will join it. Especially, we are continuing to maintain an awareness of informatics and communication convergence. Informatics Society is the organization that tries to develop informatics and technologies with this convergence. International Journal of Informatics Society (IJIS) is the journal of Informatics Society.

Areas of interest include, but are not limited to:

- Internet of Things (IoT)
- Smart Cities, Communities, and Spaces
- Big Data, Artificial Intelligence, and Data Science
- Network Systems and Protocols
- Computer Supported Cooperative Work and Groupware
- Security and Privacy in Information Systems
- Intelligent Transportation System
- Distributed Computing
- Multi-media communication
- Information systems
- Mobile computing
- Ubiquitous computing

Instruction to Authors

For detailed instructions please refer to the Authors Corner on our Web site, http://www.infsoc.org/.

Submission of manuscripts: There is no limitation of page count as full papers, each of which will be subject to a full review process. An electronic, PDF-based submission of papers is mandatory. Download and use the LaTeX2e or Microsoft Word sample IJIS formats.

http://www.infsoc.org/IJIS-Format.pdf

LaTeX2e

LaTeX2e files (ZIP) http://www.infsoc.org/template_IJIS.zip

Microsoft Word™

Sample document http://www.infsoc.org/sample_IJIS.doc

Please send the PDF file of your paper to secretariat@infsoc.org with the following information:

Title, Author: Name (Affiliation), Name (Affiliation), Corresponding Author. Address, Tel, Fax, E-mail:

Copyright

Publisher

Address: Informatics Laboratory, 3-41 Tsujimachi, Kitaku, Nagoya 462-0032, Japan

E-mail: secretariat@infsoc.org

CONTENTS