Wearable Device | Cloud Service | Intelligent Glass

Technology Reports

Expanded Mobile Services with Glasses-Type Devices – Service Usage Proposals for Intelligent Glass –

*This article presents an overview of Intelligent Glass exhibited at CEATEC JAPAN 2013. Google Glass ™ *1 has brought high expectations for glasses-type devices, although rapid popularization of these devices cannot be expected while there are only a few users. To popularize glasses-type devices, specific usage scenes must be presented to the public in stages, and the advantages of the devices must be promoted in a manner that is easy to understand. This article describes a process for popularizing glasses-type devices, and 4 prototype devices.*

Research Laboratories *Mikiko Nakanishi*
Shoji Kurakake
Tsutomu Horikoshi
Service & Solution Development Department† *Takafumi Yamazoe*

1. Introduction

Mobile terminals such as mobile phones and smartphones have enabled users to quickly acquire information. Users always carry their mobile terminals, and take them out when they want to get information. Thus, the more information they want, the more often they have to take out their mobile terminals.

Therefore, eliminating the action of taking out the mobile terminal would offer more convenience to users who use mobile terminals often in daily life. Using a wearable device all the time could address this issue, and we also believe that wearable devices will become indispensable for easily acquiring and manipulating information as the amount of it increases even more into the future [1]-[4].

As wearable devices, glasses-type devices are attracting attention and are predicted to become widely popular following the announcement of Google Glass. However, because users are not accustomed to wearing these devices, we believe it is necessary to present specific uses for them and appeal to their advantages in a way that is easy to understand.

To achieve these aims, we presented 4 prototype systems at CEATEC JAPAN 2013 (hereinafter referred to as "CEATEC") to describe a process of fostering a culture in which these devices are worn. Collectively, these 4 prototype systems are called Intelligent Glass. This article describes a process of popularizing glasses-type devices and NTT DOCOMO's Intelligent Glass.

2. Process of Popularizing Glasses-Type Devices

Figure 1 illustrates a scenario in which glasses-type devices are popularized.

1) Early Stage

To get people to use glasses-type devices, opportunities to wear them must be created. That does not mean first proposing a style in which the glasses are worn all the time, but rather focuses on eliminating initial user reluctance to wear the devices. Thus, the first step towards popularizing glasses-type devices should involve promoting usage of the

† Currently Service Innovation Department

*1 **Google Glass™**: A trademark and registered trademark of Google, Inc.

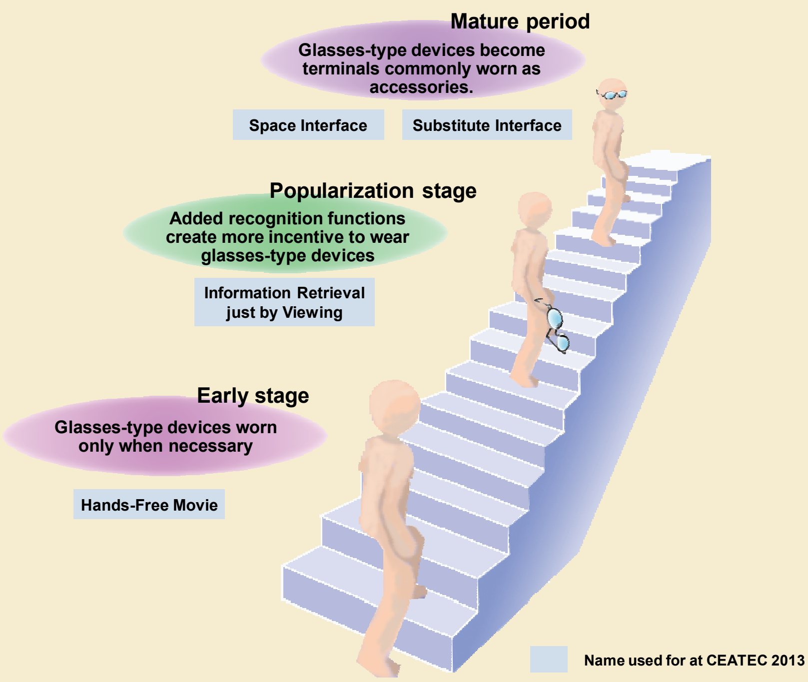


Figure 1　Popularization scenarios for glasses-type devices

devices only when necessary. Here, we propose a Hands-Free Movie experience in which the user watches a long-play movie on the glasses-type device, using the device as an external smartphone display while taking advantage of the device's hands-free characteristic.

2) Popularization Stage

As people become used to wearing glasses-type devices, they will likely demand more functions. Thus, incentives to use the glasses can be increased by including functions in addition to movie viewing. There are a variety of functions that could be added, however, one of the main advantages of glasses-type devices is their ability to display information overlaid on the object or in the direction being viewed by the user. Using this feature, we propose Information Retrieval just by Viewing – an application that includes functions for facial recognition and foreign language menu translation (using text recognition).

3) Mature Stage

The mature stage will be characterized by a world of users always wearing their glasses-type devices The glasses-type devices will become smaller, lighter in weight and more stylish, and will likely become a terminal worn as a common accessory.

Anticipating such a world, we describe two interfaces that will enable users to directly operate contents displayed in their field of vision. The first of these interfaces, called Substitute Interface, will enable users to view and manipulate contents using an everyday object such as a notepad or a folder. The second of these interfaces, called Space Interface, will enable users to have the experience of touching and manipulating virtual contents displayed in the glasses.

3. Intelligent Glass

3.1 Hands-Free Movie

Hands-Free Movie could enable movie viewing during long periods of travel on trains and so forth. Currently, to watch a movie on a smartphone, the user must keep holding the smartphone. However, using the glasses-type device as an external smartphone display will enable the user to watch long movies without having to hold anything (**Figure 2**, **3**).

To select a movie to watch, users can

use their smartphone as a track pad[*2]. Here, before putting on the glasses, the user must check the orientation of the smartphone to operate it while video is being displayed in the glasses (in the user's field of vision). Therefore, to eliminate the need to check the orientation of smartphone, the device determines the user input direction for the smartphone.

Firstly, a proximity sensor[*3] is used when the smartphone is on a flat surface such as a desk or an armrest of a chair to determine which way the user's hand is making inputs on the smartphone, then, by using the proximity and acceleration sensors[*4] and the out camera[*5], it's possible to operate the system without the need to check the smartphone orientation, as shown in **Figure 4**.

It is necessary to operate the smartphone to select a movie, but no matter how long it is, the user can view it without holding anything once it is selected.

At the exhibition, NTT DOCOMO presented a movie on a glasses-type device with a 23° Field-Of-View (FOV).

Exhibition visitors praised the glasses-type device for enabling them to view video on a screen that appears bigger than a smartphone screen. In general, a wide FOV is desirable for a binocular glasses-type display, however, we found for movie viewing there was almost no problem with the 23° FOV.

The Hands-Free Movie concept is a first step to creating a culture in which people will wear glasses-type devices, thus, the purpose of it is to create incentives to wear the glasses.

Glasses-type devices have much potential, but as long as their uses remain unclear, the incentive to wear them will remain weak. In addition, if early adopters[*6] wear glasses-type devices and onlookers don't know what they are doing with the devices, it will end with onlookers only thinking that the early

Figure 2　Terminal for Hands-Free Movie

Figure 3　Video viewed on the glasses-type device

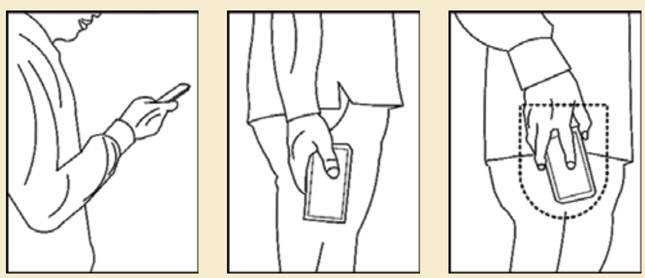


Figure 4　Holding a smartphone

*2 **Track pad**: A flat device that the user strokes with the finger to emulate mouse operation.

*3 **Proximity sensor**: A sensor which detects approaching objects.

*4 **Acceleration sensor**: A sensor that measures acceleration. A component that detects changes in position (orientation) or acceleration in linear directions.

*5 **Out camera**: A camera mounted in the rear of a mobile terminal such as mobile phone or smartphone.

*6 **Early adopter**: A consumer who takes up a new product, service or lifestyle earlier than other people, and influences others about the product, service or lifestyle they have adopted.

adopters look strange. Therefore, to eliminate reluctance to wear and use glasses-type devices, the benefits and uses of the devices must be clarified.

The glasses-type device exhibited at CEATEC were limited to viewing video, but exemplified the convenience of not having to hold a smartphone to watch video.

Many users have already taken up watching video on smartphones, while wearing of earphones to listen to music has been popular for some time. Therefore, we believe wearing glasses-type devices to watch video could be a natural extension of wearing earphones to listen to music, and thus could be accepted by the public relatively easily.

3.2 Information Retrieval just by Viewing

Looking through a glasses-type device at real objects to display information about them on the device is an example of an application of "Augmented Reality" (AR)*7. Since 1990, AR prototype systems have been developed for a range of purposes, and many technologies have been studied, although actual application has not progressed. One reason is that older types of wearable devices require the user to also carry separate equipment such as PC or power source for the device to function. However, the release of Google Glass has shown that there have been significant improvements with these devices, and conditions have become favorable for solving issues towards commercialization. For such devices, we have developed a prototype application called "Information Retrieval just by Viewing" – an AR application that recognizes targets that a user views through a glasses-type wearable device, obtains relevant information about the targets from the Internet, and displays that information on the device.

Information Retrieval just by Viewing is characterized by facial and text recognition capabilities. Thus, by developing practical applications, detailed usage scenes including ecosystems*8 can be considered, and technical issues in actual systems can be extracted.

In a basic configuration (**Figure 5**), a glasses-type device becomes a client that acquires and displays information about recognized objects from a cloud-based server. Cloud-based recognition processing will provide a range of functions so that a variety of objects can be recognized. However, due to privacy concerns with facial recognition, processing has been implemented that enables users to acquire related information only if they have been granted permission by the person recognized, who must also supply the information. Maintenance processing improves recognition performance through learning processes using captured images.

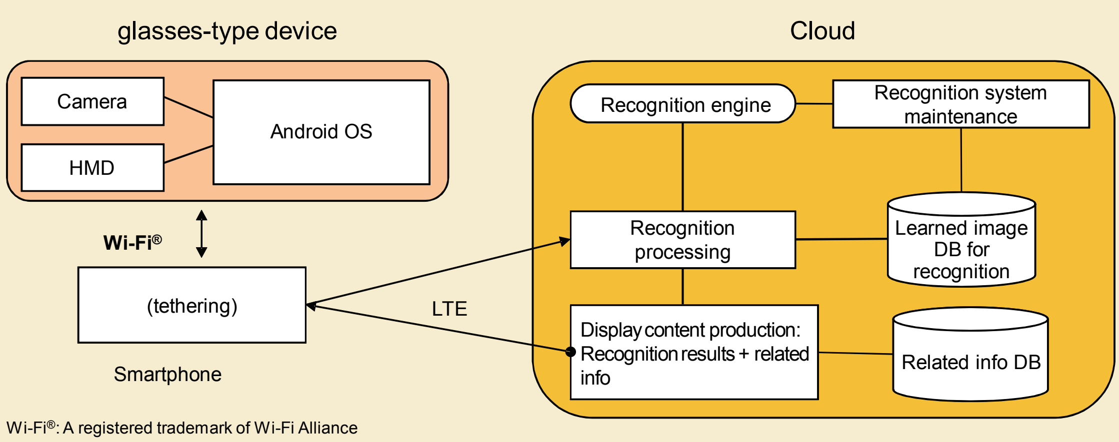


Figure 5 Information Retrieval just by Viewing system configuration

*7 **AR**: Technology for superposing digital information on the real-world in such a way that it appears to the user to be an actual part of the scene.

*8 **Ecosystem**: A symbiotic mechanism in which multiple businesses in a field partner in their business activities to make use of each other's technologies and assets, and include consumers or whole societies to create a series of flows from research and development through to sales, advertising and consumption.

As an example of an effective application of this technology, facial recognition could be useful for enriching customer relationship management in face-to-face meetings with people in specific customer groups. For instance, receptionists could use this application with a glasses-type device to confirm visitors' purposes and meeting histories, which would enable provision of prompt services, personalized services, and recommendation of options tailored to the customer (**Figure 6** (a)).

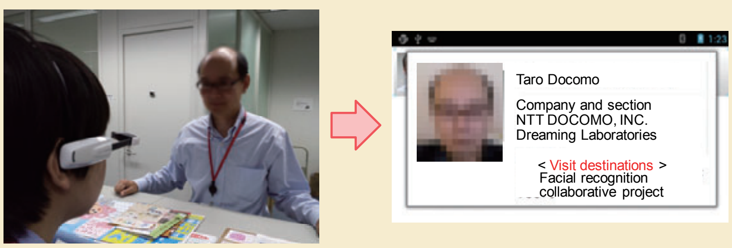
As an example of effective use of text recognition, overseas travelers could use the text recognition function to view menus or maps in foreign languages and get them translated into their own language while wearing a glasses-type device.

Through demonstrations at CEATEC and others, we obtained the following feedback about the Information Retrieval just by Viewing applications we developed.

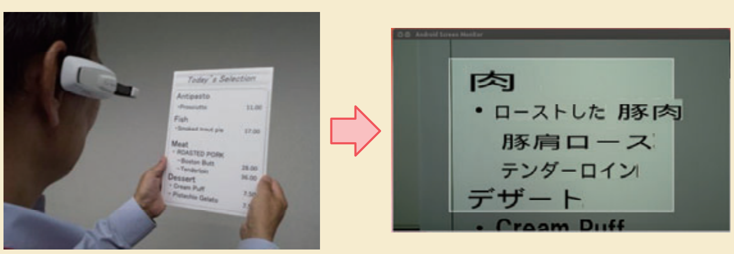
(1) Recognition results should be displayed within 0.3 seconds after viewing at an object so that user activities are not hindered. If recognition processing takes longer, a guide should indicate recognition is taking place and prompt the user to wait for the results.

(2) The number of text characters that can be displayed and viewed immediately on the device screen is extremely small. We found in our demonstration that if more than ten characters per line and more than three lines per screen are displayed, users had to stop their activities and concentrate to read the text.

(3) Because lighting varied from location to location, learning was often required in each place we performed the demonstration. With facial recognition, there are frequent demands to add or remove subjects for recognition. Therefore, the most often used functions in the system are the recognition maintenance functions that perform image sorting and deleting, and learning by the recognition system using the images. Improving the user interface and automating learning will contribute greatly to overall efficiency and make the system easier to use.



(a) Example of enriched face-to-face customer management at a reception (to promptly confirm the purpose of visit etc.)



(b) Example of text recognition (to translate a foreign language menu when traveling)

Figure 6 Usage scenes for Information Retrieval just by Viewing

Making improvements to the system based on this feedback are issues to be tackled in future.

3.3 Substitute Interface

Substitute Interface is technology that enables use of an everyday object such as a notepad or a folder as a tablet PC. By tapping a rectangular object with the finger, the user can use that object in the same manner as a tablet PC.

Mobile terminals such as laptop PCs and smartphones enable users to watch and manipulate content anywhere and anytime, and are extremely convenient. However, smartphone screens in particular are still quite small for watching video - users would like to use devices with larger screens, but these are more troublesome to carry around.

Nevertheless, our bags have schedulers, notepads and folders in them – there are many objects that are bigger than the smartphone screen. If these objects could be substituted for a device with a larger screen, users could enjoy video on a larger screen without having to carry around a bulky mobile terminal.

Substitute Interface enables the display size to be changed easily by selecting any suitably-sized object, and thus frees the user from having to carry numerous devices with different screen sizes.

Figure 7 shows a usage image of Substitute Interface. This glasses-type device is a binocular device that projects the same image into each eye. Using the camera mounted on a glass-type device and a ring-type input device worn on the finger, the system recognizes the object the user intends to use as a substitute tablet PC, and recognizes input actions such as click or swipe (to turn pages etc).

When the user taps the object, the camera mounted on the glasses-type device recognizes which object the user intends to use. Only rectangular objects are recognized by their shape and color [4].Thus, the user can use any object on hand, which eliminates the need for objects specially prepared with AR markers etc, or specifically designed objects. Content images are overlaid on the recognized object and adjusted to its size. This makes it possible to view content at the size of the object.

Figure 8 shows content (four icons) overlaid on a notepad.

The larger the object, the more icons can be displayed over it, which preserves visibility during input actions. However, if the object is small and a number of icons are displayed over it, the icons would be smaller and more difficult to manipulate for input. To solve this issue, the user interface is designed to be dependent on the size of the object used as the screen - the system recognizes the size of the object, and automatically adjusts the user interface to fit it (**Figure 9**). Using a large object means multiple icons can be displayed, which is important for visibility. With a small object, only one

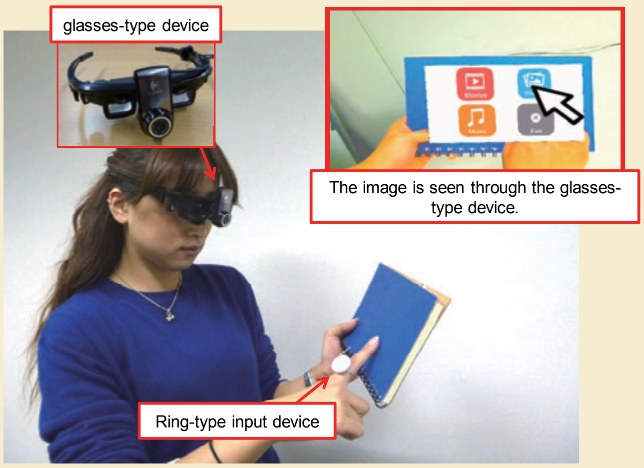


Figure 7　Substitute Interface

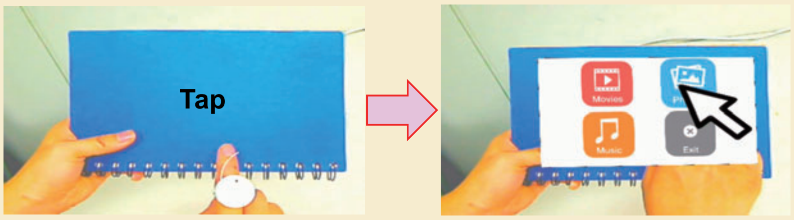


Figure 8　Content images overlaid on a tapped object

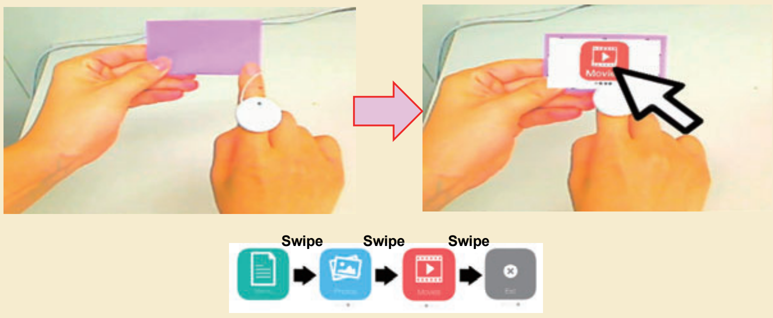


Figure 9　UI with a small object

icon is displayed at a time to make it easy to operate. In this case, the user can change the icon with a swipe action.

To recognize an input operation, the system detects the finger position and the click timing. Click timing is recognized via an acceleration sensor in the ring-type input device when the user taps the object. At the same time, the finger position is recognized by the camera. This enables detection of user input actions, and thus enables touch panel-like screen operation.

However, the following issues currently exist with this technology:

Substitute Interface uses 23°and 35° Field-of-View (FOV) glasses-type devices. With the 23° device, content can only be overlaid on part of an A5 notepad, however, with the 35° device, contents are visible over the entire A5 notepad. Despite this, the A5 size image is still small, so to view images overlaid on a larger object, glasses-type devices with a FOV greater than 35° will be required.

Visitors to the exhibition also complained that the image displayed in the glasses-type device could not be seen overlaid on the object. The focal length (position in the depth direction) of displayed images on the glasses-type device is fixed. Therefore, images might appear in front of the object the user is holding, and not appear neatly overlaid on the object. To solve this issue, stereoscopic images matched to the distance to the object could be displayed after detecting the distance whenever necessary, or a focus-free (images focused to match the focal point of the eye) retinal projection-type display could be used.

3.4 Space Interface

We developed a system to demonstrate Space Interface – an example user interface using the glasses-type device (**Figure 10**)*. By attempting to make behavior more realistic with more realistic reactions, Space Interface is technology that superposes computer graphic images in the real world. The technology enables users to directly manipulate a CG image as if it were a real object, using physical simulation technology to make real-time computations.

To enable users to manipulate the CG image projected, an infrared sensor in the glasses-type device is used to detect and recognize the near field of vision. Because the technology does not use conventional human body or hand skeleton models, it can be operated more freely.

Because the technology is based on the idea that if one moves a real object it will respond in a particular way, users can operate the technology without using any specific patterns such as gestures, even if they have no prior experience. Furthermore, because the technology enables users to manipulate real objects and reproduces their movement, it also enables a range of different expressions by simply changing CG models during content production. For example, con-

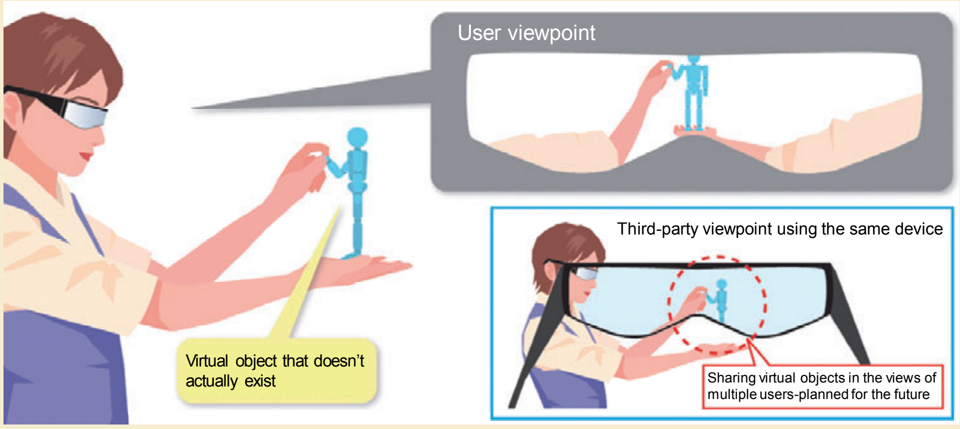


Figure 10 Overview of Space Interface

* Results of the FY 2008 IPA (Information-technology Promotion Agency) Exploratory IT Human Resources Project are used for some parts of Space Interface.

necting circles and sticks together in the shape of a human could transform into the image of doll, or creating CG cloth in the shape of clothing could turn the technology into a virtual trial fitting system.

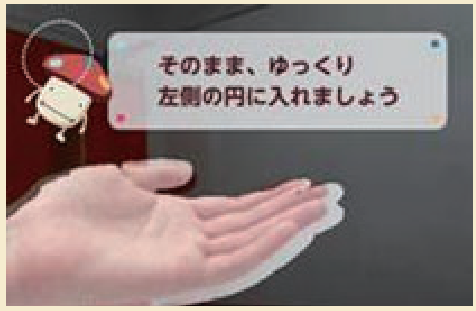
The CEATEC exhibition included a demonstration of the game featuring Docomodake (**Figure 11**), a demonstration of touching a virtual teddy bear (**Figure 12**), and a demonstration of expanding, shrinking and rotating images (**Figure 13**) which were enjoyed hands-on by many visitors. Although not demonstrated, creating CG physical simulation space for objects in the cloud and using high-speed, low-latency communications lines could enable personal communications through CG images or team work at a distance. Latency of a suitable standard for these sorts of applications has been confirmed in an LTE-Advanced[*9] experimental system handling bidirectional hi-vision images through a server processing physical simulations for the images.

Currently, due to the types of sensors that can be head mounted, because superimposing objects in real space can only be done on 1 plane, and because contents are also handled as two-dimensional objects on that plane, movement perpendicular to the plane of the object being manipulated lowers operability.

In the future we would like to develop devices with miniaturized sensors that can perceive 3-D space and improve contents. Moreover, because this technology does not have feedback mechanisms to stimulate the sense of force, users cannot feel any inertia as they manipulate objects. Therefore, we aim to enable operability that creates the illusion of touch using real-time visual feedback as users manipulate CG images using real objects. At CEATEC, many people said they could more easily manipulate the objects if they felt like they were really touching them. We would like to keep studying and evaluating the physiological aspects of these operating sensations.

4. Conclusion

Currently there are a range of issues



※This is provided only in Japanese at present.

Figure 11　Game style with Docomodake

※This is provided only in Japanese at present.

Figure 12　Manipulating a teddy bear

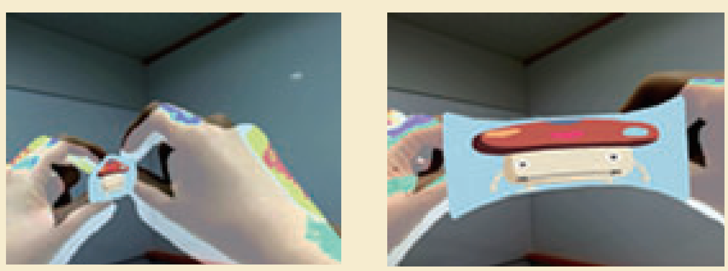


Figure 13　Pulling and stretching an image

*9　**LTE-Advanced**: A radio interface enhancing LTE to be standardized as 3GPP Release 10.

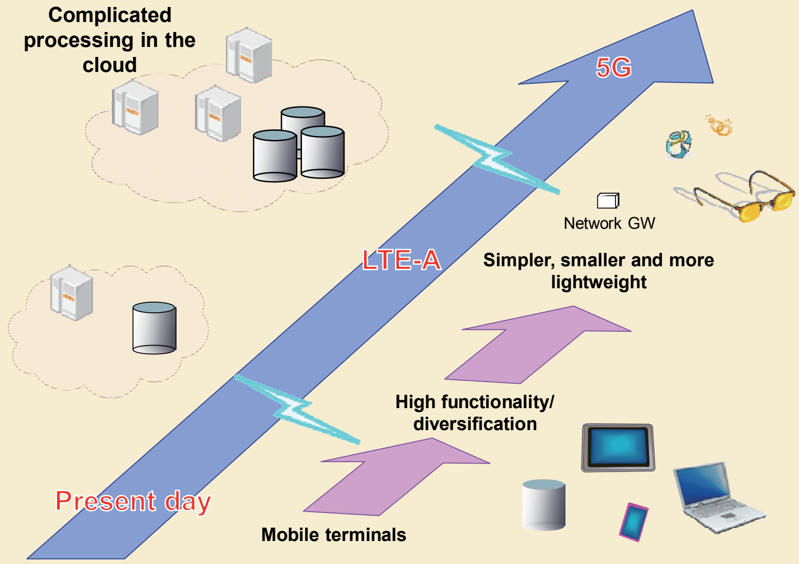


Figure 14 Terminals changing as the cloud and mobile networks evolve

that must be solved for glasses-type devices, such as the size of the device, its weight, battery life, and the size of images (FOV width). We expect it will take some time to solve these issues, but we believe these issues will be gradually tackled as these devices become more popular.

We also think that in addition to glasses, wristwatches and wristbands, wearable devices will become even more diverse, and will connect to the cloud via a network gateway. The cloud will handle complex processing, while the devices will only have simple input-output functions. With cloud-based processing, these systems will enable stress-free operation due to low latency communications enabled by superior mobile network speed and capacity, and devices with input and output functions only will enable even more miniaturization. (**Figure 14**).

Not only will linking wearable terminals to the cloud through mobile networks enable users to acquire and manipulate data easily, but will also create potential for a range of other services.

In particular, if users always wear a device, more data can be acquired about the movements and physical condition of individual users than is available using a smartphone. Based on such data, mobile services can be tailored to meet individual needs, and thus be easier to use.

Nevertheless, we have to also think about how to protect personal data obtained from wearable devices, and how to get society's understanding of their uses.

References

[1] M. Fukumoto: "A finger-ring shaped wearable handset based on bone conduction," Proc. of the Ninth IEEE International Symposium on Wearable Computers, pp.10–13, Oct, 2005.

[2] H. Manabe and M. Fukumoto: "EOG-based Gaze Input Interface Using Headphones," IPSJ Vol.52, No.4, pp. 1515–1526, 2011 (In Japanese)

[3] S. Kimura and M. Fukumoto and T. Horikoshi: "Eyeglass-based hands-free videophone," Proc. of the 2013 International Symposium on Wearable Computers, pp. 117–124, 2013.

[4] M. Nakanishi and T. Horikoshi: "Intuitive Substitute Interface," Personal and Ubiquitous Computing, Volume 17, Issue 8, pp 1797–1805, Dec. 2013.