NTT DOCOMO

Technical Journal

Vol.20 No.4 | Apr. 2019

DOCOMO Today

- Moving from a Mobile Infrastructure to a Co-creation Platform Network

Technology Reports (Special Articles)

Special Articles on Making Life More Convenient and Seamless —toward Future Lifestyles

- docomo TV Terminal as “Your TV”
- “IoT Smart Home” Supporting Daily Life Activities
- Personalized Screen Concept Using a Gesture Controlled UI
- A Face-to-face Video Calling System Facilitating Natural Eye Contact

Technology Reports

- 8K VR Video Live Streaming and Viewing System for the 5G Era
- New Platform Technology to Further Improve the Quality of Multimedia Services - MediaSDK Software Library -

NTT docomo

Moving from a Mobile Infrastructure to a Co-creation Platform Network

DOCOMO Communications Laboratories Europe GmbH
Managing Director, President & CEO
Takatoshi Okagawa

DOCOMO Communications Laboratories Europe GmbH (DOCOMO Euro-Labs) was established in 2000 in Munich, Germany. Today, we are actively involved in studying specifications for the next-generation core network and in consolidating resources for international standardization activities.

Specifically, we are participating in the work of various standardization bodies with a focus on 3rd Generation Partnership Project Service and System Aspects 2 (3GPP SA2), which formulates standards and specifications for the core network in the 5G era, and European Telecommunications Standards Institute Industry Specification Group Network Functions Virtualisation (ETSI ISG NFV), which formulates standards and specifications for network virtualization technologies. As part of this effort, we are linking up with NTT DOCOMO's Research Laboratories and development departments and collaborating with European vendors in conducting technology studies while endeavoring to formulate specifications for related standardization bodies.

The 5G system aims to accommodate many and varied requirements including high-speed and high-capacity communications, low-latency, and massive connectivity (for IoT devices, etc.). It is expected to introduce technical innovations not only in the conventional wireless network but also in the core network.

In addition, by creating new industries through co-creation with a variety of industries, we can think of 5G as a system that can help solve pressing social problems and contribute to regional revitalization. We are therefore participating in standardization activities for the automation of operations in plants and factories in diverse industries including the automobile industry (spanning automobile manufacturers, parts manufacturers, etc.) and forming alliances with individual industry groups such as the 5G Automotive Association (5GAA) and the 5G Alliance for Connected Industries and Automation (5G-ACIA). These types of activities were not previously seen in the LTE era.

In the conventional LTE-based core network, we constructed a single network for general consumers and achieved a low-cost and high-function solution while guaranteeing a level of reliability fitting a mobile carrier. In contrast, the core network in the 5G era will require diverse functions required by different industries, the ability to add functions quickly, a flexible billing system, and depending on the field, a higher level of reliability. This, in turn, will require major changes in network design and development techniques and in maintenance methods too.

Technically speaking, this means the application of microservices that can make conventional network functions even smaller in scope to enhance versatility, reusability, and convenience, network slicing technology that can flexibly arrange network functions specific to multiple domains, technology for automating the generation and maintenance of such network slices, state-discrete architecture for simplifying state management and achieving efficient deployment on cloud platforms for mobile telecom companies, and various types of container-based virtualization technologies. In this way, integration and restructuring with new technologies will also be needed in the study of international standards and specifications for the 5G network.

In June 2018, the specifications for 5G New Radio (NR) were completed at 3GPP as Release 15. This should accelerate the commercialization of 5G in various regions around the world. However, in the sense of creating a platform to enable co-creation with industry as mentioned above, specifications are still incomplete, and a variety of challenging issues have already arisen in formulating specifications for Release 16.

Furthermore, as NTT DOCOMO and other operators are beginning to commercialize network virtualization technologies, it was decided to extend the work of ETSI NFV up to 2020 as a fourth phase. This work includes the formulation of specifications that incorporate automated operations and innovative technologies from the IT industry as a platform supporting 5G network functions. All in all, international standardization activities in this area continue with increasing enthusiasm with a view to further evolution.

Amid these developments, research and development at DOCOMO Euro-Labs is focused on elemental technologies that can support the business evolution of NTT DOCOMO in the coming 5G era and on the construction of a platform network for achieving co-creation. I would like to contribute to lasting happiness in the world and a meaningful life for everyone through international standardization activities that take advantage of the benefits (knowledge) of the European region. We will move forward under the slogan "Know your customer, know the world, and think on one's own. Establish your direction, declare it to the world, and form strategic partnerships to make NTT DOCOMO into a global industry leader."

NTT DOCOMO
Technical Journal
Vol.20 No.4 Apr. 2019

[Contents]

DOCOMO Today

Moving from a Mobile Infrastructure to a Co-creation Platform Network
Takatoshi Okagawa 1

Technology Reports (Special Articles)

Special Articles on Making Life More Convenient and Seamless—toward Future Lifestyles

ユーザーを選択してください。
選択されたユーザーのdアカウントでログインします。

(P.4)

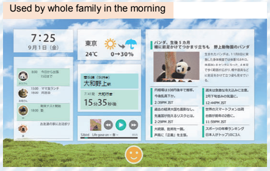
docomo TV Terminal as “Your TV” 04

Video Delivery | User Switching | Viewing History

“IoT Smart Home” Supporting Daily Life Activities 16

IoT | Smart Home | WEB API

Used by whole family in the morning



(P.25)

Personalized Screen Concept Using a Gesture Controlled UI 25

Personalization | Home Devices | Gesture Controlled UI

(P.34)

A Face-to-face Video Calling System Facilitating Natural Eye Contact 34

Eye Contact | Face-to-face | Video Calling

(P.43)

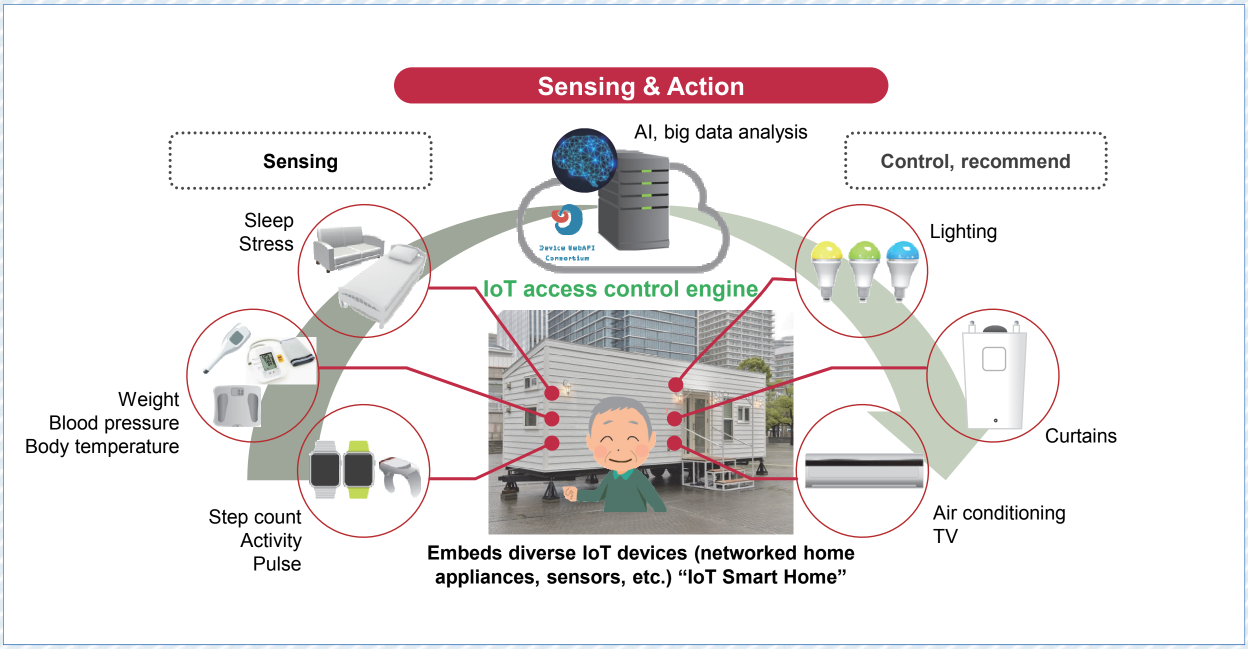
Technology Reports

8K VR Video Live Streaming and Viewing System for the 5G Era 43

VR Live Video Streaming FPGA

New Platform Technology to Further Improve the Quality of Multimedia Services
—MediaSDK Software Library— 51

Player QoE Quality Improvement



Technology Reports (Special Articles) “IoT Smart Home” Supporting Daily Life Activities (P.16)

Overview of IoT Smart Home

Technology Reports

Video Delivery　User Switching　Viewing History

Special Articles on Making Life More Convenient and Seamless—toward Future Lifestyles

docomo TV terminal as “Your TV”

Communication Device Development Department　Kenichiro Masami　Chihiro Suzuki
Yuya Tanaka　Kasumi Araki

In recent years, set top boxes oriented to video delivery have become increasingly popular around the world. However, they have yet to be adequately equipped with content recommendation functions tailored to user preferences based on viewing history. To achieve such a function, NTT DOCOMO has developed a home device as “Your TV” that features zapping viewing based on user viewing history centered about d ACCOUNT®[*1] and a multi-account function envisioning family usage.

1. Introduction

In light of the recent spread of home devices centered about video services, NTT DOCOMO has developed the docomo TV terminal®[*2] set top box as “Your TV” based on the product concept of “broadening the enjoyment of all NTT DOCOMO video services on the family’s home TV as desired by each member of the family” [1]. In other words, this product, while assuming family usage, can also meet individual viewing needs.

An external view of the docomo TV terminal is shown in **Photo 1** and main specifications are

Photo 1　External view

[*1] d ACCOUNT®: A free common ID for using a variety of services provided by NTT DOCOMO such as net shopping and digital content. A registered trademark of NTT DOCOMO, INC.

[*2] docomo TV terminal®: A trademark or registered trademark of NTT DOCOMO, INC.

listed in **Table 1**.

Set top boxes up to now have featured a usage format in which users actively select the content that they wish to watch from a program schedule that includes standard recommendations from the service provider. Consequently, with the aim of encouraging passive viewing tailored to individual users, NTT DOCOMO introduces automatic zapping made possible by inferring the content or programs that each user would like to watch at the present moment from viewing history and presenting those recommendations on the home screen as the user's first view (**Figure 1**).

Conventional set top boxes have also suffered from a variety of issues including the difficulty of setting multiple accounts for family usage and of using a remote control unit in addition to hard-to-understand Frequently Asked Questions (FAQ) on terminal operation.

This article provides an overview of docomo TV terminal, explains the mechanism of d ACCOUNT and the user experience[*3] with the Home app that resolves the above issues, and describes the technology and specific usage scenarios of the “Osusume hint (Recommended Usage Hints)” function.

Table 1　Main specifications

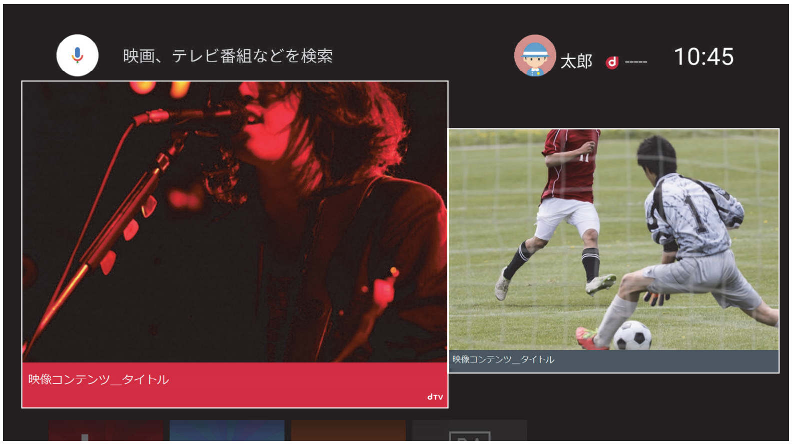
Color	White
Size	107 mm (W) × 107 mm (D) × 25.5 mm (H)
Weight	209 g
OS	Android TV 7.0
CPU	Quad Core 1.6GHz
Internal memory capacity (RAM/ROM)	RAM3GB/ROM16GB
HDR	HDR10, HLG, Dolby Vision
DLNA	Supports only the DMS function. docomo TV terminal apps are needed for viewing. DLNA/DTCP-IP (Hikari TV for docomo only)/DTCP+ (Hikari TV for docomo only)
LTE	LTE/3G/GSM not supported (no insertion of UIM card)
External ports	HDMI2.0a Gigabit Ethernet USB2.0×1 port, USB3.0×1 port
Wi-Fi	IEEE802.11ac/a/b/g/n, MIMO supported
Bluetooth	Bluetooth4.2
Remote control	Built-in mike for voice input, Bluetooth, infrared supported
Supported services	DTV, d anime store, dTV channel, DAZN for docomo, Hikari TV for docomo

DLNA: Digital Living Network Alliance
DMS : Digital Media Server
DTCP: Digital Transmission Content Protection
HDR : High Dynamic Range
HLG : Hybrid Log Gamma
MIMO: Multiple Input Multiple Output

[*3] User experience: Everything a user feels when using, consuming, or owning a product or service.

2. Application Configuration of docomo TV terminal

The application configuration of docomo TV terminal is shown in **Figure 2**. Loaded on the Android™[*4] TV OS application layer, these applications are divided into basic functions and video services. The former consists of the “Home app,” “d ACCOUNT setting app” for authenticating and managing the user’s d ACCOUNT, and “Osusume hint” for presenting advice on terminal operation, while the latter consists of various applications for running video services such as “dTV®[*5]” and “Hikari TV®[*6]” that output video content. The following



*The screen shows sample images.
*This is provided only in Japanese at present.

Figure 1　Zapping UI

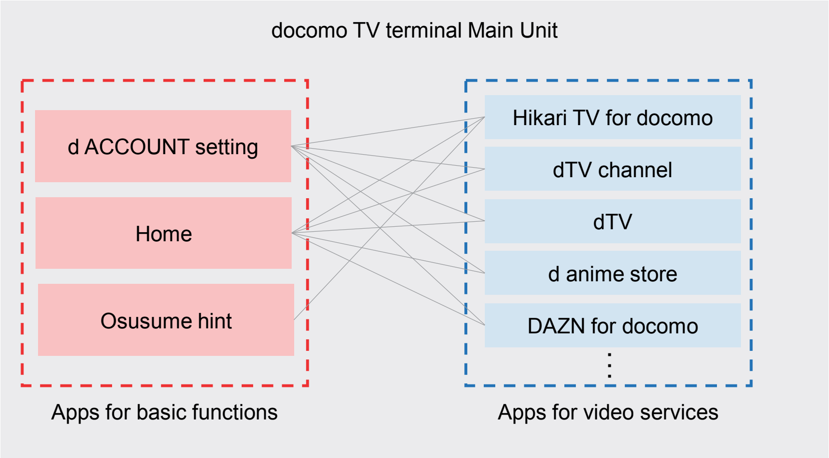


Figure 2　Application configuration

[*4] Android™: An open source platform targeted mainly at mobile terminals and promoted by Google Inc. in the United States. A trademark or registered trademark of Google LLC in the United States.

[*5] dTV®: dTV, dTV channel, and dTV terminal are registered trademarks of NTT DOCOMO, INC.

[*6] Hikari TV®: A video delivery service operated by NTT Plala Inc. A registered trademark of NTT Plala Inc.

describes the technology for achieving “Your TV” and specific usage scenarios.

3. User Management by d ACCOUNT

3.1 Multi-account Function

Contract information and usage conditions for each user with respect to the various types of services provided by NTT DOCOMO are managed using d ACCOUNT, a free common ID [2]. With docomo TV terminal, the user only has to log into the terminal once. There is no need to log in every time a different service is used since account information can be referenced from the d ACCOUNT setting app.

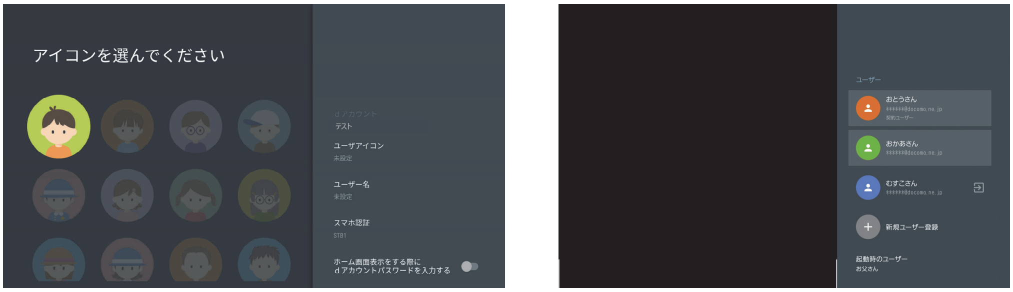
As described above, account management by the d ACCOUNT setting app is similar to that of a smartphone, but while a smartphone corresponds to an individual account, docomo TV terminal, which envisions family usage at home, was designed to manage multiple accounts. This is called a multi-account function.

In the case of family usage, selecting your own icon from among multiple user icons when the terminal starts up enables you to log in to all services from your own account as shown in **Figure 3**. The docomo TV terminal uses this multi-account function to present recommended content in a zapping-type format on the home screen based on the viewing history of each account. In this way, “Your TV” comes to life immediately after starting up the terminal.

3.2 Authentication Processing with d ACCOUNT Setting App

A user’s d ACCOUNT, which is used as login information for many services, is securely managed in the form of a token*7 at login time on that terminal. The docomo TV terminal as well uses a token for the d ACCOUNT setting app to perform authentication processing for various services thereby enabling single sign-on*8 within the terminal.

The docomo TV terminal also displays an icon and user name as account information on the screen.



アイコンを選んでください

dアカウント
テスト

ユーザアイコン
未設定

ユーザー名
未設定

スマホ認証
STB1

ホーム画面表示をする際に
dアカウントパスワードを入力する

ユーザー

おとうさん
******@docomo.ne.jp
契約ユーザー

おかあさん
******@docomo.ne.jp

むすこさん
******@docomo.ne.jp

新規ユーザー登録

起動時のユーザー
お父さん

*This is provided only in Japanese at present.

Figure 3　Account-registration and account-viewing screens

*7 Token: The result of converting information into a character string, used here to transform d account information into a character string that cannot be understood by another party.

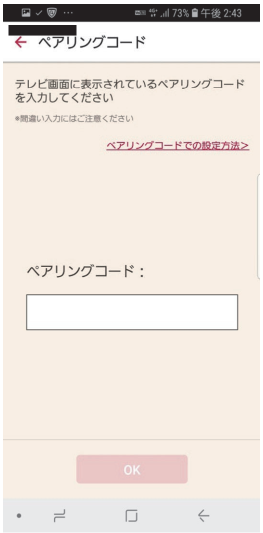
*8 Single sign-on: The ability of logging into multiple services with a single account.

Switching from this icon and user name to another user changes the active user (the account currently using the terminal).

3.3 Simple Account Registration

Account registration results in automatic authentication. For example, when a user purchases a terminal at a store, a d ACCOUNT will be registered on the terminal on the store side making it unnecessary for the user to register a d ACCOUNT again when connecting from home (although an authentication key must be input from the user's smartphone). This scheme saves the user the trouble of making initial d ACCOUNT settings on the terminal and enables the user to start using services anytime after making the purchase simply by connecting the terminal to a power supply and the Internet.

Of course, d ACCOUNT registration on the terminal may also be performed manually, and for this case, NTT DOCOMO provides a registration method using a pairing code[*9] that simplifies input by linking with the user's smartphone. In this method, the user inputs the code displayed on docomo TV terminal using the d ACCOUNT setting app on the user's smartphone. This action registers the d ACCOUNT registered on the smartphone with the doocmo TV terminal as well (**Figure 4**).



テレビ画面に表示されているペアリングコードを入力してください

※間違い入力にはご注意ください

*This is provided only in Japanese at present.

Figure 4　Registration by pairing code

3.4 Smartphone Authentication

The doocmo TV terminal takes into account the use case of switching among multiple accounts. It therefore recognizes the need for privacy and requires a password when switching accounts to prevent other users from using one's own account.

However, the need for inputting a password may frequently occur in actual use. With this in mind, NTT DOCOMO also provides a function for performing authentication by smartphone to avoid the security-related problem of displaying a password on a screen at the time of input and to eliminate the annoyance of having to input a password by remote control every time. In this regard, authentication by smartphone can already be performed by a function provided on the d ACCOUNT setting app for smartphones. This function can also be used to perform d ACCOUNT authentication on a computer browser or a set top box such as doocmo TV terminal by having that equipment send a notification to the smartphone and then using biometric authentication processing (fingerprint/iris authentication) on the smartphone. This scheme negates the need for inputting a password on the screen and makes the authentication process simpler and more secure. The docomo TV terminal

[*9] Pairing code: An identifier for performing d ACCOUNT authentication.

provides the same sort of function enabling authentication by smartphone to be used as an alternative to password input on the terminal (**Figure 5**).

4. User Experience Features

4.1 Product Concept and Provided Value

The product concept can be broken down into three features: all NTT DOCOMO video services, home TV, and support for each family member. Each of these features is summarized below.

1) All NTT DOCOMO Video Services

This feature means exactly what it says: purchasing this product enables the user to experience the freedom of enjoying all NTT DOCOMO video services. As of September 2018, these video services included DTV, d anime store, dTV channel®, DAZN®*10 for docomo, and Hikari TV for docomo, but there had not been a device in the NTT DOCOMO product lineup that could use all of these services. For example, Hikari TV that assumes the use of an optical circuit could not be viewed with a smartphone, and dTV terminal® sold in the past supported the viewing of only DTV and d anime. This development of doocmo TV terminal has made it possible to view all video services.

2) Home TV

This feature emphasizes the use of a TV terminal for viewing video services instead of a smartphone. However, in contrast to using a smartphone that assumes input by a touch panel, using a TV terminal assumes input operations by a remote control corresponding to the TV screen display, which has a negative effect on operability (degree of simplicity, degree of freedom, etc.). For example, when selecting content in a carousel format*11, only one step is needed when using a mobile device such as a smartphone while two or more steps are needed when using a remote control. There is therefore a need for improving operability with a remote control when a TV terminal is the target device.

3) Support for Each Family Member

Differences in viewing habits emerge in relation to the number of users. “Home TV” presents

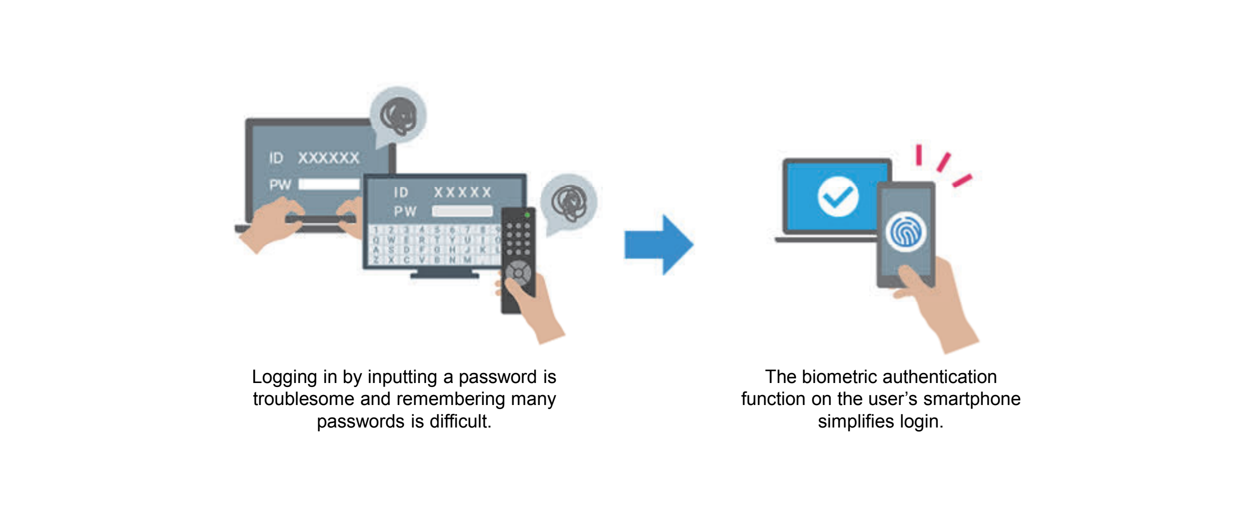

ID XXXXXX
PW
ID XXXXX
PW
Logging in by inputting a password is troublesome and remembering many passwords is difficult.
The biometric authentication function on the user’s smartphone simplifies login.

Figure 5 Authentication by smartphone

*10 DAZN: A trademark or registered trademark of Perform.

*11 Carousel format: A scheme for displaying multiple objects in a linked manner and for selecting a particular object by sliding to it.

an image of a TV set in a family’s shared living space like a living room in contrast to that of an individually owned smartphone. It means that the whole family may view content together or that an individual member of the family may view personally preferred content. In short, users’ needs change according to usage scenario, so there is a need for an environment that can meet the needs of a variety of users.

In light of the above, the value that docomo TV terminal aims to provide can be summed up as “the ability to comfortably enjoy the NTT DOCOMO video services that I as a member of my family would like to watch right now on our TV screen.” A User Interface (UI)[*12] has been designed to achieve this value.

4.2 UI Designs on the Home Screen

Two UI designs called “multiuser personalization” and “zapping UI” were proposed for the docomo TV terminal as described below.

1) Multiuser Personalization

In the case of a family, this refers to a state in which the information required for the father, mother, and each child is optimized for the father, mother, and each child, respectively. Consequently, in addition to the scenario in which the family enjoys content together, this implies the provision of personalized information so as to satisfy the preferences of the user operating docomo TV terminal. Personalization of content is thought to be achieved by providing “recommendations based on the current service contract[*13]” and “recommendations based on a viewing log[*14].” With this in mind, the following UI elements are required.

- The user (personalized information) who wishes to log in when starting up the terminal must be easy to select
- After login, the user must be treated as being one and the same for all services
- Information on users not logged in must not be displayed

In the case of a dedicated device, user authentication can be easily performed at a prescribed position within the startup sequence. In an Android environment, however, all applications can operate in an asynchronous manner, which makes it difficult to implement a function for suppressing the launch of other applications until user authentication completes. To solve this problem, this product is designed with a function for logging into the Home app—the starting point of user operations—so that other applications cannot be used in a non-logged-in state. The startup sequence of the docomo TV terminal is shown in **Figure 6** (1) – (3) and described below.

(1) Although corresponding to a return from SLEEP mode in a smartphone, pressing the power button always launches the Home app on the framework[*15] layer so that the screen of another application does not accidentally appear in front of the user. It must also be considered that multiple startup modes exist and that there are cases in which terminal startup is not equal to the Home app startup, so extra information is given so that terminal startup is understood. This scheme achieves a function that ensures startup of the Home app on the first screen.

(2) To provide for the case in which another

[*12] **UI:** Operation screen and operation method for exchanging information between the user and computer.

[*13] **Recommendations based on the current service contract:** Recommendations that prioritize user-contracted services from among all video services provided by NTT DOCOMO.

[*14] **Recommendations based on a viewing log:** Recommendations that specify content that the user would like to watch next based on a viewing log.

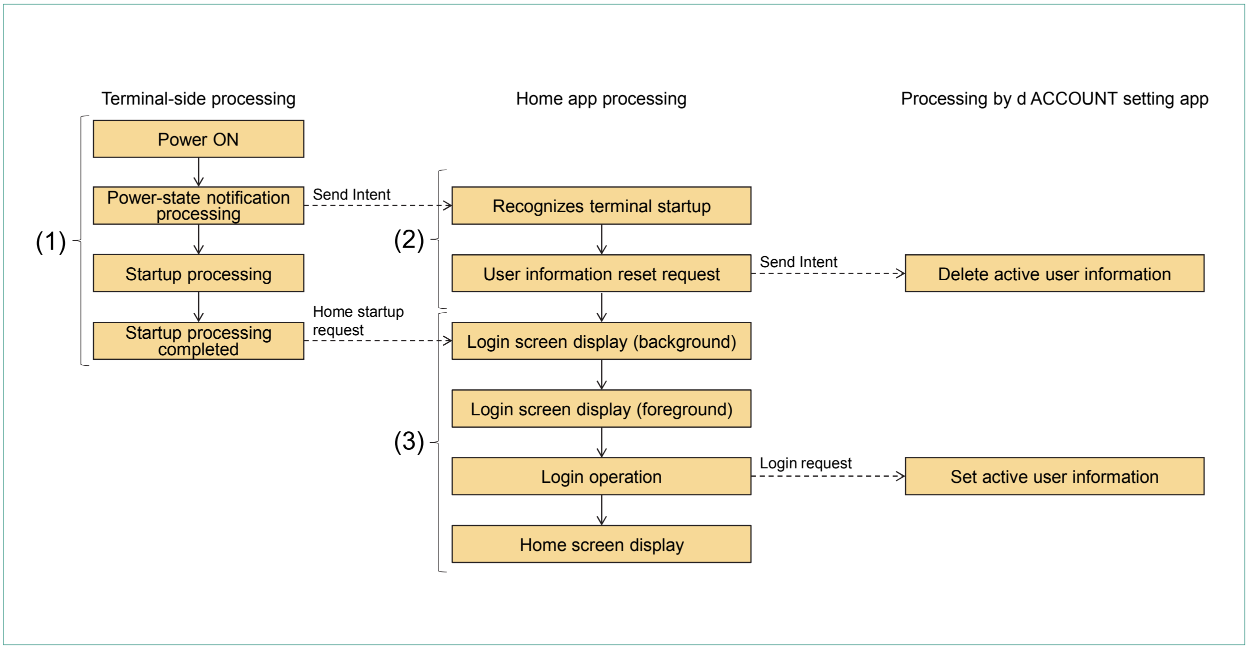

Terminal-side processing
Home app processing
Processing by d ACCOUNT setting app
Power ON
Power-state notification processing
Startup processing
Startup processing completed
(1)
Send Intent
Home startup request
(2)
(3)
Recognizes terminal startup
User information reset request
Login screen display (background)
Login screen display (foreground)
Login operation
Home screen display
Send Intent
Login request
Delete active user information
Set active user information

Figure 6　Terminal startup sequence

application has obtained user information through an interrupt during the user authentication process, this step performs processing that sets active user information to 0. The idea here is to prevent a certain service application from using, for example, the mother’s user information at the instant when the father is trying to log in.

(3) At this point, login processing of the designated account begins. This processing resets active user information thereby enabling each service application to make use of that user information.

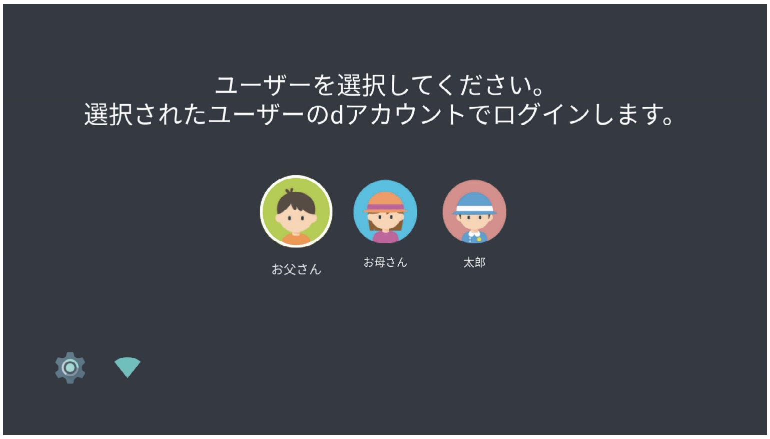
An example of the user experience with the login screen of the docomo TV terminal is shown in **Figure 7**. After the user selects his or her own icon, NTT DOCOMO video services become available and content oriented to that user will be recommended. Additionally, for users that value privacy, specifications call for the input of a password when logging in to an account that the corresponding user wishes to be closed off.

2) Zapping UI

This UI design refers to operations that enable the user to select the content to be used next while actually viewing that content.

In the case of television broadcasts, the numerical buttons on the remote control unit can be used to switch instantly from one program to another. In comparison, a number of operations are needed in the case of video services before viewing can begin. In general, these consist of the following three steps that can feel burdensome to the user: (A) search for video using keywords that come to mind, (B) select the content of interest from the search results presented, and C: decide whether to continue watching the selected content.

*15 Framework: Software that encompasses functionality and control structures generally required for software in a given domain. With a library, the developer calls individual functions, but with a framework, it handles the overall control and calls individual functions added by the developer.



ユーザーを選択してください。
選択されたユーザーのdアカウントでログインします。

*This is provided only in Japanese at present.

Figure 7　Login screen

To eliminate this burden, it was proposed that this product include the experience of video content playback on the home screen.

The user experience with the zapping UI of the docomo TV terminal is shown in Fig. 1. Here, the content most likely to be of interest to the user is displayed in the first focus[16] after starting up the Home app. The playback of that content then begins so that the user can decide whether to continue watching. Then, once the playback of that content completes, the UI automatically advances to the next item of content and begins playback again so that the user can decide in a passive manner whether to continue watching that content. Furthermore, so as not to hinder users who wish to perform operations in an active manner, this UI features a layout that places startup icons of various NTT DOCOMO service applications at the position of the second focus thereby enabling the user to start up a target service in one step.

This zapping UI requires highly accurate recommendations, but since the Home app cannot control the recommendation function provided by the Android standard, such high accuracy is difficult to guarantee. As a consequence, this product is combined with a server that performs integrated analysis of user information to prepare recommendations. The function configuration of this recommendation process is shown in **Figure 8**. In this process, the Home app passes only the d ACCOUNT identifying information to the server as an argument. The server, in turn, uses this information to return optimal recommendations based on the user’s service contract conditions, viewing log, etc.

Incorporating these “multiuser personalization” and “zapping UI” UI designs in the Home app achieves a UI that fulfills the product concept.

5. Osusume hint (Recommended Usage Hints)

5.1 Overview of Osusume hint

The docomo TV terminal incorporates the Osusume hint [3] function first incorporated in smartphone

[16] Focus: Highlighting an icon etc. to confirm a process before inputting or executing.

models of the 2016 summer season. This service displays helpful information on how to use the target terminal in a more comfortable and enjoyable manner according to the usage habits and conditions of each and every customer. In this regard, the docomo TV terminal uses the name “information” instead of “hints” since the plan for the future is to display “information” for various services in the form of an “Information List” in addition to hints on terminal operation (**Figure 9**).

We note here that this is the only application—a NTT DOCOMO original function—in the Android TV OS that encourages terminal operation and usage appropriate to the customer.

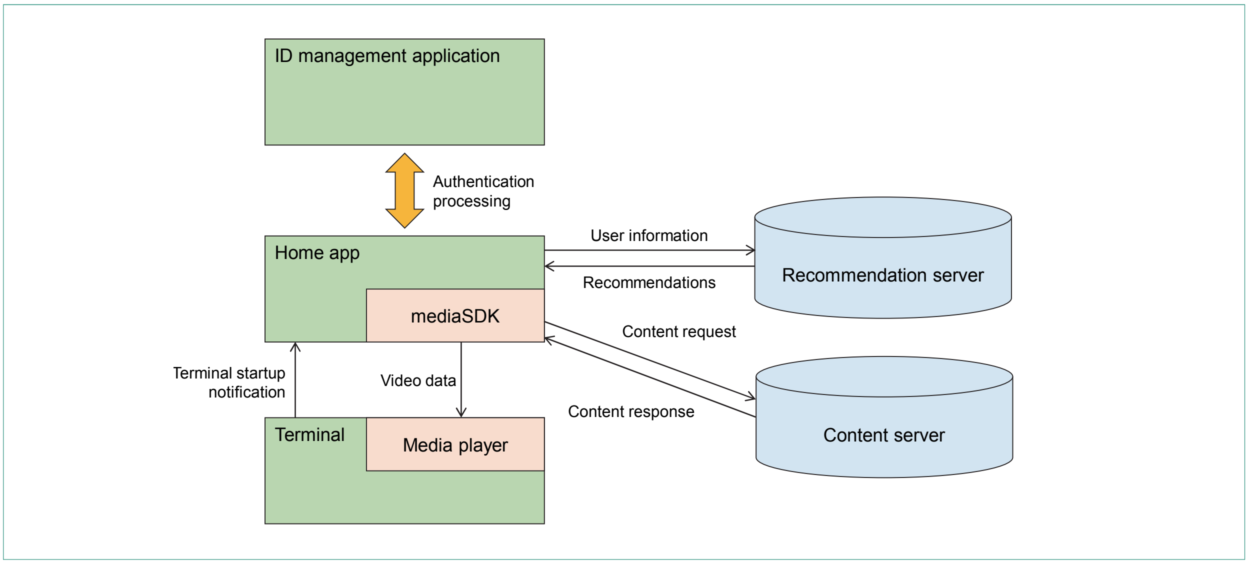
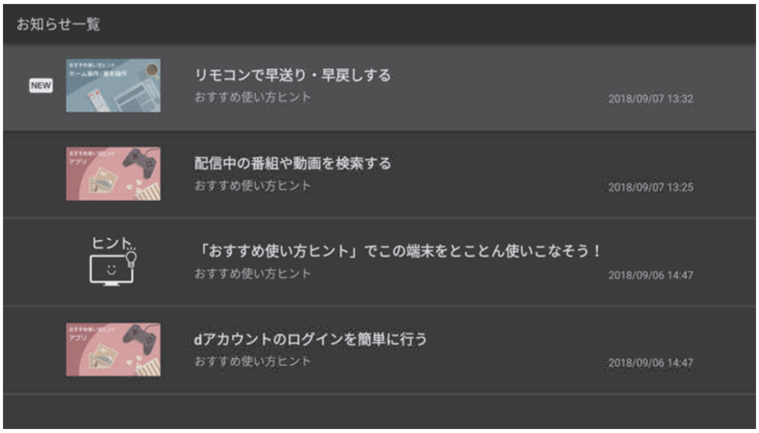


Figure 8　Function configuration chart



お知らせ一覧

NEW リモコンで早送り・早戻しする
おすすめ使い方ヒント　2018/09/07 13:32

配信中の番組や動画を検索する
おすすめ使い方ヒント　2018/09/07 13:25

「おすすめ使い方ヒント」でこの端末をとことん使いこなそう！
おすすめ使い方ヒント　2018/09/06 14:47

dアカウントのログインを簡単に行う
おすすめ使い方ヒント　2018/09/06 14:47

*This is provided only in Japanese at present.

Figure 9　Information list

5.2 Information Display Methods

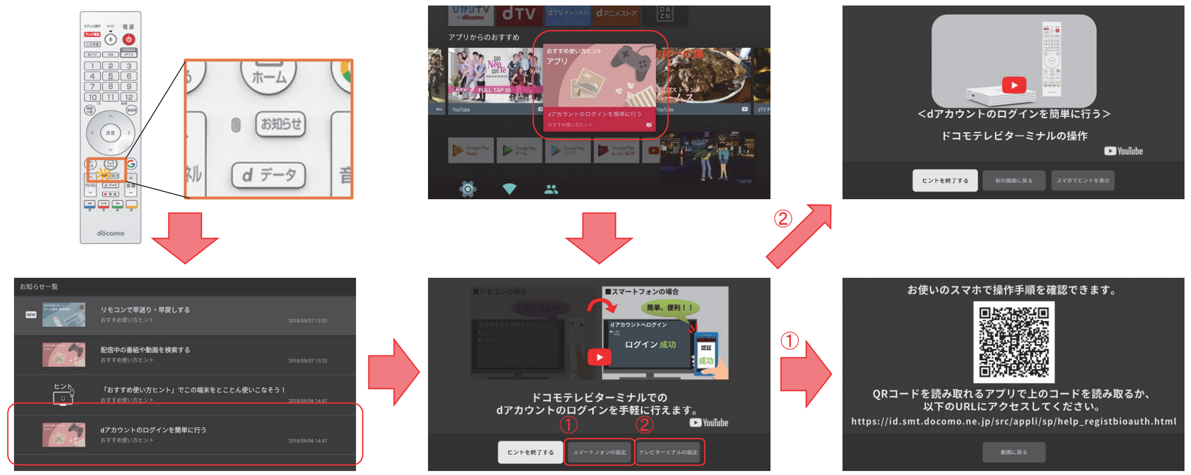
Two methods are provided for viewing helpful information. The first is pressing Information displayed on a recommendation and the second is to press the “Information” button on the remote control and selecting the information desired (**Figure 10**). The information so selected is designed to provide a more intuitive, easy-to-understand information display by playing back a video and encouraging the user to operate the terminal in the way shown. The video shown, which is uploaded on YouTube™*17, is played back on YouTube using WebView*18 from the Osusume hint app screen. However, considering that screen operations cannot be simultaneously performed while watching the video, a QR code®*19 function has been incorporated so that the video can also be viewed on the user’s smartphone for convenience sake.

5.3 Information Updating

The content of Information needs to be modified and optimized whenever a new version of the terminal OS is released or when user usage patterns change. For this reason, the Osusume hint app periodically checks the server for the presence of a new rule-set database and updates the app’s database with the latest information if necessary. In this way, the terminal can always display the latest information. Furthermore, since information content is delivered in accordance with each user’s usage history, switching to another account when multiple d ACCOUNTs are registered will display the information optimized for that account.

5.4 Built-in Information Button

The docomo TV terminal features a built-in “Information” button on the remote control unit to



お知らせ一覧

NEW リモコンで早送り・早戻しする おすすめ使い方ヒント 2018/09/07 13:32

配信中の番組や動画を検索する おすすめ使い方ヒント 2018/09/07 13:25

ヒント 「おすすめ使い方ヒント」でこの端末をとことん使いこなそう！ おすすめ使い方ヒント 2018/09/06 14:47

dアカウントのログインを簡単に行う おすすめ使い方ヒント 2018/09/06 14:47

ドコモテレビターミナルでの
dアカウントのログインを手軽に行えます。

ヒントを終了する | スマートフォンの設定 | テレビターミナルの設定

＜dアカウントのログインを簡単に行う＞

ドコモテレビターミナルの操作

ヒントを終了する | 前の画面に戻る | スマホでヒントを表示

お使いのスマホで操作手順を確認できます。

QRコードを読み取れるアプリで上のコードを読み取るか、
以下のURLにアクセスしてください。
https://id.smt.docomo.ne.jp/src/appli/sp/help_registbioauth.html

動画に戻る

*This is provided only in Japanese at present.

Figure 10 Methods of using Osusume hint

[17] YouTube™: A trademark of Google, LLC.

[18] WebView: A function for displaying a Web page within an application.

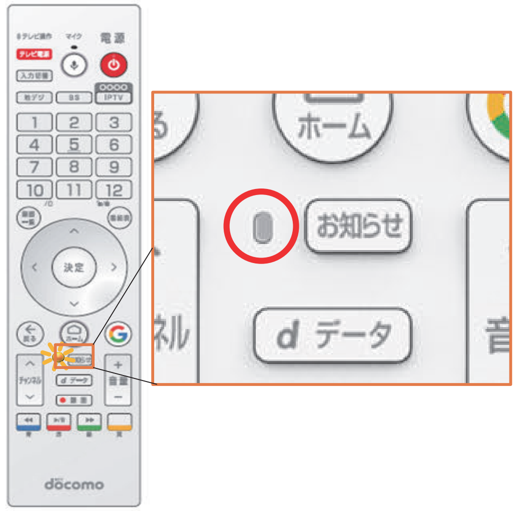
[19] QR code®: A type of two-dimensional bar code. A registered trademark of Denso Wave Incorporated.

make it easy to watch previously displayed information at any time. By pressing this button, the user can view the latest item of information as well as a list of previously displayed information in the form of an “Information History.”

The latest update to docomo TV terminal adds a function for turning on an LED situated next to the “Information” button whenever “important information” is being displayed (**Figure 11**). The purpose of this LED function is to advise the user that this is content that NTT DOCOMO would like the customer to grasp as soon as possible. This function can also be handled from a database. Additionally, for users who were not able to check such “important information” in real time, docomo TV terminal also incorporates a mechanism for lighting the LED as a reminder to prevent that information from being missed. Going forward, the plan is to make the content displayed in “Information” all the more convenient by displaying not only terminal operation hints but also information on various NTT DOCOMO services.

6. Conclusion

This article described in detail the means and technologies used for achieving “Your TV” as one concept of the doocmo TV terminal. Going forward, we plan to study mechanisms for providing new viewing experiences tailored to individual user needs while expanding built-in services not only in smartphones but also in home devices as new user contact points.



*This is provided only in Japanese at present.

Figure 11　Information LED

REFERENCES

[1] NTT DOCOMO: “docomo TV terminal,” (in Japanese). https://www.nttdocomo.co.jp/product/docomo_select/tt01/index.html

[2] NTT DOCOMO: “d ACCOUNT,” (in Japanese). https://id.smt.docomo.ne.jp/

[3] NTT DOCOMO: “Osusume hint,” (in Japanese). https://www.nttdocomo.co.jp/service/osusume_hint/

Technology Reports
IoT Smart Home Web API
Special Articles on Making Life More Convenient and Seamless—toward Future Lifestyles

“IoT Smart Home” Supporting Daily Life Activities

Service Innovation Department Ken Yamashita Takashi Yoshikawa
Takafumi Yamazoe Shoichi Horiguchi Shinichi Mokutani

IoT is attracting much interest in diverse fields. It is seen as promising technology for creating value in various ways such as the remote control of devices and the analysis of collected data. As part of this trend, NTT DOCOMO is focusing on the “home” as the center of life and is constructing an IoT Smart Home®*1 that installs IoT devices throughout the home. The IoT Smart Home performs integrated management of home IoT devices through an IoT access control engine and achieves information sensing and device control within the home. NTT DOCOMO has conducted a demonstration experiment using this IoT Smart Home with the aim of achieving a home that can support a person’s daily life.

1. Introduction

The Internet of Things (IoT) is becoming increasingly known and accepted in society. In the beginning, the targets of IoT were mainly fields such as factory automation and productivity improvement, but more recently, attention has been focusing on use cases more familiar to the general public such as integrated control of home appliances and home security.

Nevertheless, IoT is surrounded by a variety of technical issues such as data volume, communication networks, security, data analysis techniques, and cost [1]. At NTT DOCOMO, we have been promoting research and development in this area focusing particularly on the problem of interconnectivity.

*1 IoT Smart Home®: A trademark or registered trademark of NTT DOCOMO, INC.

At the same time, diverse social issues are making their appearance as a result of Japan's aging society such as an increase in elderly single-person households [2], increase in the population of elderly requiring nursing care [3], and rising healthcare costs [4]. In fact, the difference between average lifespan and healthy lifespan in Japan is widening, which suggests that these social problems will become increasingly severe in the years to come. In the light of the above, NTT DOCOMO is constructing an IoT Smart Home to support people's lives from various perspectives including comfort and health by applying the company's expertise and know-how in IoT technology to the "home" as the center of life.

In the past, the "smart home" was often talked about in terms of power management and energy efficiency using such keywords as Net Zero Energy House (ZEH)[*2], Home Energy Management System (HEMS)[*3], smart meter[*4], and demand response[*5].

In contrast, NTT DOCOMO's IoT Smart Home has been constructed based on the concept that the home can support its occupants from the viewpoints of comfort, health, peace of mind, safety, and beauty. The IoT Smart Home features a mechanism that collects and analyzes everyday life-related data of its occupants through IoT devices in the home and that provides them with an appropriate living space again through IoT devices based on analysis results. The aim here is to achieve a home that supports daily life through this mechanism.

To provide an IoT system that can achieve the IoT Smart Home, we have constructed an IoT access control engine[*6] as a cloud-based platform that can coordinate diverse IoT devices in an integrated manner and have developed a variety of applications. This article describes the IoT Smart Home and the technical features of the IoT access control engine, presents the results of a demonstration experiment using the IoT Smart Home, and touches upon future developments.

2. Configuration of IoT Smart Home

2.1 Overview of IoT Smart Home

NTT DOCOMO has constructed an IoT Smart Home to achieve a home that supports daily life. Exterior and interior views of the IoT Smart Home are shown in **Figure 1**. This IoT Smart Home was implemented as a mobile home that could be towed and moved as desired. As shown in **Figure 2**, the IoT Smart Home embeds a variety of IoT devices that are managed and controlled by NTT DOCOMO's IoT access control engine to create a comfortable and healthy living space.

2.2 System Configuration of IoT Smart Home

The system configuration of the IoT Smart Home

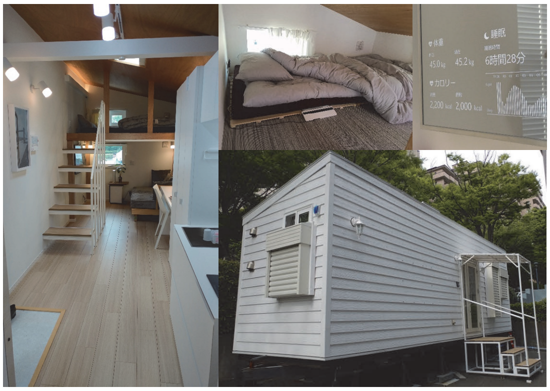


Figure 1 Exterior and interior views of IoT Smart Home

[*2] ZEH: According to the Ministry of Economy, Trade and Industry (METI), a house that can simultaneously achieve a "comfortable indoor environment" and "net energy consumption of zero or less on an annual basis."

[*3] HEMS: A system for visualizing and managing the amount of electric power, gas, and other forms of energy used by home appliances and household equipment through monitors, etc.

[*4] Smart meter: A power meter equipped with a communications function.

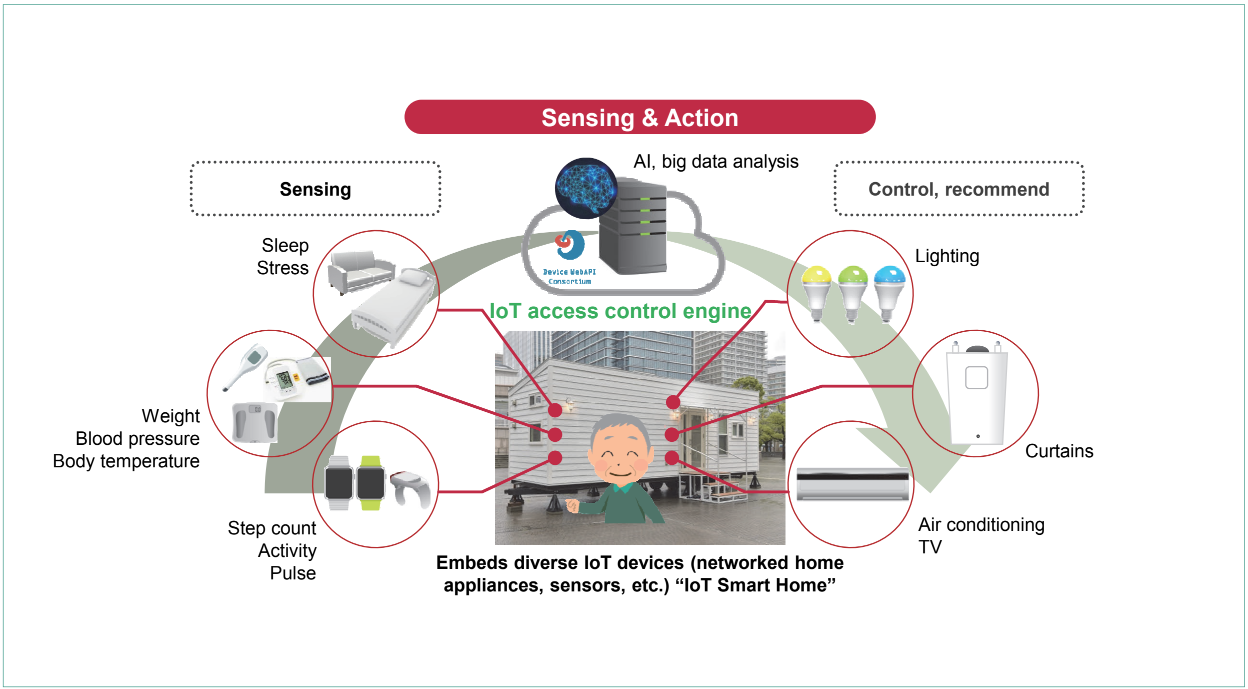


Figure 2　Overview of IoT Smart Home

is shown in **Figure 3**. The IoT Smart Home consists of four main elements: IoT devices, home gateway, IoT access control engine, and user application.

In this article, IoT devices refer to things targeted for control over the Internet and things that can collect information. IoT devices installed in the IoT Smart Home consist of commercially available products such as a body weight scale and lighting and custom-made products such as a smart mirror (a mirror with a head-up display). These IoT devices are managed and controlled by NTT DOCOMO's IoT access control engine. In this regard, many IoT devices perform communications based on Near-Field Communication (NFC) standards such as Bluetooth®*7 and Wi-Fi®*8, so communications between IoT devices in the IoT Smart Home and the cloud are performed via a home gateway installed in the home. The user application is implemented as a Web application that provides functions for controlling IoT devices and visualizing information through the IoT access control engine.

2.3 Connection of IoT Devices through IoT Access Control Engine

The IoT Smart Home currently installs about 20 types of IoT devices. The IoT access control engine developed by NTT DOCOMO manages and controls these IoT devices having different manufacturers and specifications. The IoT Smart Home achieves the five functions described below through use of the IoT access control engine.

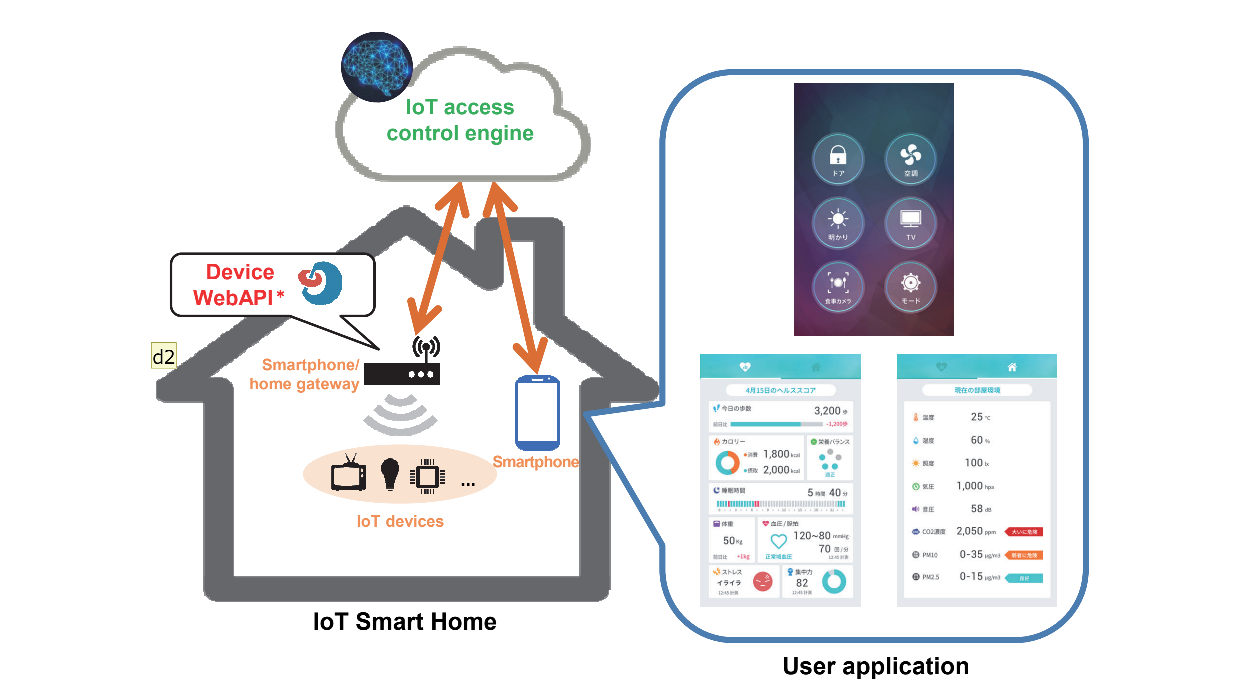
1) Collection of Information from IoT Devices

The IoT access control engine can obtain life-related data collected from various types of IoT sensors via the home gateway equipment in the home. For example, an IoT sleep mat can be placed

[*5] Demand response: According to the Ministry of Economy, Trade and Industry (METI), the adjustment of power demand patterns by holders of consumer energy resources or third parties by controlling those energy resources.

[*6] IoT access control engine: A cloud platform developed by NTT DOCOMO for controlling and managing various type of IoT devices.

[*7] Bluetooth®: A short-range wireless communication standard for interconnecting mobile terminals such as mobile phones and notebook computers. A registered trademark of Bluetooth SIG Inc. in the United States.



*Device WebAPI: Generic name for Web-based APIs used within the API framework GotAPI standardized by Open Mobile Alliance (OMA), an international standardization organization.

Figure 3 System configuration of IoT Smart Home

under bedding to collect data on an occupant's sleep patterns such as the quality of sleep, breathing rate, and frequency of tossing and turning in sleep. In addition, a body weight scale can be imbedded in the floor in front of a washbasin to collect biological data such as weight in a natural, unencumbered manner.

2) Control of IoT Devices

The IoT access control engine enables all sorts of home devices to be controlled. For example, it enables the front door to be locked and unlocked through a smart digital key and infrared-controllable curtains to be opened and closed. Home appliances such as air conditioners and televisions can also be controlled. Furthermore, such IoT devices can not only be individually controlled but also controlled in groups such as by turning off power, shutting down air conditioners, and locking doors before going to bed.

3) Remote Management

Constructed as a cloud system, the IoT access control engine can collect information from IoT devices and control them from remote locations. This makes possible a variety of functions, such as setting the air conditioning system before returning home and checking home conditions from a remote location.

4) Device Extendibility

The IoT access control engine enables supported IoT devices to be added simply by adding plug-in software to the home gateway. This ease of extendibility means that devices within the IoT Smart

*8 Wi-Fi®: The name used for devices that interconnect on a wireless LAN using the IEEE802.11 standard specifications, as recognized by the Wi-Fi Alliance. A registered trademark of the Wi-Fi Alliance.

Home can be extended and modified as needed.

5) Diverse Access Rights Management

The IoT access control engine enables a variety of access rights management functions to be achieved in terms of users, time periods, and functions. In the demonstration experiment that we conducted using the IoT Smart Home, a different subject lives in the house for one week during which time the subject's account is given the right to use IoT devices while the manager is given the right to view IoT device status. In this way, only the subject and manager can access the functions required for the IoT Smart Home.

6) Accumulation of Life Data

All operation history and data logs of IoT devices in the IoT Smart Home are stored in a database within the IoT access control engine. This scheme enables the huge amount of daily life-related data to be automatically accumulated and diverse types of data to be analyzed resulting in value-added data. The aim here is to automatically create a comfortable and healthy living space for the occupant of the IoT Smart Home by analyzing the data collected from IoT devices on the cloud and controlling those devices based on analysis results.

2.4 Installed Devices

The devices currently installed in the IoT Smart Home are listed in **Table 1**.

2.5 Functions

The IoT Smart Home is currently made up of six main functions as summarized below.

1) Visualization of Healthcare Information

Healthcare information collected by various healthcare devices in the IoT Smart Home can be visualized on the occupant's smartphone. In this way, the occupant can view comprehensive healthcare information that combines the data of multiple devices instead of viewing information from individual devices separately.

2) Environment Monitoring

This function provides real-time visualization of indoor/outdoor environment information such as dust concentrations (PM10, PM2.5), temperature/humidity, CO_2 concentrations, and wind direction/speed, plus open/closed status of each door in the home and locations where someone is present. It can be used to watch over family members living separately and to view indoor/outdoor environment information invisible to the naked eye.

3) Smart Mirror

The IoT Smart Home adopts a smart mirror for the washbasin. This mirror can display various types of information for the occupant such as yesterday's and today's body weight, one week's worth of sleep data, and today's weather. A body weight scale is also imbedded in the floor in front of the washbasin. This type of natural information collection and display brings health matters to the attention of the occupant in a casual, trouble-free manner.

4) Meal Analysis

This function can determine the content of a meal and calculate calories and nutritional balance by having the occupant take a photo of the meal with a smartphone camera. It can also offer advice on meals based on the content of one day's worth of meals.

5) Remote Control

IoT devices within the IoT Smart Home can be controlled by smartphone. This function enables

Table 1 Devices installed in IoT Smart Home

IoT device	Communication scheme	Function
Blood pressure monitor	BLE	Measure blood pressure
Body weight scale	BLE	Measure body weight
Sleep meter	Wi-Fi	Measure sleep status
Human sensor	EnOcean	Detects human presence (bedroom, front door, bathroom, sofa)
Door open/closed sensor	EnOcean	Detects whether a door is open or closed (refrigerator, freezer, microwave oven, closet, front door)
Smart wristband	BLE	Measures number of steps, burned calories, etc.
Meal camera (smartphone)	LTE	Estimates meal content, calories, nutritional elements
Dust sensor	BLE	Measures PM10, PM2.5
Light	Wi-Fi	Turns light ON/OFF, change color
Infrared learning remote control	Wi-Fi	Controls air purifier, aroma diffuser, air conditioning, TV, curtains, skylight curtains
CO_2 sensor	EnOcean	Measures CO_2 concentration
Smart lock	BLE	Gets locked/unlocked status and performs locking/unlocking
Power distribution board	Wired LAN (ECHONET Lite)	Measures power
Smart mirror	Wi-Fi	Displays information such as sleep data, body weight, weather, time, etc.
Position detecting floor	BLE	Extracts position of occupant
Shutter	Wi-Fi (ECHONET Lite)	Opens/closes shutter, adjusts angle
Indoor/outdoor environment sensor	Wired LAN	Measures NO, NO_2, SMP, PM2.5, wind direction/ speed, temperature/humidity, HCHO, VOC, CO_2
Health advisor	–	Not yet connected to IoT access control engine
Cosmetic dispenser	–	Not yet connected to IoT access control engine

BLE (Bluetooth® Low Energy): Extended specification of Bluetooth near-field communication standard, added to Bluetooth ver. 4.0. Features low-power communications. Bluetooth is a registered trademark of Bluetooth SIG Inc.

ECHONET® Lite: A communication protocol specified by the ECHONET Consortium mainly for home systems. ECHONET is a registered trademark of ECHONET Consortium.

EnOcean®: A wireless communication technology using the sub-gigahertz band featuring self-powered, battery-free data communications. EnOcean is a registered trademark of EnOcean GmbH.

IoT devices to be controlled separately as well as in groups according to various life scenarios such as when returning home. This capability eliminates the bother of operating the remote control unit for each IoT device.

6) Chatbot Conversation

IoT devices can be controlled and information visualized via a chat-oriented User Interface (UI). This function also supports conversation on other than IoT-related matters, which makes it appear as if the home has a personality of its own.

3. Life Monitoring Demonstration Experiment Using the IoT Smart Home

3.1 Overview of Demonstration Experiment

We conducted a demonstration experiment using the IoT Smart Home to determine whether a home could provide its occupant with a comfortable and healthy space using IoT and AI technologies.

In the experiment, each subject spent one week living in the IoT Smart Home, and at the end of this week, we performed an evaluation to check for any before-and-after changes in the condition, awareness, and behavior of the subject. Other than living in the IoT Smart Home, each subject went about daily life as usual such as going to work or working from home. This experiment has so far been conducted two times at different locations with a total of 20 subjects, each of whom were asked to spend one week living in the IoT Smart Home. **Table 2** provides an overview of this life monitoring experiment.

3.2 Experimental Results (Questionnaire)

We administered a questionnaire to the 20 subjects. The replies from each subject were obtained from a website on the last day of living in the IoT Smart Home for one week. The response rate was 100% (20 out of 20 subjects). Results are shown in **Figure 4**.

As a result of living in the IoT Smart Home, it was found that 75% of responders reported a rise in health awareness and that 65% noticed something about their own state of health. It was also found that living in the IoT Smart Home resulted not only in a change in awareness but also in specific behavioral changes such as greater concern about meals and proactive use of stairs as reflected by statements in the comment field of the questionnaire.

3.3 Experimental Results (Data)

Among the data obtained from the experiment, the human-sensor values for the time that two subjects lived in the IoT Smart Home are visualized

Table 2 Overview of life monitoring experiment

	Time period	Experiment location	No. of subjects
1st session	2017. 12~2018. 02	Sotetsu Rosen Mini Sachigaoka store, parking lot (near Futamatagawa Station)	6
2nd session	2018. 06~2018. 09	Sotetsu Bunka Kaikan, parking lot (near Ryokuen-toshi Station)	14

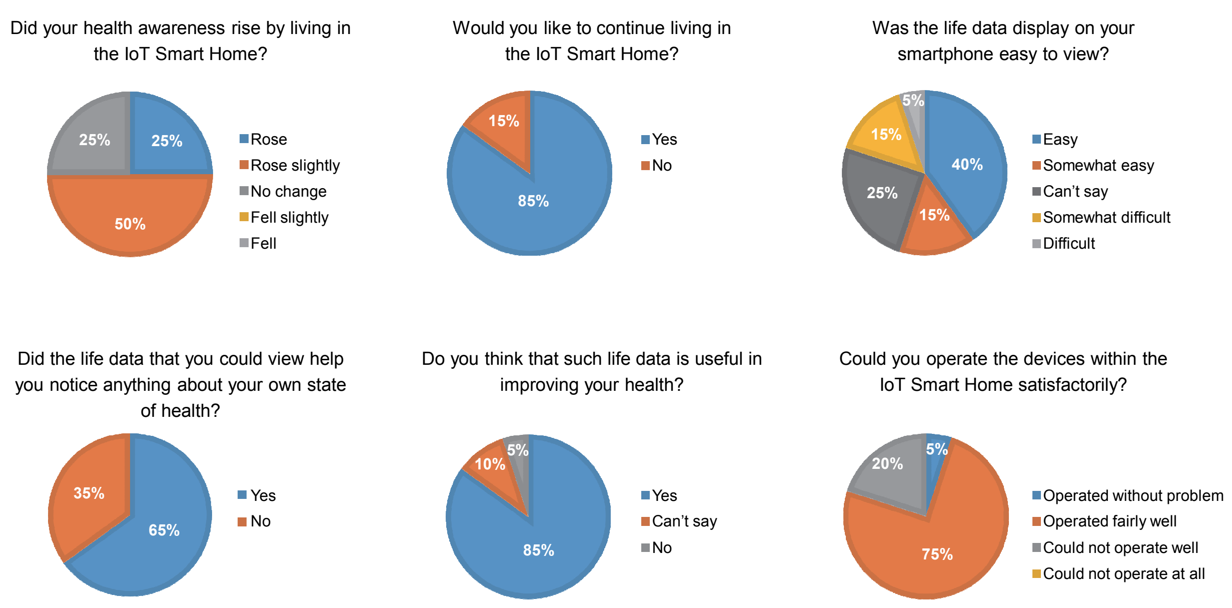


Figure 4　Results of questionnaire

in **Figure 5**. The following three functions can be considered from this data.

1) Extraction of Living Patterns

For subject A, the living pattern that could be observed from one week of data was one of going back and forth between the sofa and front door before sleeping and after getting out of bed. The data for subject B, meanwhile, revealed a pattern of spending some time on the sofa before sleeping but going back and forth between the sofa and front door after getting out of bed similar to subject A. These results suggest the possibility of classifying a person’s daily life in terms of a certain pattern based solely on data obtained from human sensors. In this way, it should be possible to propose control schemes for IoT devices or provide optimal home-automation functions that take daily living patterns into account.

2) Detection of Abnormal Condition

As described in 1) above, the possibility exists that an occupant’s living pattern can be extracted, but conversely, the possibility also exists of detecting movements out of the ordinary, that is, of detecting an abnormal condition. This is a capability that could be applied to keeping watch of family members living separately.

3) Distinguishing Occupants

If we assume that the daily living pattern of an occupant can be extracted as described in 1), it should also be possible to distinguish one occupant from another in the case of a home with multiple occupants. This suggests the technical feasibility

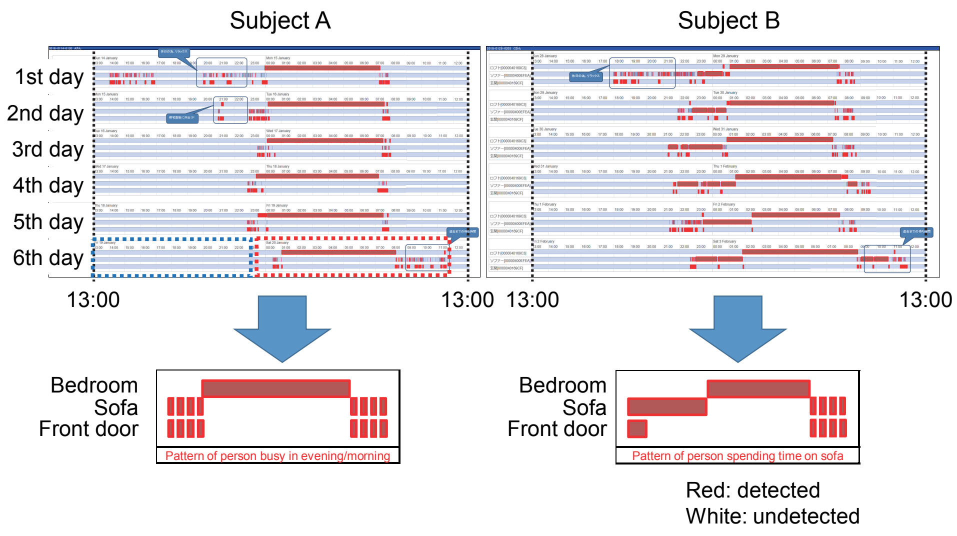


Figure 5　Visualization of human-sensor values

of applying this feature not only to single-person households but also to general households, communal homes, etc.

4. Conclusion

This article described a life-monitoring demonstration experiment using NTT DOCOMO's IoT Smart Home. This experiment demonstrated the possibility of managing a wide variety of IoT devices in an integrated manner through the use of an IoT access control engine and of creating a comfortable and healthy living space for the home's occupant. Next, to prove these hypotheses, we plan to perform data testing using combinations of IoT devices and conduct more demonstration experiments with a greater number of subjects. Finally, we aim to contribute to the solving of diverse social problems in the aging society by constructing a "home that supports daily life," that is, a home that takes on a personality of its own and automatically creates a comfortable and healthy living space by understanding the home's occupants.

REFERENCES

[1] G. D. Abowd: "Software engineering issues for ubiquitous computing," Proc. of the 21st international conference on Software engineering, pp.75–84, ACM, May 1999.

[2] MIC: "2018 White Paper on Information and Communications in Japan," p.151, 2018 (In Japanese).

[3] Cabinet Office, Government of Japan: "2017 Annual Report on the Aging Society," 2017 (In Japanese).

[4] Ministry of Health, Labor and Welfare: "2016 Trends in Medical Expenditures," p.1, Sep. 2017 (In Japanese).

Technology Reports

Personalization Home Devices Gesture Controlled UI

Special Articles on Making Life More Convenient and Seamless—toward Future Lifestyles

Personalized Screen Concept Using a Gesture Controlled UI

Communication Device Development Department Yuki Matsunaga
Consumer Business Department Aya Murakami

Smartphones have become widespread as devices for collecting a wide range of data, but users can feel stress at not always being able to get the information they need accurately. It can require significant time and effort to get such information. To solve this issue, NTT DOCOMO has proposed the “Personalized Screen” concept, which provides information in a natural form, even without the user actively try to get it.

1. Introduction

Mobile phones are currently the most widespread information devices and are used very frequently. Smartphones in particular have reached a penetration of 75.1% of all households in Japan [1]. They have become indispensable to users for getting all kinds of information and for communicating with others through SNS. On the other hand, an issue with smartphones is that users can feel stress if there are obstacles to communicating or getting information.

In a survey of smartphone users ranging from 9 to 25 years old, 35.4% reported, “I would say that I am dependent on the Internet” [2]. Another study reported that 47.4% of smartphone users reported, “I feel uncomfortable when I cannot use my smartphone as often as usual” [3]. Users cannot be without their smartphones, even when they are at home.

Users are able to get information using their smartphones anywhere and at any time, but there is a huge amount of information, and they must make appropriate choices to get the information they desire. Users must actively access information with a smartphone, performing searches and other operations, and this can require significant time and effort to obtain the information they want. For these reasons, some users subscribe to information distribution services so they can passively receive the information they need, or they receive notifications from specialized applications, but this does not necessarily always get them the information they want. In another survey, 33.4% of smartphone users reported that information services such as e-mail magazines that they subscribed to had become "Annoying" or "No longer useful" [4].

Another issue that has been identified with smartphones is that many users feel that the screen is too small for collecting information. 36.7% of users reported, "The screen is too small," when browsing on their smartphones [5], indicating that they would like a larger screen to display more information at once.

To reduce such stress felt by users, the information they desire needs to be displayed appropriately and on a large screen, so that they can enjoy it more passively. However, it is fundamentally difficult to provide appropriate information on a large screen as described above while away from home, so smartphones, which are very portable, are indispensable as information gathering tools.

Issues and stress in obtaining information as described above also occur within users' own homes. As such, NTT DOCOMO has proposed the "Personalized Screen" concept for home environments, as a solution that can display the desired information naturally, without requiring active effort from the user.

Personalized Screen is a home device designed to integrate comfortably with information gathering behaviors in the flow of users' daily lives. It enables them to obtain information without stress, and realizes a user experience which presents them with the desired information when it occurs to them, while they are relaxing at home. Another important element is to provide a User Interface (UI)*1 that is comfortable to operate and enables users to get more detailed information on their interests and concerns.

NTT DOCOMO has proposed other home devices on the theme of improving communication, such as petoco*2 and Tomokaku*3 [6] [7], but Personalized Screen is a new device concept, oriented to improving lifestyle environments.

This article introduces the Personalized Screen concept and a prototype, and describes evaluation of the device by users.

2. Personalized Screen Concept

The premise of Personalized Screen is to enable users to obtain information passively, so a key point is to display desired information naturally, without requiring users to actively retrieve the information themselves. Doing so, requires consideration of what information to display and how.

A comfortable way to operate the system is also necessary, to enable users to get more detailed information on their interests or concerns if they want to. As such, the following three points are important in realizing the Personalized Screen concept.

*1 UI: Operation screen and operation method for exchanging information between the user and computer.

*2 petoco: A home communication device developed, mass produced and marketed by E3 Inc., using technology developed by and licensed from NTT DOCOMO.

*3 Tomokaku: A handwritten communication concept proposed by NTT DOCOMO.

(1) Display of information is integrated smoothly into the flow of daily life

(2) Information is personalized

(3) A comfortable method of operation

2.1 Information Display Integrated Smoothly into the Flow of Daily Life

To integrate smoothly into their daily lives and provide them with information naturally, without causing stress, the desired information needs to be displayed at the desired location. A large screen is also preferable, so that more information can be displayed at once. There are many flat surfaces in a home, such as walls, tables, windows, and mirrors, but in most cases they serve a single purpose and are not used for displaying information. With Personalized Screen, these types of large surface, which exist naturally in our living environments, are used to display information (**Figure 1**).

Scenarios will differ for each user, but possible examples include using a living room wall or large-screen television to display information for the whole family, and walls in the bathroom, kitchen or bedrooms to display more personalized information. The display location is not restricted and is selectable by users, enabling them to get information naturally, within the flow of their daily lives.

2.2 Personalizing Information

The information users desire varies greatly, and can change according to place and time, even for the same user. To provide information to users without stress, it is important to discern the user and scenario, and optimize the information displayed from one minute to the next.

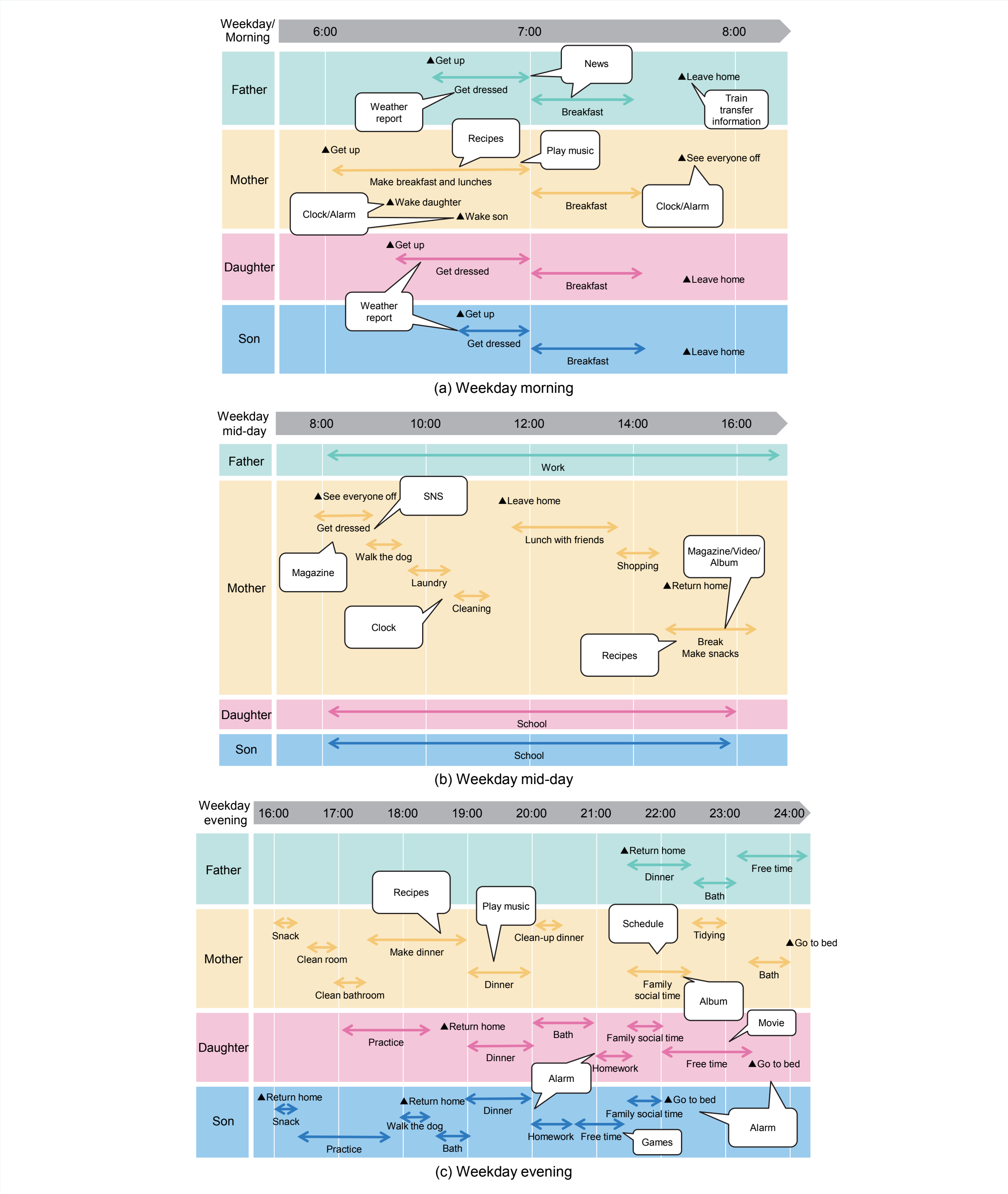
As an example, consider a family of four, with the father working in an office during the day, the mother a homemaker, a daughter in middle school and a son in elementary school. An example of their pattern of behavior and the information they need is shown in **Figure 2**. To give the whole family the information they need without causing stress, the information displayed must change automatically, without operation by family members. For example, weather forecasts could be displayed when the whole family is present, news and commuting information for the father, and recipes and alarms for the mother. During the day when only the mother

(a) Display on a wall

(b) Display in a window

(c) Display on a desk

Figure 1　Use cases for the Personalized Screen



(a) Weekday morning

(b) Weekday mid-day

(c) Weekday evening

Figure 2　Example of family behavior patterns and required information

is home and the rest of the family are out, the information is specialized for her, which may include SNS, magazines, or movie information. Then, in the evening, the system resumes displaying information for the whole family again. The son and daughter can also continue enjoying entertainment content they were viewing on their smartphones, such as a game or a movie. And of course, the information needed will be different on the weekend than on weekdays.

The ability to display information optimized for individual or multiple users, and for the day and time, is an important element of Personalized Screen.

2.3 Comfortable Control Operations

Users will be able to get information they need from Personalized Screen without stress as described above, but they also need a way to get more detailed information about their interests and concerns comfortably.

Input InterFaces (IF) currently in common use include remote controls for television and other household electronics, mouse and keyboard for personal computers, and touch screens for smartphones. However, users must always have the remote control or mouse at hand in order to control a device. For touch screen operations, users need to touch the screen directly, which is difficult when far from the screen, or for larger screens that may not be within reach. There is also a risk that a device could be damaged when touched directly. All of these input interfaces require troublesome actions such as searching for the remote, or approaching the device to perform operations. When users cannot reach the controller or device, such as while cooking or taking a bath, they cannot perform operations at all.

As such, we have proposed using gestures as a comfortable mode of operation. Gestures allow operation without a controller and from a distance, so they support a range of use cases within a home, and are an optimal input interface for a home device integrated smoothly into the flow of daily life.

3. Prototype and Evaluation

To evaluate the Personalized Screen concept, we prototyped an application capable of gesture operations. We conducted user reviews and exhibits projecting images onto a large screen and obtained feedback from users. We describe the prototype and evaluation results below.

3.1 Information Display Functionality

When the application is launched, the home screen is displayed. It then updates, periodically and automatically, arranging various content on the screen so that users can get information by just looking at it. The system uses a camera with facial recognition to determine the number of users and differentiate them, to provide content that is expected to be of interest to the users who are present. For example, if several users are present, content of common interest, such as news or a shared scheduler could be displayed, but for a single user, content such as their SNS feed could be displayed.

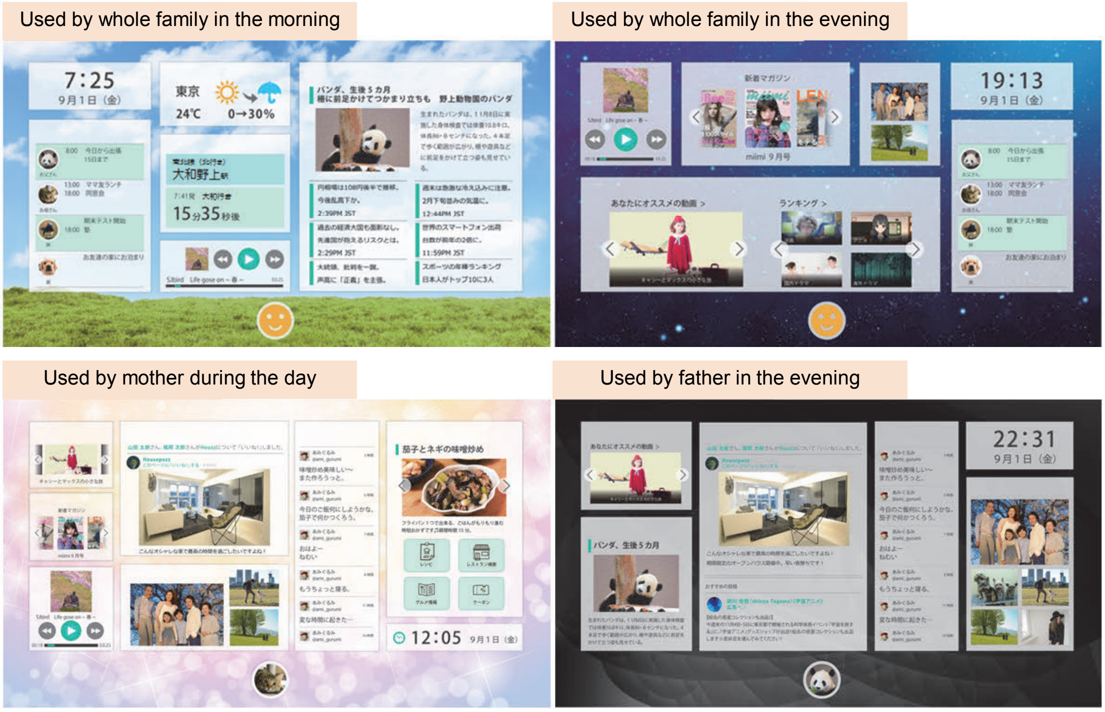
The system can also display content synchronized with a user's smartphone history, such as the continuation of a movie that was being watched on a smartphone, or it can notify of updates to products that the user searched for earlier.

To enable users to get more detailed information regarding their interests or concerns, operations can also be done on any of the content on the home screen. By waving a hand up and down or left and right in front of the gesture recognition camera, a cursor can be moved over the content. Then, for example, more information can be found by scrolling on the news screen, or playback of music or video can be started or paused. When any of the content is selected, it is launched and displayed using the whole screen, so that the user can browse content or watch video on the large screen.

Images of actual screens are shown in **Figure 3**. They represent weekday scenarios for the family described in Section 2.2. The screen for the whole family in the morning is in the upper-left, the family screen in the evening in the upper-right, the mother's daytime screen in the lower-left, and the father's evening screen in the lower-right.

3.2 Gesture Operation

We used gestures for the input interface, as a comfortable way of operating Personalized Screen. In the process of testing the prototype, we learned that to implement comfortable gesture operations, we needed to define a gesture scheme and improve the accuracy of gesture recognition, but we would


Used by whole family in the morning

Used by whole family in the evening

Used by mother during the day

Used by father in the evening

Figure 3　Prototype application screen shots (content depends on the user and the time of day)

also need to optimize the screen display for gesture operations. Thus, for this prototype, we optimized both the gestures and the screen display as described below.

The greatest difficulty we found with gesture operations was that behaviors unintended by the user were selected and performed. To prevent this with the prototype, we first recognize the user's palm, and then only allow three different gestures, which were grasping, waving the hand left and right, and moving the hand up, down, left and right. We did not define gestures that were similar, making it more difficult to recognize gestures incorrectly and perform unintended operations, and we required recognition of the palm before accepting operations to reduce the margin for such operations.

We also did not display a pointer on the screen, and used a cursor to indicate which content to operate on. This was because compared to using a mouse or touch screen, we found that it is difficult to perform detailed operations using gestures, and that inconsistencies between the pointer and their sense of their own body made performing operations tiring. We also arranged the home screen with a small number of large rectangles rather than many small icons, as on a smartphone home screen, to avoid the need for fine operations.

We also played sound effects and magnified the display of the content indicated as the cursor was moved by the user. This was done to provide feedback to the user, because gestures do not provide feedback in the way that touching a screen or clicking a mouse does.

We also defined gestures to start and stop scrolling on the screen. We did this rather than defining screen scrolling operations as on a smartphone, because such gestures would require large movements with the whole arm, which could cause fatigue. We were able to reduce such fatigue by designing the prototype to continue scrolling automatically once started.

Finally, we improved operability by confining the recognizable motions to the 2D plane of up, down, left and right. Initially we studied gestures with motion in 3D, such as selecting content by pushing your hand forward in the depth direction. However, unlike actions in the real world, using a gesture does not provide any sensation of actually touching the content, so users had no intuitive sense of how far they needed to move their hands to perform an operation. This made it very difficult to control. Users also actually moved their hands forward or back, even when they only intended to move them in the up/down/left/right plane, and this caused them to select content that they had not intended. For these reasons, we decided that using 2D operations was optimal.

3.3 Evaluation

The prototype described above was exhibited at an event titled, "Encounter the revelation of "near" future: - 5G creates lifestyles of future -" jointly held by NTT DOCOMO and the National Museum of Emerging Science and Innovation ("Miraikan") from November 9 to 11, 2017. Approximately 600 users were able to experience our concept. We conducted a user survey at the same time, and discuss the resulting comments and evaluation below.

1) Evaluation of the Personalized Screen Concept

(1) Favorable comments

We received many favorable responses. On opinion was that the idea of automatically displaying information optimized to the user's needs while synchronizing with their smartphone was very interesting. Another user said that they only watch the television in the morning to see the time, so displaying other useful information in this way was very good. We received positive comments regarding the concept from almost all participants.

(2) Points for improvement

Some users expressed concern for privacy, because the concept is that the content displayed depends on whether the display is shared or used by one person. For example, the system could display on a large screen personal information such as text messages that the user does not want to share with other people in the room.

2) Evaluation of Gesture Control

(1) Favorable comments

There were many comments saying that the system was easier to operate than expected, and that being able to operate it remotely using gestures felt futuristic. Many participants agreed that being able to operate a large screen from a distance was good, confirming that using gestures for operation fit very well with our concept.

We also received a comment that the concept could be applied as a solution for the elderly, due to its simple, intuitive operation.

(2) Points for improvement

We received mostly positive responses regarding usability, because we worked hard to address issues with gestures, but we still received a few comments such as, "It didn't always work as I expected," and "I found it tiring," when compared with mouse or touch operations.

3) Evaluation Summary

Generally, we received positive responses to our Personalized Screen concept, but improving personalization and protecting privacy remain as issues, so we want to continue studying improvements in these areas.

We also confirmed the effectiveness of operation using gestures. On the other hand, although we designed the system with consideration for preventing operation errors and fatigue, we also received a few comments that this was not adequate, so more work to improve operability is needed.

4. Conclusion

We have described the Personalized Screen concept, which is a home device that is able to display information needed by users naturally, integrating smoothly into the flow of their daily lives. The positive response it received at "Encounter the revelation of "near" future: - 5G creates lifestyles of future -," exhibition and positive evaluations from users suggest that users were receptive to the concept. In the future, we will continue study of issues identified in the evaluation, working toward commercialization of the concept, and also examine applications in domains other than home devices.

REFERENCES

[1] Ministry of Internal Affairs and Communications: "Results of 2018 Survey of Telecommunication Usage Trends (June 22, 2018 Rev.)," May 2018.
http://www.soumu.go.jp/johotsusintokei/statistics/data/180525_1.pdf

[2] Ministry of Internal Affairs and Communications: "Institute for Information and Communications Policy – Survey of Internet Use and Dependency Trends among Youth," Jun. 2013.
http://www.soumu.go.jp/iicp/chousakenkyu/data/research/survey/telecom/2013/internet-addiction.pdf

[3] Just Systems Inc.: "Field Survey of Smartphone Dependency," Sep. 2012.
https://marketing-rc.com/?_ppp=8a9cfe684b&c=881515ae4823e912-00&p=3617&preview=1

[4] PR TIMES: "Report on Survey of Smartphone Notifications. 70% of People Check Push Notifications! 60% of People Store up e-mail newsletters without reading them! | Emprize Press Release," Jun. 2015.
https://prtimes.jp/main/html/rd/p/000000008.000012737.html

[5] Impress: "Gender differences in dissatisfaction with Smartphone Web browsing; Improvements must match the target users/Smartphone Report Vol. 5-1 | Smartphone Report | Web Administrators Forum," Feb. 2013.
https://webtan.impress.co.jp/e/2013/07/10/15118

[6] K. Murakami: "petoco": A Home Communication Device," NTT DOCOMO Technical Journal, Vol.20, No.2, pp.32–41, Nov. 2018.

[7] K. Ishiguro, et al.: "Tomokaku": A Handwritten Communication Concept," NTT DOCOMO Technical Journal, Vol.20, No.2, pp.42–51, Nov. 2018.

Technology Reports

Eye Contact　Face-to-face　Video Calling

Special Articles on Making Life More Convenient and Seamless—toward Future Lifestyles

A Face-to-face Video Calling System Facilitating Natural Eye Contact

Communication Device Development Department　Shinji Kimura　Eriko Ooseki

Advances in information and communication technology have promoted the spread of video calling, and it is becoming more important to improve the user experience by enhancing a sense of being face-to-face. Improving image quality, increasing the screen size and many other technologies contribute to this, but to provide a commercial service will require a system that can realize a sense of being face-to-face that is adequate on a practical level, at a reasonable cost. NTT DOCOMO has evaluated the contribution of various video conferencing parameters on this face-to-face sense, identified eye contact as having a particularly important contribution, and developed a video calling system that achieves eye contact, based on a front-and-center image capture technology. This article describes details of the system.

1. Introduction

With advancements in devices such as cameras and displays, increasing network speeds, and the spread of tools such as smartphones and PCs, video calling between distant locations has become common for casual communication among friends and also for meetings in business. In such video calling, achieving a sense that the other person is actually in the same room with you and having conversations, which we are calling a "face-to-face sense," is a major goal in enhancing the user experience, but it is difficult to say we have reached a point where it can completely replace actual face-to-face conversations and meetings.

Beyond psychological and cultural reasons, this is because, compared with actually being face-to-face, (1) we do not feel a sense of realism or presence from our counterpart, and (2) it is difficult to reach mutual understanding smoothly. The former is considered to be due to deficiencies in video quality and 3D perception [1], and the lack of eye contact is understood to have a large effect on the latter [2].

The speed of networks will continue to increase with 5G in the future, high quality video will be sent back and forth, and we can expect real commercial video calling services [3] with a strong face-to-face sense to be realized. On the other hand, considering the technical feasibility and cost of a commercial service, we cannot expect all of the deficient elements of video calling systems described above to be completed perfectly, and a system that realizes a practically face-to-face sense that is sufficient, at a reasonable cost, is needed. As such, NTT DOCOMO has identified the contribution of various video conferencing parameters enhancing face-to-face sense, and has developed a system that addresses elements that have a particularly large contribution.

This article describes evaluation experiments conducted in preparation for building a video calling system, the system developed based on the results of those experiments, and extended functions added to promote active communication.

2. Evaluation Experiments

Before building the system, we evaluated the degree of contribution of various video conferencing parameters on face-to-face sense, to identify parameters that need to be prioritized.

2.1 Evaluation Procedure

For these evaluations, we defined face-to-face sense as "the feeling of being in the same place and conversing with a friend or family member in ordinary communication," and evaluated the degree to which subjects felt a face-to-face sense from evaluation video. To obtain stable evaluation results, we compared reference conditions (video conditions presumed to yield the highest face-to-face sense among all patterns) with other patterns, varying each of the parameters. We used a nine-step Likert scale[*1] to compute a Mean Opinion Score (MOS)[*2]. Our subjects were members of the public aged 18 to 30, 20 males and 20 females. During evaluation, video was viewed from a distance of 1.5 m. Four parameters that have been shown to increase face-to-face sense in earlier research [1] [2] were varied, as shown in **Table 1**. Subjects watched a total of 168 patterns of video

Table 1 Parameters varied during evaluation

Parameter type	Conditions of variation
Resolution (horizontal) (pix)	1,920*, 1,440, 1,280, 960, 540, 480
Person display scale (%)	100*, 67, 50, 33
Size of projected image (in)	100*, 75, 55, 42, 32
Line-of-sight mismatch (cm)	0*, 10, 20, 30, 40

*Reference conditions

[*1] Likert scale: A type of response metric for psychological testing and used in surveys and other types of studies. A statement is presented to subjects and they indicate the degree to which they agree with the statement. Generally, the scale has five steps, but seven and nine step scales are also used.

[*2] MOS: A widely used measure of subjective quality representing the average value of subjective evaluations given by multiple subjects.

with three different actors and composed of 56 patterns that varied either a single parameter or two parameters at once from the reference conditions. Subjects then evaluated face-to-face sense when compared with the reference video. Photos of the actual evaluation are shown in **Photo 1**.

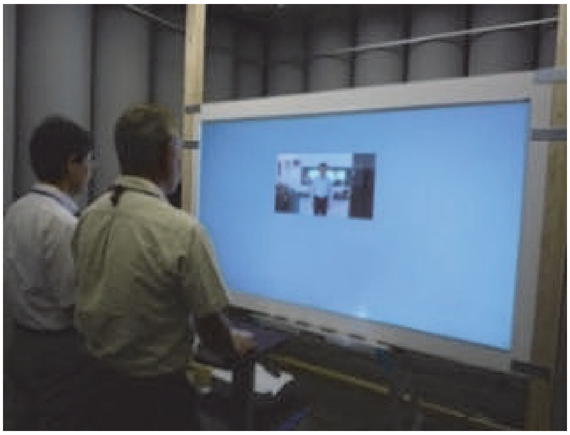
2.2 Evaluation Results

To derive the rate of contribution to face-to-face sense for each parameter, we conducted a multiple regression analysis*3. The results, shown in **Figure 1** (a), had a determination coefficient*4, R^2, of 0.86 (≥0.8), indicating that we were able to estimate face-to-face sense using a multiple regression equation (a model equation) that was highly correlated to actual evaluation values. Deriving the rates of contribution to face-to-face sense for each parameter from this multiple regression equation yielded, in decreasing order: display scale of person (33.0%), line-of-sight mismatch (26.0%), projection screen size (25.7%), and resolution (15.3%). To measure receptivity to such systems as a service, we also asked subjects whether or not they would use it as a tool for everyday communication with friends and family. The results, shown in Fig. 1 (b), had a determination coefficient, R^2, of 0.82 (≥0.8), with contribution rates in decreasing order of: line-of-sight mismatch (33.4%), display scale of person (30.0%), projection screen size (19.3%), and resolution (17.3%). These results show that reducing the amount of line-of-sight mismatch (i.e.: enhancing eye contact) and implementing person display scale closer to life-size will have a greater effect on increasing the face-to-face sense and acceptability of a video call system than increasing the projection screen size or resolution.

3. Video Calling System Capable of Front-and-center Imaging

To display at life-sized scale, there are methods that extract the person from video in real time and change the scale to maintain life-size, even if the person moves [4]. However, we did not use such a method for our system. We assumed that both parties would stay at a fixed distance, and

(a) Reference conditions

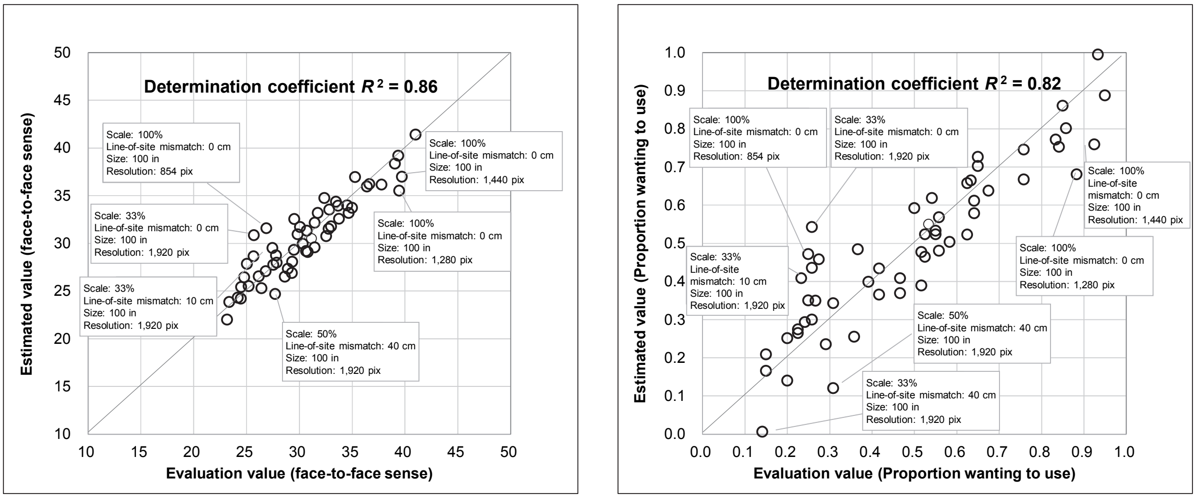


(b) Person display scale: 33%, projected size: 32 in

Photo 1　Evaluation conditions

*3　**Multiple regression analysis:** A data analysis method that attempts to predict a single objective variable using a linear combination of multiple explanatory variables. This predictive equation is called a multiple-regression equation (or model equation).

*4　**Determination coefficient:** An index of the correlation between values estimated by a multiple regression equation and real values. Generally, if the determination coefficient is 0.8 or greater, we say that the predicted and measured values are highly correlated, meaning that the regression equation predicts measured values accurately.



(a) Relation between evaluation and estimated values: Face-to-face sense

(b) Relation between evaluation and estimated values: Service acceptance

*Nine-step evaluation converted to score from 10 to 50 (standard=40 points, 5-point increments)

Figure 1　Graph of Face-to-face sense and service acceptance

arranged the camera to have a viewing angle to display the image at a life-sized scale, taking the size of the display screen into consideration. We also studied projection and display systems for a front-and-center camera capture technology that would realize eye contact (able to display the image of the other user on a screen, while also capturing a front-and-center image of the user looking at the screen), and built such a system.

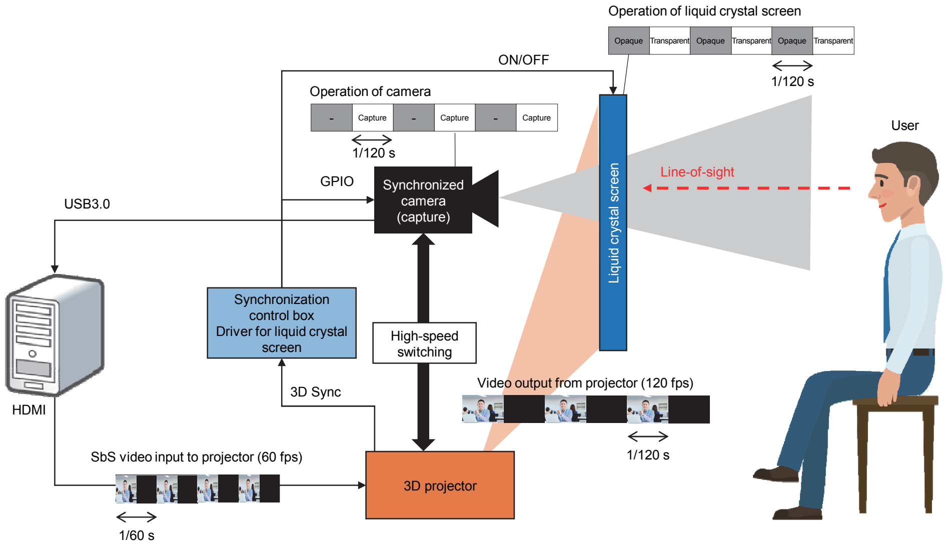
Note that we used Web Real-Time Communication (WebRTC) software, which implements video calling in a browser, to implement video calling.

3.1　Projection Methods

Projection schemes, which use a projector and a screen to display video, have the benefit that it is easy to expand to a large screen. One way to realize front-and-center image capture with a projection scheme is to use an liquid crystal screen with time-division processing[*5] [5]. To implement time multiplexing with ordinary devices, we built a system using a projector capable of 3D stereoscopic display, light-modulating glass capable of switching electronically between transparent and opaque, and a camera with externally controllable shutter timing. With a screen of light-modulating glass, brightness and detail of the image are better when projected from the rear, so we used an ultra-short focus projector to reduce the overall depth of the system, and to make installation easier.

The front-and-center capture system is shown in **Figure 2**. The 3D projector can project 120 fps

[*5] Time-division processing: For projection or capture of 3D video, this is a method whereby the images for the left and right eye are both projected by partitioning along the time axis. Other ways of projecting 3D images include partitioning spatially using deflection, and partitioning by frequency.



GPIO: General-Purpose Input/Output

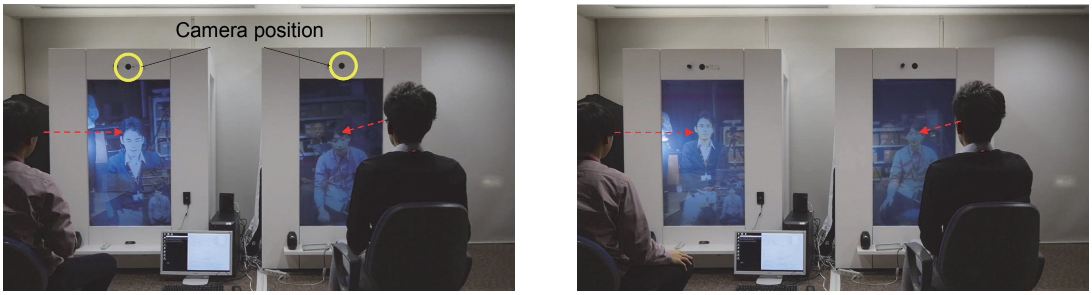
Figure 2 Face-to-face projection mechanism using time partitioning

video by inputting 60 fps Side-by-Side (SbS)*6 format video. Humans perceive any flashing over about 50 Hz as always-on, because this exceeds the flicker-fusion frequency. Thus, for this system, we prepared SbS video with the calling video beside a black screen as input to the projector, and projected it in 3D display mode, simulating 60 Hz video. In 3D display mode, a signal for synchronizing multiple projectors (3D Sync) is output 60 times per second, so by inputting this signal to both the camera and the screen, camera capture can be done by making the screen opaque when the projector is projecting the image, and making it transparent when the projector is not projecting the image (it is projecting a black image). With this time-division processing, the user can see the image projected on the screen without perceiving flicker, while the camera positioned behind the screen can take front-and-center video of the user. An image of actually using this system for a video call is shown in **Photo 2**. The system realizes conversation with the remote person, while looking at them displayed at near life-sized scale, and facilitates more natural eye contact than is possible with existing systems that capture video from peripheral cameras.

3.2 Display Method

We confirmed the effects of front-and-center capture with this screen method, but the method uses a rear-projection scheme, so a certain amount of space is needed behind the screen. Also, due to the time-division processing, the brightness of the projected video is theoretically half that of normal projection without time-division processing. The

*6 SbS: A format that includes two different images within a single video frame by reducing the horizontal resolution to half of the original images and aligning them beside each other within the frame. Used mainly for 3D display, to accommodate images for the left (L) and right (R) eyes within a single frame.



(a) Capture with camera above screen (existing system)

(b) Capture with face-to-face camera (this system)

Photo 2 Face-to-face projection effect

response of the liquid crystal screen switching between transparent and opaque, and measures taken in the projector to prevent crosstalk*7 in 3D display mode also contribute to reducing brightness, so that in practical terms, it is approximately one quarter that of normal projection. Thus, the displayed video was darker, reducing the sense of presence. As such, a system capable of front-and-center video capture, while displaying brighter video and using less space was needed. To achieve this, we developed a system using a transparent Organic Light Emitting Diode (OLED) display.

The transparent OLED display emits its own light and images are very bright when displaying (emitting), while the display has high transparency of approximately 40% when not displaying an image (non-emitting). The transparent OLED we used for our system also has very directional light emission, providing a field of view of approximately 180 degrees to the front, while the image is almost invisible from the rear. As such, front-and-center video capture can be done even if time-division processing is not used, by simply placing the camera behind the transparent OLED display. Note that we do not need the OLED to be transparent from the user side (the side viewing the video), so we covered the back of the display with a black mask, except in front of the camera, to improve contrast when displaying video and to reduce awareness of the camera for users of the system. An overview of the system is shown in **Figure 3** and a view of the system in use is in **Photo 3**. It shows how the system saves space and realizes front-and-center video capture, while displaying a brighter image compared to the projection method.

4. Facilitating Active Communication

The objective of video calling systems is to improve a sense of presence and enable parties to understand each other with less effort, facilitating communication between remote locations. The ability to share an experience, such as looking at a photograph together, is an important element in promoting active communication. To facilitate such shared experiences, we implemented a function that links with an application on a participant's smartphone, and shares photographs from the smartphone through the system. To enhance the sense that users are looking at the same photograph when sharing it, the photograph is shown in mirror image to one user (left-right reversed).

*7 Crosstalk: Ideally for 3D display, the right eye sees only the right-eye image, and the left eye sees only the left-eye image, but in some cases, the right eye may see the left-eye image and vice versa. This occurrence is called cross talk, and can be a cause of motion sickness or fatigue. 3D display projectors take various measures to reduce cross talk due to low LCD response times in 3D glasses, such as inserting black frames while switching between left-eye and right-eye images. These measures can result in the 3D display being less than half of the brightness of 2D display.

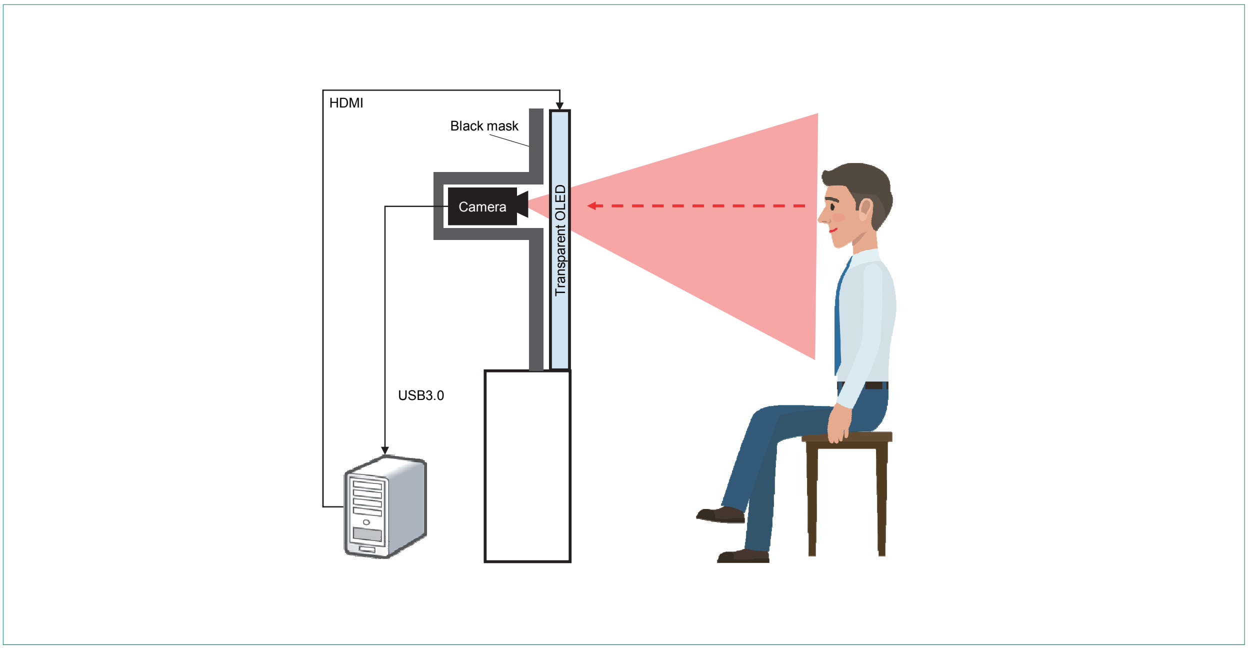


Figure 3　Display method

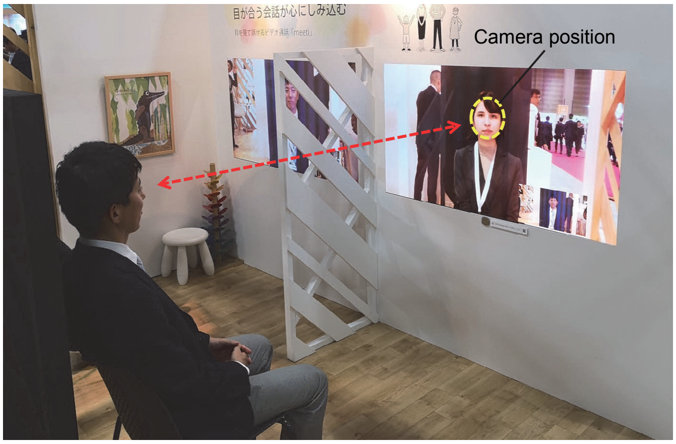


Photo 3　Using the display format system

An example of this is shown in **Photo 4**.

To further promote active communication requires systems with extended functionality to meet a wide range of user needs, such as video calling between differing languages, or functionality to apply virtual makeup [6] for use when telecommuting. For such purposes, our system supports plug-ins, providing an interface to pass the video and audio data transmitted during a video call to other programs. This enables additional functionality using the data to be added later. **Photo 5** shows the system being used with a prototype translation plug-in. The plug-in converts speech to text and uses an Application Programming Interface

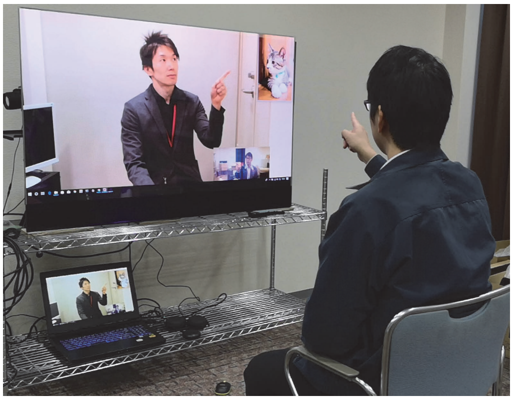


Photo 4　Object sharing function

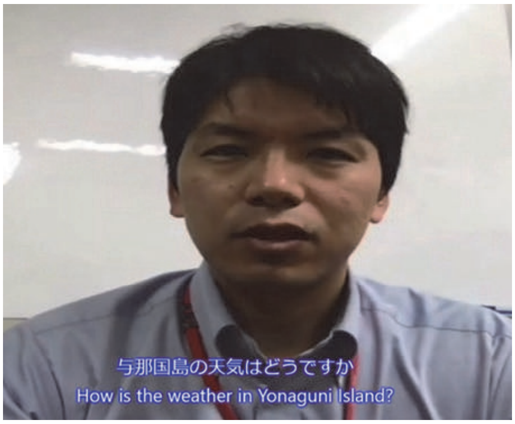

与那国島の天気はどうですか
How is the weather in Yonaguni Island?

Photo 5　Operation of the translation plug-in

(API)[*8] to translate it to a specified language, so that captions in the receiver's language are shown on the screen for video calls between users speaking different languages, just as though a simultaneous interpreter was being used.

5. Conclusion

This article has described a study in which users evaluated the contribution of various video calling parameters on a sense of being face-to-face, and the need for a "Face-to-face video calling system that facilitates natural conversation with eye contact," based on a result of the study, indicating the need for eye contact. As the speed of networks increases and devices such as displays and cameras continue to advance, we expect video calling to be used in an increasing range of scenarios, and our intension was to build a system that provides a strong sense of being face-to-face in practical terms.

A system adopting display method was demonstrated in the Smart Home Communication booth at "DOCOMO Open House 2018: Revolutionizing business and the world with 5G," held on December 6-7, 2018, and was very well received.

In the future, we will continue working with our

[*8] API: An interface that enables software functions to be used by another program.

partners toward commercialization of this technology, with testing, demonstrations and other activities.

REFERENCES

[1] A. Prussog, L. Mühlbach and M. Böcker: "Telepresence in Videocommunications," Proc. of the Human Factors and Ergonomics Society Annual Meeting, Vol.38, No.3, pp.180–184, 1994.

[2] L. S. Bohannon, A. M. Herbert, J. B. Pelz and E. M. Rantanen: "Eye contact and video-mediated communication: A review," Displays, Vol.34, No.2, pp.177–185, Apr. 2013.

[3] KDDI: "Sync Dinner," (In Japanese).
http://connect.kddi.com/sync/dinner/

[4] S. Uchida, E. Ashikaga, M. Imoto, M. Wagatsuma and K. Hidaka: "A Concept of Immersive Telepresence "Kirari!"," Proc of the 43rd IIEEJ Media Computing Call, T2-2, 2015.

[5] H. Ishii and M. Kobayashi: "ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact," Proc. of CHI'92, pp. 525–532, May 1992.

[6] Shiseido Corp.: "Shiseido develops 'TeleBeauty', an automatic makeup application for online meetings," Oct. 2016 (In Japanese).
https://www.shiseidogroup.jp/news/detail.html?n=00000000002041

Technology Reports

VR Live Video Streaming FPGA

8K VR Video Live Streaming and Viewing System for the 5G Era

Communication Device Development Department Naoto Matoba

NTT DOCOMO has developed the world's first VR 360° 8K video live streaming and viewing system. This system enables real-time operation through stitching equipment to synthesize video from a number of cameras into 360° 8K video without visible seams and encoding equipment that compresses and uploads the 360° 8K video to a streaming server. Both pieces of equipment use FPGA technology. Also, HMD using Panorama Cho Engine technology and high-resolution liquid crystal displays enable 360° 8K video viewing.

1. Introduction

In recent years, advances in Virtual Reality (VR)[*1] technology have led to the creation of viewing environments with 360° video[*2] that offer high level sensations of presence and immersion. The resolution of cameras capable of capturing 360° video has advanced from full High Definition (HD)[*3] to 4K[*4], while professional cameras capable of capturing 360° 8K[*5] video have also started appearing.

The resolution of Head Mounted Displays (HMDs)[*6] for 360° video viewing also continues to advance, and the fast communications speeds of the coming 5G technologies will enable streaming[*7] and live video streaming of the large amounts of data required for 4K and 8K. Hence, the combination of VR and 5G technology holds promise for services offering high-presence entertainment experiences such as sports and live performances to users in remote locations, as if the viewer is actually in the venue.

Live streaming of 360° video has almost never been attempted, because compared to normal video, the wide area of the 360° video display can suffer from unsatisfactory resolution, which made it impossible to deliver video with a quality good enough

[*1] VR: Technology that gives the user the illusion of being in a virtual world. In recent years, this illusion is mainly achieved using HMD (see*6) technologies that affect the user's visual perception.

[*2] 360° video: Video that covers the entire available field of vision in all directions - front and back, right and left, and up and down.

[*3] Full HD: A video screen format consisting of approximately 1,000×2,000 pixels in the vertical/horizontal directions.

to satisfy users.

Also, because it's not possible to capture 360° video with one camera, video signals from multiple cameras must be stitched together (stitching[*8]), encoded[*9] and uploaded in real time. However, real-time stitching and encoding entails very high processing loads, which means so far 360° video has been limited to 4K, even with professional equipment. In addition, delivering 360° 8K video from a server also would require very expensive specialized equipment to decode[*10] the 8K video so that it can be viewed, which makes it unachievable within the bounds of equipment configurations available to ordinary users. Additionally, user HMDs do not have sufficient resolution to reproduce 360° 8K video.

Hence, aiming for live streaming of 360° 8K video for the first time, we prototyped equipment to enable 360° 8K video capture, streaming and viewing in real time.

Specifically, we developed (1) stitching equipment to synthesize video signals from multiple 4K cameras and generate 360° 8K video in real time, (2) encoding equipment to compress 360° 8K video and upload it to a server in real time, (3) Panorama Cho Engine®[*11] (PCE) technology-based PCE encoding equipment for real-time streaming of 360° 8K video to users in a viewable format, (4) a PCE player to play back the PCE encoded data, and (5) an HMD using one 2K-resolution liquid crystal display per eye. Furthermore, to verify the usability of the equipment, we installed outdoor cameras connected to 5G equipment to test 360° 8K video live streaming.

This article describes the structure of the 360° 8K video live streaming and viewing system we developed, and describes a live video streaming demonstration experiment and its results.

2. Structure of the 360° 8K Video Live Streaming and Viewing System

2.1 System Structure and Objectives

Generally, equirectangular format[*12], an intermediate format, is used for 360° video streaming, because of the necessity to convert 360° video to regular video formats to compress and deliver 360° video with existing software and hardware, and because of the simpler player-side processing for viewing with an HMD. Generally, 360° video resolution is indicated by the resolution of the conversion to equirectangular format.

We are aiming to improve the resolution of 360° video using this streaming system to commercialize 360° live video streaming around 2020. Currently, HMD resolution for 360° video is mostly around 1K to 1.5 K per eye. Thus, we assume 2K display resolution per eye will be widely used around 2020.

Comparing the resolution of the HMD display with video resolution and assuming approximately 90° viewing angle in the horizontal direction, 8K resolution with the equirectangular format is necessary to display 360° video on displays with approximately 2K resolution per eye. Therefore, with this development, we set our technical objective as 360° 8K video live streaming.

With 360° video live streaming, video converted to the equirectangular format is transmitted in the same way as normal live streaming, and then played back at the viewer side after conversion for the

[*4] 4K: A video screen format consisting of approximately 2,000 × 4,000 pixels in the vertical/horizontal directions.

[*5] 8K: A video screen format consisting of approximately 4,000 × 8,000 pixels in the vertical/horizontal directions.

[*6] HMD: A device worn on the head that covers the field of vision, and displays a virtual world to the wearer. The wide viewing angle of these devices gives the wearer more intense sensations of emotion and presence.

[*7] Streaming: A communication method for sending and receiving audio and video data over the network, whereby data is received and played back simultaneously.

[*8] Stitching: Creating 360° video from videos captured with a number of cameras so that the seams between the videos are invisible. The 360° video output after stitching is often displayed in equirectangular format (see*12).

HMD. For this reason, more video processes are involved than normal live streaming and viewing, and these processes have to be done in real time, even though the greater the video resolution, the larger the processing load. Here, “real time” means the time equal to or shorter than the time calculated as the reciprocal of the frame rate (the number of frames displayed per second) for processing one frame[*13] of video with the viewing equipment. In other words, for a frame rate of 30 frames per second (fps), processing for each frame of video must finish in each piece of equipment within 1/30 of a second.

Figure 1 shows a schematic of the prototype system we developed. First, video captured with several cameras is combined and converted to equirectangular format, encoding is performed to compress the data into a size that can be transmitted, which is then uploaded to the server. Then, the encoded data is delivered from the server, decoded at the viewer side, and converted to panorama format to be played back on the HMD.

2.2 360° 8K Video Capture, Stitching and Encoding

First, we describe the method of capturing 360°

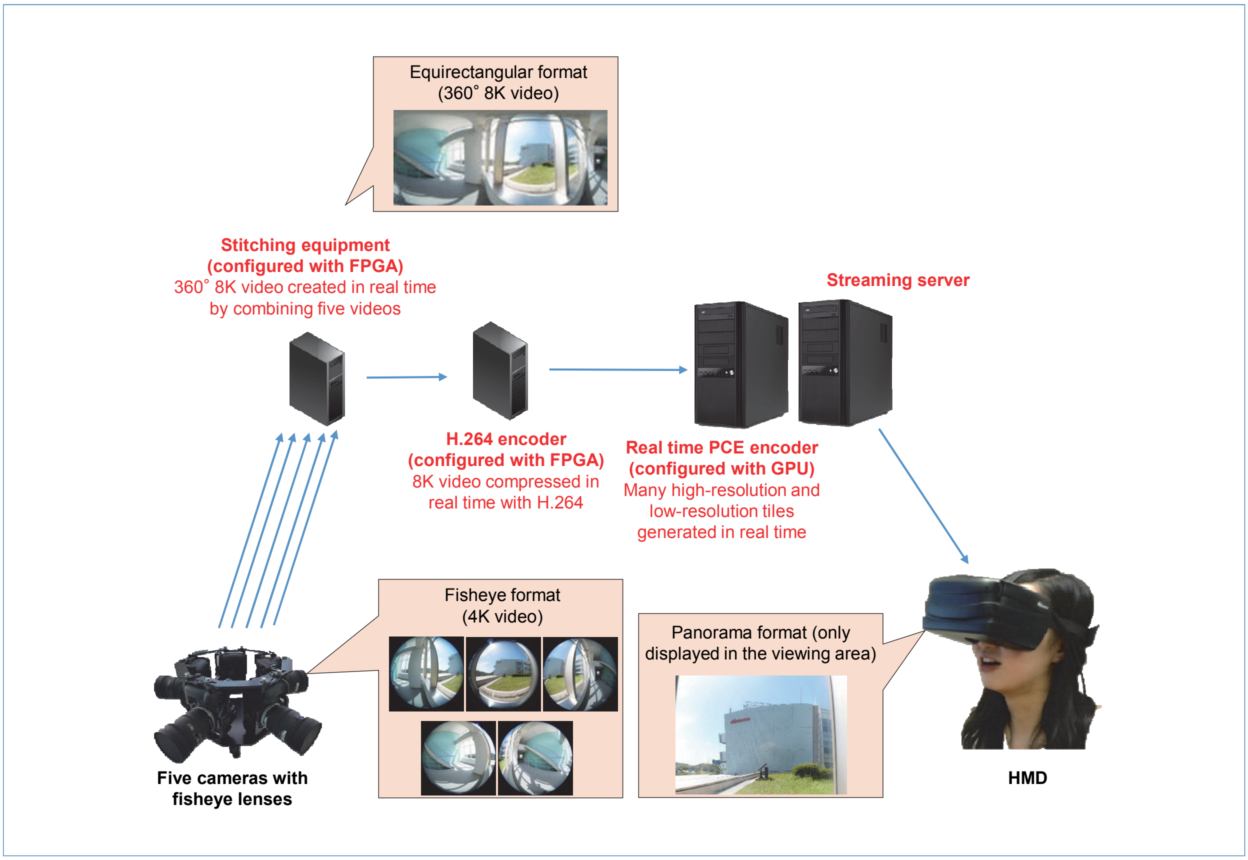


Figure 1　Overall equipment configuration diagram

[*9] Encoding: In this article, encoding means compressing large amounts of video data so that it can be sent over the network.

[*10] Decoding: Reconverting data compressed with encoding equipment back to video data.

[*11] Panorama Cho Engine®: Technology that uses streaming technology developed by NTT Media Intelligence Laboratories. A registered trademark of NTT TechnoCross Corporation.

[*12] Equirectangular format: A format for projecting spherical 360° video on a flat surface. Normally, this format has a vertical/horizontal ratio of 1:2, and is characterized by projection with latitude and longitude intersecting vertically.

[*13] Frame: One of the many single still images that make up video.

video. Because it isn't possible to cover the 360° using a single camera, even with a fisheye lens enabling capture 180 to 200° or greater, multiple cameras are used to capture video covering the 360° [1]. Also, it's not possible to use bulky professional cameras because the distance between cameras is large and stitching becomes impossible. For this reason, a combination of the now common compact 4K cameras is used.

Because the resolution of normal capture in the vertical direction is not sufficient to capture 360° 8K video, we arranged multiple 4K cameras tilted vertically 90° around the circumference and positioned horizontally outward. For the lens, we used a circular fisheye lens capable of capturing a 180° viewing angle. Many of the targets for capture exists close to the horizontal, and since it's not really necessary to install cameras to capture the ceiling or the floor, we only installed horizontally oriented cameras but were able to capture everything up to the ceiling with the fisheye lens. Also, video taken from the center of the fisheye lens with as many cameras as possible would be ideal because of the large distortion around the circumference of the fisheye lens. However, since more cameras means more processing load on stitching equipment due to the increased number of video signal inputs, we used only five cameras in this development.

Next, we described the stitching equipment. Using commercially available equipment for real-time stitching is limited to 4K video output. Currently, no equipment exists that can convert and output video signals with the resolution higher than 4K in real time. Although methods have been suggested that entail multiple stitching of 4K video [2], these methods don't take into account stitching of video from fisheye lenses and would be difficult to apply to this development. **Figure 2** describes the algorithm we used for stitching [3]. First, captured video is rotated 90° because cameras are oriented in the vertical direction, then signals are converted from fisheye format to equirectangular format. Next, by moving the video in vertical, horizontal and rotational directions of the optical axis of the lenses, the five video signals are synthesized and corrected so that video differences between cameras caused by camera fixing jigs are not noticeable. Also, so that the borders between cameras are not apparent, blending processing is performed on multiple parts of the video between the cameras.

Since the stitching processing must be done in real time, we settled on using Field Programmable Gate Array (FPGA)*14 technology because we learnt that it enables high-speed computation processing and can achieve 30 fps.

There are examples of achieving 8K encoding equipment such as [4]. However, considering costs including those of decoding equipment, we decided to use FPGA with H.264*15 IP core*16 to achieve encoding for this development.

A Serial Digital Interface (SDI)*17 is used for transferring video signals between the camera, stitching FPGA and encoding FPGA. With 360° video, it's not possible to position the stitching and encoding equipment close to the camera because the range of capture is in all directions. Also, flexibility is required for arranging equipment for outdoor experimentation to suit the conditions of the site. By using an SDI, it's possible to interconnect equipment using coaxial cable over distances

*14 FPGA: An integrated circuit that is configurable after manufacture.

*15 H.264: A video data compression encoding method standard recommended by ITU, and widely used in broadcast and Internet streaming, etc.

*16 IP core: Partial circuit information summarized as functional units for developing FPGAs, etc.

*17 SDI: A widely-used video signal transmission standard, mainly with professional video equipment, and that enables transmission of non-compressed video signals on coaxial cable.

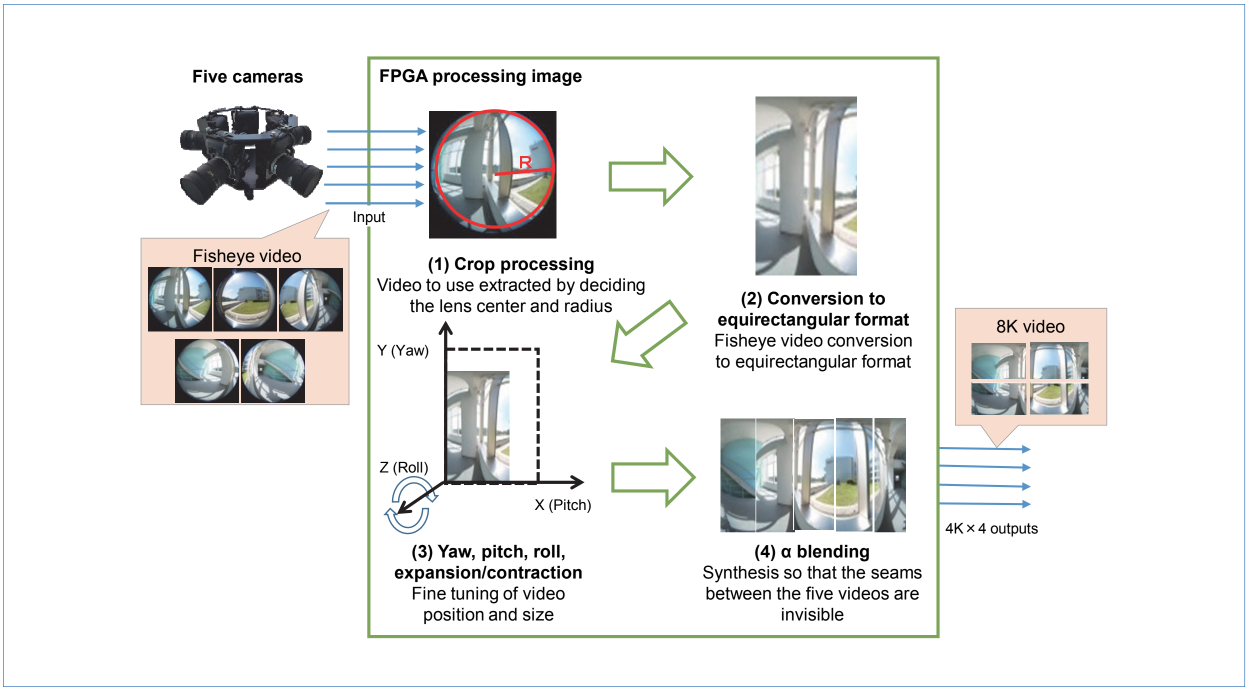

Five cameras
FPGA processing image
Input
Fisheye video
R
(1) Crop processing
Video to use extracted by deciding the lens center and radius
(2) Conversion to equirectangular format
Fisheye video conversion to equirectangular format
8K video
Y (Yaw)
Z (Roll)
X (Pitch)
(3) Yaw, pitch, roll, expansion/contraction
Fine tuning of video position and size
(4) α blending
Synthesis so that the seams between the five videos are invisible
4K × 4 outputs

Figure 2 Stitching algorithm

up to 100 m.

Figure 3 shows the external appearance of the stitching and encoding equipment we prototyped.

2.3 360° 8K Video Streaming and Decoding

Next, we describe the method of delivering and decoding the 360° 8K video.

Because of the large load involved with decoding 8K video delivered from the server for viewing at 30 fps, play back in real time is problematic even using a high-end PC with good decoding performance. Also, transmission capabilities of 80 to 100 Mbps are required because the bit rate for encoding must be sufficient to transmit signals without large drops in picture quality. Hence, aiming for commercialization around 2020, both the necessary transmission performance requirements and viewing equipment processing load must be reduced.

We used PCE technology in this development to solve these issues. **Figure 4** describes the principle of PCE. When viewing 360° video with an HMD with the horizontal viewing angle at 90°, the remaining 270° of video cannot be seen. Also, the direction that the user is viewing the 360° video is measured using sensors such as gyroscopes so that only the video in the viewing direction is displayed after decoding. Here, with PCE, the data for the detected viewing direction is sent to the streaming server, and the video in that direction is sent with the same resolution as 360° 8K video (a high-resolution partial tile). In addition, considering that the user might move his or her head suddenly, the 360° video is simultaneously sent with the resolution

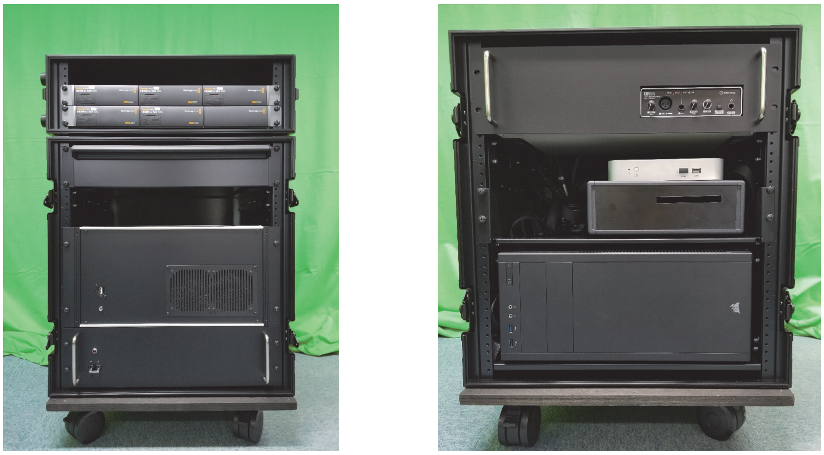


Figure 3　Stitching equipment (left), encoding equipment (right)

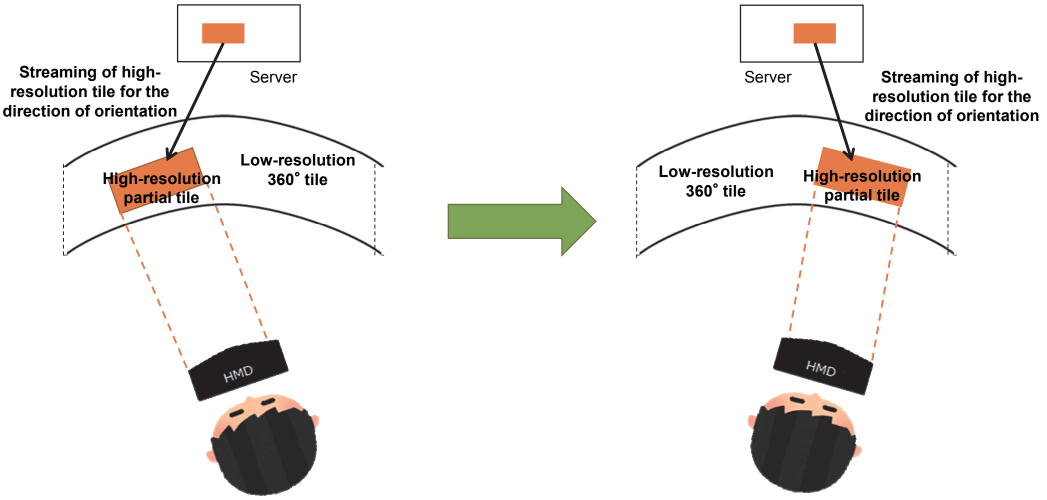

Server
Streaming of high-resolution tile for the direction of orientation
High-resolution partial tile
Low-resolution 360° tile
HMD
Server
Streaming of high-resolution tile for the direction of orientation
Low-resolution 360° tile
High-resolution partial tile
HMD

Figure 4　PCE principle

reduced (a low-resolution 360° tile). With this system, the video is always 8K resolution in the line of sight of a static user, but maintained at a lower resolution when the user suddenly moves his or her head.

In this development, considering the viewing angle, we used 2K for the resolution of both the high-resolution partial tiles and the low-resolution 360° tiles. This enables the decoding load to be two videos at 2K resolution, which is a significant reduction in processing load compared to 8K video, and enables play back through an HMD using processing capabilities equivalent to the latest smartphones or similar devices.

As it's necessary to generate these tiles in real time at the server, in this development, 16 high-resolution partial tiles in various directions and one low-resolution 360° tile are generated in real time in consideration of processing load at the server.

PCE technology enables viewing of video with the same quality as 8K.

2.4 Viewing with HMD

HMDs used for viewing 360° video consist of displays for the video, lenses adjusted for wide viewing angle focus and a housing designed to maintain suitable display and lens positioning when worn on the head. As discussed, as there are no products with displays suitable for viewing 360° 8K video available on the market, this is a newly prototyped system.

This HMD (**Figure 5**) uses a 2K-resolution liquid crystal display per eye matched to the resolution of 360° 8K video. The pixel density is 1,008 pixels per inch (ppi)[*18].

Both displays and lenses influence the quality of the video experienced when viewing with an HMD, but since there were no suitable lens combinations available, we designed and developed a new lens to suit the liquid crystal display.

To raise the sense of immersion and presence, requirements for the lens for this HMD include widening the viewing angle, making the lens small and light for mounting on the head, and minimizing peripheral aberration of the lens.

In this development, we used plastic to design a highly refractive aspheric lens to reduce weight and rectify aberration in the lens periphery. Although combining more of these lenses further improves peripheral aberration, we used three to strike a balance between the overall weight of the HMD and the extent of aberration improvement.

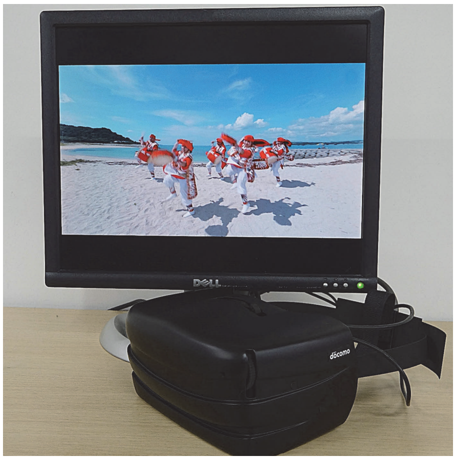


Figure 5　Prototype HMD

3. Verification Experiments of 360° 8K Video Live Streaming Using 5G Equipment

As stated, transmission speeds required to send 360° 8K video will be possible in the 5G era, meaning there is anticipation for high quality 360° video live streaming services.

To verify this technology, we performed a demonstration experiment of 360° 8K video live streaming using 5G at the Niigata Soh Odori event held on September 16, 2018 as part of the Niigata City demonstration experiment project. **Figure 6** describes

[*18] ppi: The number of pixels per inch.

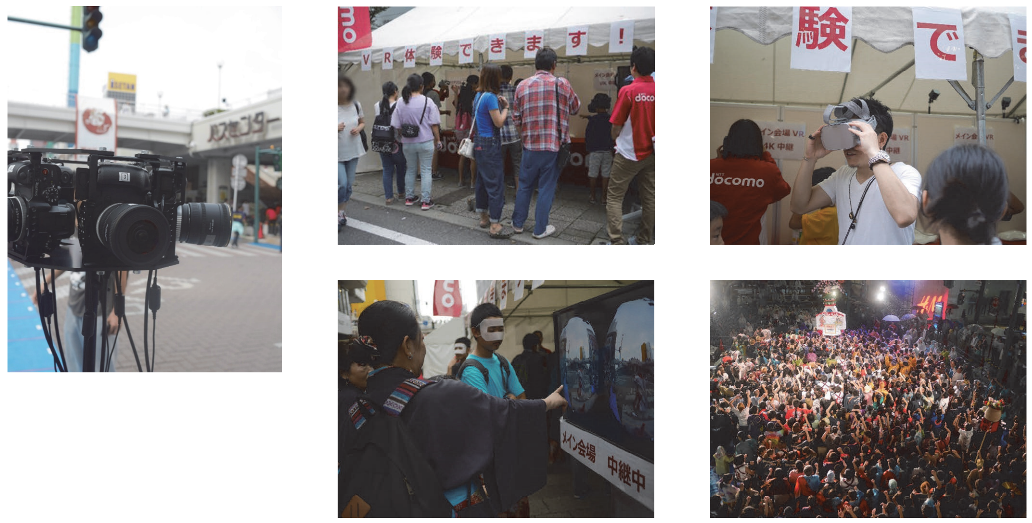


Figure 6 The scene at the Niigata Soh Odori demonstration experiment

this demonstration experiment.

During the experiment, we successfully provided stable live streaming of a 360° 8K video experience through HMDs to approximately 400 visitors over almost a full day.

4. Conclusion

We developed a prototype 360° 8K video live streaming and viewing system as an initiative to improve the quality of 360° video streaming for the 5G era. We also conducted outdoor technical demonstration experiments using this prototype equipment. This enabled us to garner a wide range of knowledge about the achievability and effectiveness of 360° 8K video live streaming and clarify a range of issues towards commercialization. Firstly, in solving these issues, we will continue to work on development for even higher quality to improve the user experience. At the same time, we also plan to study how to make the overall system cheaper.

REFERENCES

[1] M. Shohara, H. Sato and K. Yamamoto: "Special edition A: A New Audio Visual Experience, Chapter 2 360° Capture," Journal of the Institute of Image Information and Television, Vol.69, No.7, pp.652-657, Sep. 2015 (In Japanese).

[2] T. Sato, K. Namba, M. Ono, Y. Kikuchi, T. Yamaguchi and A. Ono: "Surround Video Stitching and Synchronous Transmission Technology for Immersive Live Broadcasting of Entire Sports Venues," NTT Technical Review, Vol.15, No.12, Dec. 2017.

[3] Kolor: "Autopano - Tutorials - Quick Start Guide." http://www.kolor.com/wiki-en/action/view/Autopano_-_Tutorials_-_Quick_Start_Guide

[4] Y. Nakajima, Y. Nishida, M. Ikeda, K. Nakamura, T. Ohnishi, T. Sano, H. Iwasaki and A. Shimizu: "Development of 8K video encoder with HEVC realtime encoder LSIs," Vol.40, No.35, BCT2016-76, pp.13-16, Oct. 2016 (In Japanese).

Technology Reports
Player QoE Quality Improvement

New Platform Technology to Further Improve the Quality of Multimedia Services —MediaSDK Software Library—

Communication Device Development Department Ginpei Okada Kazuki Asai Yunsang Oh

The multimedia service applications such as "dTV®*1, dTV channel®*2, dAnime store, and Hikari TV®*3 for docomo include a software library called MediaSDK which commonizes all functions required to provide services such as streaming processing. MediaSDK is a player required for video and audio data delivery, and because it has a significant impact on the quality of playback experienced by service users, its development significantly contributes to improving the quality of NTT DOCOMO's media services. This article describes an overview of MediaSDK and initiatives to improve the quality of related software.

1. Introduction

NTT DOCOMO has developed a number of services and applications respectively for the dramatically changing and highly competitive world of video delivery services, although there are issues with development efficiency. Hence, we developed MediaSDK (a software development kit) as common library with common functions for media applications (**Table 1**) to improve development efficiency, quality and maintainability.

NTT DOCOMO aims to improve the quality of MediaSDK from the perspectives of (1) the quality of the actual software, and (2) the Quality of Experience (QoE) of streaming*4. This article describes the developmental methods used to improve the quality of software, and analysis of quality data to improve QoE. The article also introduces issues and initiatives planned for the future.

2. Developmental Methods of MediaSDK

Currently, DOCOMO provides MediaSDK for the services below (**Table 2**). Since many services are provided through MediaSDK, the different

*1 dTV®: A trademark or registered trademark of NTT DOCOMO, INC.

*2 dTV channel®: A trademark or registered trademark of NTT DOCOMO, INC.

*3 Hikari TV®: A trademark or registered trademark of NTT Plala Inc.

Table 1 Overview of functions available with MediaSDK

Type	Details
Applicable media	Video, audio, subtitle
Delivery format	Streaming playback, local playback
Service model	Linear live playback, live catch-up playback, VoD
Delivery quality adjustment	Adaptive bitrate streaming, fixed bitrate streaming
Playback functions	Playback speed adjustment (0.5 to 32x), pause, seek, fast-forward/rewind
Supported devices	Android, Android TV, iOS/tvOS™, Web browsers
Digital rights management	Device-mounted DRM used (PlayReady, Widevine, FairPlay, etc.)
Codec	Hardware decoder used (H.264, H.265, etc.)
Data analysis	QoE data reporting function
API	APIs provided to application developers for each OS
Others	Device function linkage such as HDR, high-resolution audio

VoD: Video on Demand
tvOS™: A trademark of Apple Inc.

Table 2 Services (as of December 2018)

Service	Android	iOS	PC (HTML5)
dTV	○	–	–
dAnime store	○	–	–
dTV channel	○	○	○
DOCOMO TV Terminal® home app	○	–	–
Hikari TV for docomo	○	–	–
Other company's VoD service (service name not disclosed)	○	–	○

DOCOMO TV terminal®: A registered trademark of NTT DOCOMO INC.

hardware, operating systems, Digital Rights Management (DRM)[*5] methods, encoding settings, and contents formats, etc. involved must be handled individually. There are also demands for regular releases of MediaSDK with the flexibility to handle a wide range of service requirements.

2.1 Cross-platform Support

MediaSDK consists of functions required for video and audio playback, and plug-ins to customize those functions for various services. Normally, development of a new application requires complex implementations on an OS. In contrast, MediaSDK modularizes these complex implementations, and

[*4] Streaming: A communication method for sending and receiving audio and video data over the network, whereby data is received and played back simultaneously.

[*5] DRM: Functions for protecting copyrights of digital content by restricting redistribution and preventing unauthorized copies, etc.

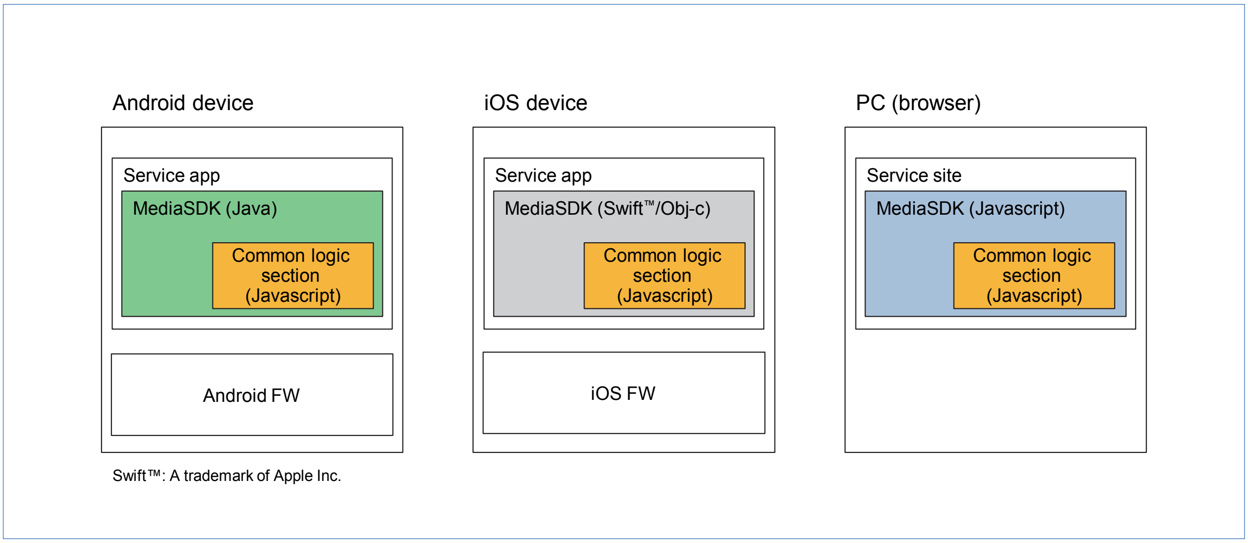
provides a commonized Application Programming Interface (API)*6, which makes it easier for developers to develop service applications.

Regarding plug-ins, our development must proceed efficiently on multiple platforms such as Android™*7, iOS*8 and PC, so the MediaSDK software includes plug-ins written in JavaScript*9 language for the cross-platform common logic section which enables commonization across operating systems (**Figure 1**).

Changing the software logic of each service can be done by updating the JavaScript section (**Table 3**). Furthermore, with modern Android and iOS smartphones, if conditions such as "primary purpose of the application is not changed," "a different storefront is not created," and "security is maintained" are satisfied, some sections of logic or parameters may be changed after delivery to the application stores (App Store®*10, Google Play™*11) by overwriting the script*12 (such as JavaScript) without altering the application itself. Therefore, a JavaScript plug-in hot patch can be delivered from the server to change the operation of an application using MediaSDK.

2.2 Agile Development Initiatives

MediaSDK entails development requests from various services received simultaneously and in



Swift™: A trademark of Apple Inc.

Figure 1 MediaSDK software structure

Table 3 JavaScript plug-ins

Plug-in type	Details
QoE plug-in	Sends QoE quality report and updates report data
ABR plug-in	Changes adaptive bitrate adjustment logic
Network plug-in	Changes destination content server connection
DRM plug-in	Changes DRM to use

*6 API: An interface that enables other software to use the functions available with an OS or middleware.

*7 Android™: A trademark or registered trademark of Google, LLC., in the United States.

*8 iOS: A trademark or registered trademark of Cisco Corp. in the U.S.A. and other countries, and used under license.

*9 JavaScript: A script (see *12) language appropriate for use in Web browsers. JavaScript is a registered trademark or trademark of Oracle Corporation, its subsidiaries and affiliates in the United States and other countries.

*10 App Store®: A trademark or registered trademark of Apple Inc. in the United States and other countries.

parallel as well as frequent requests for functions to be released, but because these cannot be properly handled with normal waterfall development*13 schemes, they are done with agile development*14 (Scrum development*15). In actual fact, the introduction of agile development schemes enabled 20 releases in FY 2017 including evaluation versions.

2.3 Testing and Release Automation

To achieve 20 releases per year, software must be evaluated rapidly. Time taken for manual testing will also detract from development. To solve these issues, we automated 93.2% of testing (2,750 items) to reduce the work load involved in evaluation.

We have also reduced the work load for creating release items (library, sample applications, porting guides, API specifications) by making automatic generation possible with a one-click linkage to Source Repository*16 systems such as Git*17. **Figure 2** shows an image of the automation of SDK release package creation.

2.4 Linkages to Functions in Devices (Decoder, DRM)

We have designed MediaSDK for playback using secure DRM methods and decoders supported by various operating systems and devices, and developed it with support for the DRM methods commonly used in recent years such as Widevine™*18/PlayReady®*19/FairPlay®*20, etc. so that services can be provided on a wide range of devices including PC browsers. For this reason, services such as those in **Table 4** are able to smoothly deliver high-quality audiovisual contents that require provision of high security level digital rights management technology.

However, particularly with Android terminals, depending on the handset model and chipsets*21 installed, there is a lot of variation in playback performance and quality which can affect the services users are experiencing.

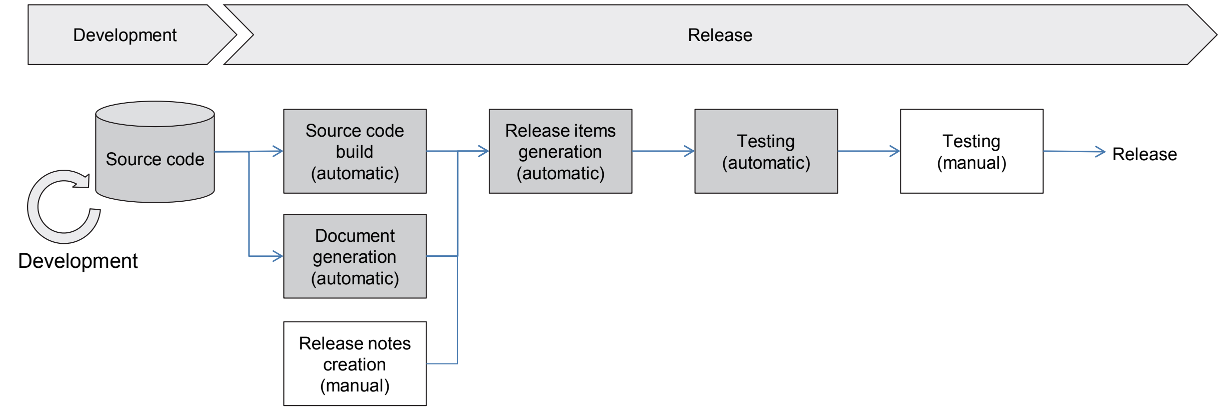

Development
Release
Source code
Development
Source code build (automatic)
Document generation (automatic)
Release notes creation (manual)
Release items generation (automatic)
Testing (automatic)
Testing (manual)
Release

Figure 2 Image of automation of SDK release package creation

*11 Google Play™: A service from Google for delivering applications, video, music and books to Android terminals. Google Play™ is a trademark or registered trademark of Google, LLC. U.S.A.

*12 Script: A simple programming language for describing programs for simple processes. A program described by a script may also be called a script.

*13 Waterfall development: A development method in which the processes of definition of requirements, design, implementation and evaluation are performed in order.

*14 Agile development: A development methodology based on the Agile development declaration, a generic name for light development methods for rapid and adaptive software development.

3. Quality Data Analysis Initiatives

Apart from managing quality when developing software, there is also a mechanism in MediaSDK to further improve the quality of services after commercial release. The plug-in function that achieves quality management from the perspective of maintenance operations is described below.

3.1 Playback Quality Data Collection

Here, we describe initiatives to improve the quality of user experience with the functions of the QoE and Adaptive Bit Rate (ABR)*22 plug-ins. The former enables QoE improvement by building base for collection and analysis of data about video playback quality, while the latter enables QoE improvement through changes to its ABR logic.

Figure 3 shows an image of the operations of the QoE and ABR plug-ins, while **Table 5** shows examples of media playback quality data reported from the QoE plug-in.

The data from the QoE plug-in is in the JSON*23 format. In December 2018, approximately 30 GB of data (all Android smartphone users) were collected every 24 hours for the dAnime store. This data is analyzed to extract issues to improve playback quality, and then the extracted issues are reflected in the playback logic using the ABR plug-in to improve the playback quality.

Table 4 Example of applied high security level digital rights management technology

Service	Contents	Remarks
dTV	4K HDR10	
dTV channel	Dolby Atmos®	a-nation event live delivery on August 25 and 26, 2018
Hikari TV for docomo	Dolby Vision® HDR	

HDR: High Dynamic Range
Dolby Atmos®: Dolby Atmos and Dolby Vision are registered trademarks of Dolby Laboratories.

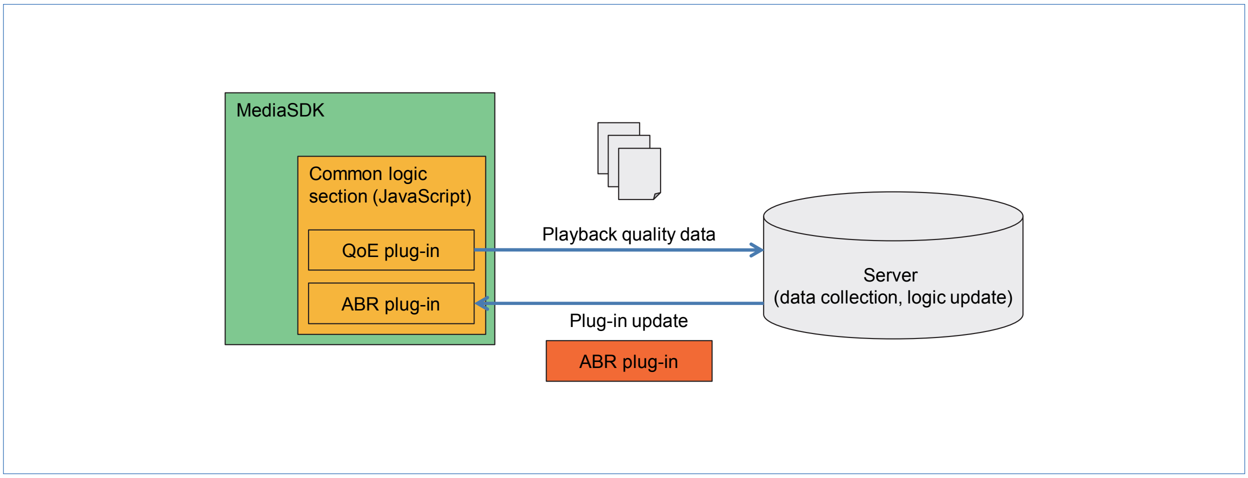


Figure 3 QoE, ABR plug-in operations system image

*15 Scrum development: An agile development method.
*16 Source Repository: An area where source code is stored.
*17 Git: A source code management tool.
*18 Widevine™: A trademark or registered trademark of Google, LLC., in the United States.
*19 PlayReady®: A trademark or registered trademark of Microsoft Corp. in the United States and other countries.
*20 FairPlay®: A trademark or registered trademark of Apple Inc. in the United States and other countries.
*21 Chipset: Devices that control mobile terminal software and various hardware processing. Devices such as the CPUs and control circuits are collectively referred to as "the chipset."

3.2 Playback Quality Data Analysis

We use the Plan, Do, Check, Act (PDCA) cycle*24 in MediaSDK maintenance operations. We describe the process of improving playback quality through quality data analysis based on PDCA (**Table 6**).

1) Planning

First, data reported from the QoE plug-in is modified in the server so that it can be visualized. **Table 7** shows an example of visualization of the state of playback quality.

By analyzing visualized data, user QoE can be understood. The next step is to set the target values required for improvements. Main improvement targets include "time taken for playback to start," "time for re-buffering during playback," and "adjustment for selected ABR picture quality."

2) Doing

ABR logic is used to optimize playback by changing the picture quality to shorten buffering time. We create proposals in consideration of each case because although the ABR plug-in can be applied to all applications, it's also possible to apply customized logic for particular groups (of Internet Service Providers (ISP), device models, operating systems, etc.).

3) Checking

The created ABR plug-in is delivered to a specific user group, and a comparison is made of its quality with the existing logic (by A/B test). Because it's possible to perform the evaluation with the aforementioned visualization, the quality targets in Table 7 are used as the main evaluation items.

Table 5 Example of playback quality report data

Data	Details
Startup time	Time until playback of first video frame starts
Buffered duration	Cached continuous buffer time
Download time	Time to download a segment of video data
Size	Size of downloaded segment
Consecutive drops	Number of serial frame drops (value for measuring terminal performance)
Action	User event
Network errors	Network errors during playback
Others	Others, reporting for QoE-related data

Table 6 Maintenance operations with the PDCA cycle

Planning	Specific targets for improvements are set through data analysis with data visualization. Improvement proposals are studied from data analysis based on hypotheses.
Doing	An improved version of software is created based on the results of studies in the planning stage
Checking	A/B testing of the existing software and the improved version, and evaluation of results
Action (improvement)	Software is updated based on the results of the A/B testing

*22 ABR: Technology to dynamically change picture quality for playback to match communication speeds.

*23 JSON: A data description language based on object notation in JavaScript®.

*24 PDCA cycle: A method of ensuring smooth running of business. The PDCA cycle entails repeatedly and continually running through the four steps of (1) Plan (planning), (2) Do (performing), (3) Check (measuring results) and (4) Act (making improvements).

Figure 4 shows an image of the ABR plug-in A/B test

4) Action (improvements)

The optimal playback logic ABR plug-in is selected from the results of playback quality evaluation, which can be reflected in the player as required to improve quality. The PDCA cycle does not have to finish with one cycle but can be repeated to further improve quality.

3.3 Overall Evaluation of Playback Quality

The results of quality targets must be properly judged for their merits to advance the PDCA cycle for quality improvement by collecting and analyzing quality data, as mentioned. However, there are many quality targets that are variables in evaluations, and many variables in video playback are in trade-off relationships. Hence, there may be problematic cases where improving one target

Table 7 Playback quality visualization example

Playback quality target	Details
Average playback time	Average playback session (total playback time/total number of playback sessions)
Average buffer size	Size of data accumulated in device before playback
Number of error events, frequency	Error event trends
Start time	The time taken from the user's start playback operation to the actual start
Speed	The network environment throughput at the user side
Bitrate ratio	Trends of video and sound quality selected with ABR logic
Frame drop	Ratio of occurrence of data that cannot be processed due to the terminal performance

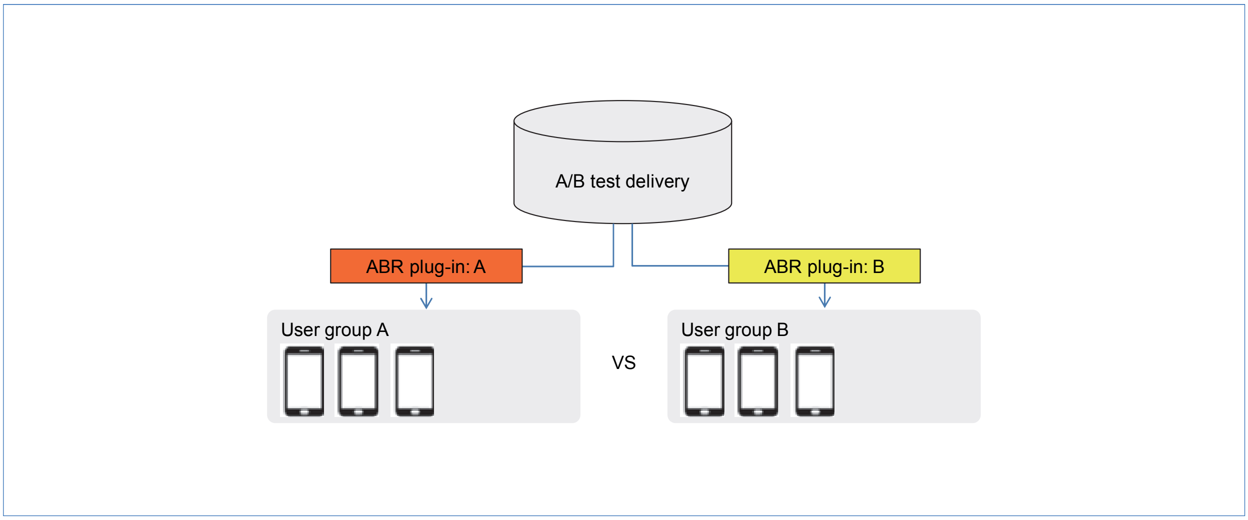


Figure 4 Image of the ABR plug-in A/B test

degrades another, making it difficult to judge merits fairly with A/B testing.

Hence, we use NTT Network Technology Laboratories' playback quality evaluation technology [1] to perform an overall evaluation. Using the technology described in [1], it's possible to calculate overall QoE values for points 1 to 5 using information about buffering with video playback or media picture quality. Comparing these QoE values makes it possible to maximize the levels of user satisfaction by maximizing the values through the PDCA cycle.

4. Issues

4.1 Differences in Device Performance

It's not always possible to make great improvements of streaming playback performance just by improving the ABR algorithm, because streaming playback performance is heavily dependent on device performance. Particularly with Android devices, where there is a lot of variation of the quality and performance of functions required for video services (decoding, DRM encoding processing, etc.) with each manufacturer, it can be difficult to apply uniform improvements. For this reason, we plan to study methods to measure device performance with playback and apply algorithms tailored for performance.

4.2 Overall System Optimization

This article has described improving playback quality, although QoE targets also depend on factors such as Content Delivery Network (CDN)*25, delivery servers and content encoding methods. Therefore, we intend to optimize overall service systems by repeating A/B testing with various combinations of servers, clients and contents.

4.3 PDCA Cycle Efficiency

The A/B testing and so forth we have discussed currently all require human operations. However, by promoting automated data analysis with the advances in quantitative scores such as QoE values and with AI technologies, we intend to make the PDCA cycle more efficient and put efforts into automating quality improvement of video delivery services.

5. Conclusion

This article has provided a general description of the development methods for the software library included in DOCOMO's multimedia service applications, and described how their quality is improved. We implemented agile development schemes for flexible cross-platform development. We also reduced working load and achieved greater efficiency by automating release. Going forward, we aim to further raise quality through repeated analysis, evaluation and improvements with the PDCA cycle using NTT Network Technology Laboratories' QoE visualization technology to analyze playback quality data.

REFERENCES

[1] K. Yamagashi and T. Hayashi: "Parametric Quality-Estimation Model for Adaptive-Bitrate-Streaming Services," IEEE Transactions on Multimedia, Vol.19, No.7, pp.1545-1557, Feb. 2017.

*25 CDN: A network solution optimized for fast and stable distribution of large files such as images and video.

NTT DOCOMO
Technical Journal Vol.20 No.4

Editorship and Publication

NTT DOCOMO Technical Journal is a quarterly journal edited by NTT DOCOMO, INC. and published by The Telecommunications Association.

Editorial Correspondence

NTT DOCOMO Technical Journal Editorial Office
R&D Strategy Department
NTT DOCOMO, INC.
Sanno Park Tower
2-11-1, Nagata-cho, Chiyoda-ku, Tokyo 100-6150, Japan
e-mail: dtj@nttdocomo.com