Development of USB Transcoder

Tsuyoshi YAMASHITA*, Toshihiro TAKASHIMA and Takahiro KUSUMOTO

Due to the spread of mobile devices and the evolution of mobile broadband networks, a large number of mobile phone users watch video content on their own devices. The authors have developed a USB connected transcoder to convert high-quality content into the suitable format and qualities for mobile devices. This development allows existing recorders and broadcast receivers to make MP4 files and to send streaming video data to mobile devices via IP networks.

Keywords: transcoder, mobile, video, USB

1. Introduction

With the popularization and functional sophistication of mobile phones, smartphones and other personal digital assistances, as well as advances in mobile broadband network technology, the number of mobile users who watch video content on their own devices is increasing today. On the other hand, the quality of the video content itself is being improved as large floor-standing video display units are in wide use. This level of content is of high quality and is difficult to watch directly on a mobile device with a smaller screen and limited signal-processing capability. Thereby the market is looking to transcoding technology that converts high-quality video content to a format that makes it possible to watch on conventional mobile devices. To respond to this market trend, we have developed a USB transcoder that transcodes the content when connected externally to conventional video recorders or receiver devices. This paper gives an outline of the transcoder and its evaluation results.

2. Distributing Video Content to Mobile Devices

2-1 Expansion of the mobile content market

Many years have passed since the distribution of multimedia content to fixed broadband networks was generalized. In the field of mobile devices such as mobile phones and smartphones, widespread use of high-performance terminals and increased wireless network capacity (data transmission speed) have also enabled the distribution of Chaku-uta (ringtones) and other music content. In particular, the recent increase in network transmission speeds has encouraged video content distribution services to mobile phones such as "EZ channel" and "BeeTV." In addition to the above distribution services for mobile phones, services for watching video are also becoming popular among smartphone, game machine, portable music player, and digital photo frame users. Because of this change in user attitudes, the size of the mobile content distribution market, which exceeded 500 billion yen in 2009, is expected to expand.

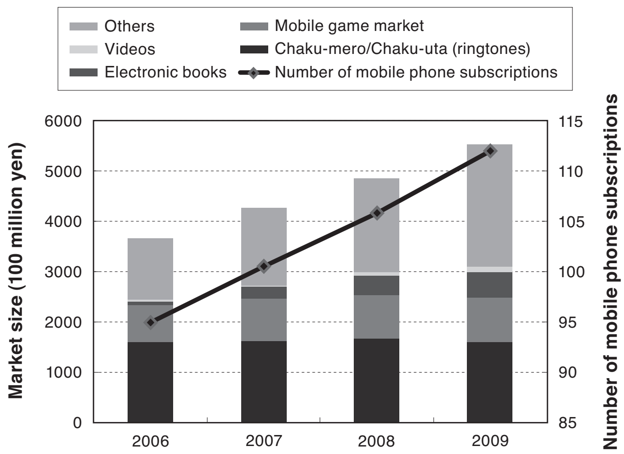

Fig. 1. Expansion of Mobile Content Distribution Market

2-2 Using content on mobile devices

Along with an increase in the number of users who watch video content on their mobile devices, there is increasing demand for watching video content which has been recorded in their video recorders or is distributed for floor-standing electric devices such as TV units. However, it is difficult to watch general broadcast content that contains high-quality video data on mobile devices due to the inferior processing performance of mobile devices and differences in video formats. In addition, there are many differences in display size, supported format or video parameter depending on the mobile device (**Table 1**). This means that to share the same video content on different mobile devices, the content must be converted to a format available on each terminal.

The process of converting digital video and/or audio from one format and/or codec to another and converting the resolution, bit rate, and other video parameters is generally called "transcoding." Transcoding is a key technology for the distribution of video and audio content to mobile devices.

Table 1. Video Terminal Specifications

	Television	Mobile phone (*)	Portable media player supplied by company A	Portable game machine supplied by company S
Display resolution	HD (1920×1080)	QVGA (320×240) VGA (640×480) WVGA (800×480) FWVGA (854×480)	480×320	720×480
Codec	MPEG2 (MP)	MPEG4 Video (SP) H.264/AVC (BP) VC1	MPEG4 Video (SP) H.264/AVC (BP)	MPEG4 Video (SP) H.264/AVC (BP/MP)
Format	MPEG2-TS	MP4 WMV One-Seg (TS)	MP4	MP4 avi

* Specifications of mobile phones depend on their terminal types.
SP = Simple Profile, BP = Baseline Profile, MP = Main Profile, HP = High Profile

3. Development of USB Transcoders

3-1 Outline of the transcoder

We have prototyped a dongle type USB transcoder that externally adds a transcoding function to conventional video devices without a transcoding function. This USB transcoder will enable users to convert video content recorded in a video recorder into a suitable format and save it to their mobile phone to watch later at their convenience. The USB transcoder can also convert, in real time, the broadcast program that the user is watching on a floor-standing TV set to a format suitable for a mobile device or electric devices in other rooms. The USB transcoder's input specifications and the output specifications for each function are shown in **Table 2**.

3-2 Hardware specifications

Table 3 gives the hardware specifications of the prototype USB transcoder and **Fig. 2** shows the hardware block diagram. The transcoder has been built as a USB dongle so that it can be plugged directly into the USB port of a conventional video device. Its chassis size is slightly larger than currently popular USB memory devices (**Fig. 3**). To improve the serviceability, the transcoder has been designed to be powered through the USB bus (except for the MicroSD card interface, which must be externally powered through an AC adapter.) This transcoder contains a memory system consisting of a 256 MB DDR3 SDRAM and 1 MB PROM. To reduce power consumption and support bus powered operation via USB, only a boot loader program has been embedded in the PROM. After the boot loader is started, firmware is downloaded from the USB host onto the RAM and then booted.

Table 2. Input/Output Specifications of the USB Transcoder

Input specifications

Item	Specification
Format	TTS
Video codec	H.264/AVC MPEG2
Audio codec	MPEG2 AAC MPEG1 Layer II
Video bit rate	20 Mbps max.
Audio bit rate	350 kbps max.
Resolution	1920×1080 1440×1080 1280×720 720×480

Output specifications (for downloading video)

Item	Specification
Format	MP4
Video codec	H.264/AVC
Video bit rate	64~384kbps
Audio codec	MPEG4 AAC
Audio bit rate	32~128kbps
Resolution	QVGA (320×240) QCIF (176×144)
Frame rate	30 fps, 15fps

Output specifications (for real-time streaming)

Item	Specification
Format	TTS
Video codec	H.264/AVC
Video bit rate	1~3Mbps
Audio codec	Pass through
Audio bit rate	Pass through
Resolution	VGA (640×480) QVGA (320×240)
Frame rate	30 fps

Table 3. Hardware Specifications of the USB Transcoder

Item	Specification
Memory	256 MB (DDR3 × 2)
PROM	1 MB
Power consumption	800 mA max.
Enclosure dimensions	117 mm × 33 mm × 13.6 mm (see **Fig. 3**) (without protrusions)
External I/F and others	LED × 1 USB 2.0 I/F (plug type A) × 1 MicroSD card slot × 1 5 V external power supply × 1 * External power supply is required only when a MicroSD card is used.

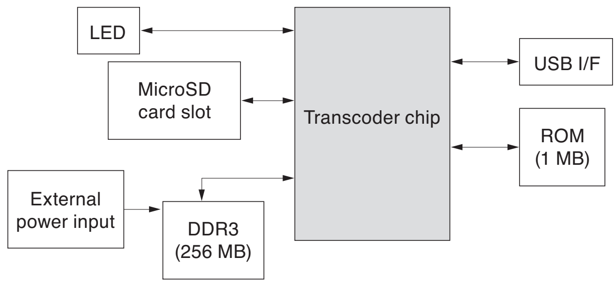

Fig. 2. Hardware Block Diagram

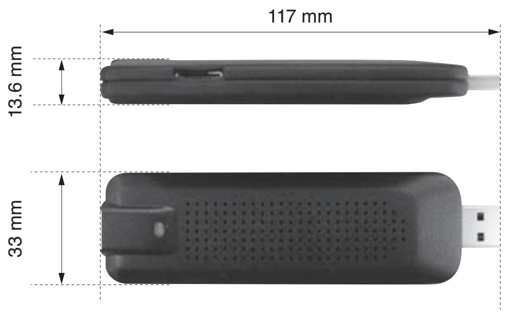

Fig. 3. External Appearance of USB Transcoder

3-3 Software and system structure

The new USB transcoder communicates with the host by IP protocol. Since the parameter setting and video stream transfer between the USB host and USB transcoder are performed over an IP communication channel, it is easy to build application software on the host side. Since the USB transcoder is still a prototype, the data transmitted between the host and the USB transcoder is not encrypted.

The system structure necessary for use of the transcoder is described below. **Figure 4** shows an example of the system structure to convert the format of the stored content for viewing on mobile terminals. The host is a video recorder, which reads out the content saved on its HDD or other storage medium and inputs it to the USB transcoder. The transcoder converts the input content to the MP4 file format and returns the converted content to the host. The host stores the content received from the transcoder in its storage medium. After conversion of the content is complete, the converted content file that has been stored in the storage medium is copied to a MicroSD card or other suitable recording medium to make it available for watching on a mobile phone or other mobile devices.

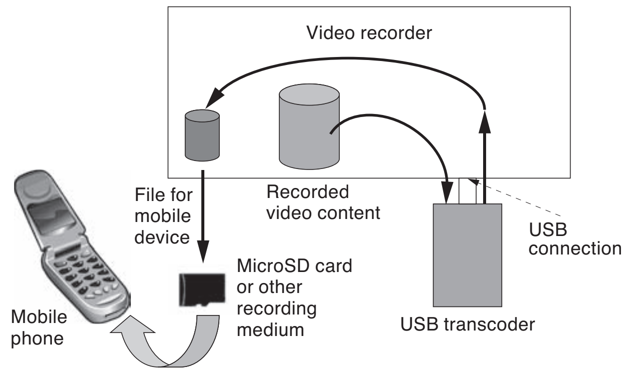

Fig. 4. Example of System Structure for Converting Stored Content

Figure 5 gives an example of the system structure for real-time transcoding.

In this case, a digital broadcasting receiver serves as the host, and it provides the copy of received stream to the USB transcoder. The transcoder converts the received stream to the TTS file format in real time and returns the converted stream to the host. The host transfers the transcoded stream to an external IP network to make the content available for watching in real time on mobile devices which cannot receive broadcast waves directly.

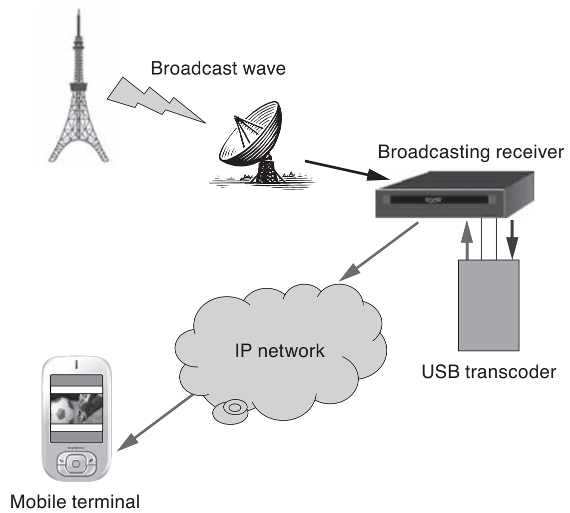

Fig. 5. Example of System Structure for Real-time Transcoding

4. Evaluating the USB Transcoder

This chapter deals with evaluation of the new USB transcoder and the evaluation results.

4-1 Evaluation of the combinational transcoding function

To check whether or not the new transcoder can transcode typical visual/audio parameter streams that are used widely for various broadcasting services, we carried out a test for different combinations of input streams and output parameters. The input stream parameters used for the test are given in **Table 4**, while the output parameters are given in **Table 5**.

In addition to the combinational test, we carried out an aging test (for approximately 3 hours), a selective transcoding test for each program of a multiple-program broadcasting, and a switching test between multiple-program broadcasting and single-program broadcasting. The objective of these tests was to evaluate the operational stability of the transcoder and its response to behaviors peculiar to the broadcasting stream.

The combinational transcoding test did not reveal any problems for all combinations of input streams and output parameters. However, in the switching test between multiple-program and single-program streams, the transcoder was found to be unable to follow up the change of video codec types at the switching point of the programs (**Table 6**). This shortcoming should be corrected in the near future.

Table 4. Test Stream Parameters

Video codec	Resolution	Audio codec	Type of audio
MPEG2	1920×1080i	AAC-LC	Stereo
MPEG2	1440×1080i	AAC-LC	Stereo 5.1 ch
MPEG2	720×480i	AAC-LC	Stereo Monaural
MPEG2	720×480i	MPEG1 Layer2	Stereo
H.264/AVC	1920×1080i	AAC-LC	Stereo
H.264/AVC	1440×1080i	AAC-LC	Stereo 5.1 ch
H.264/AVC	720×480i	AAC-LC	Stereo Monaural

Table 5. Output Parameters

Operating mode	Format	Video parameter	Audio parameter
Downloading to mobile phone at high quality	MP4	H.264/AVC, 320×240 30 fps, 384 kbps	AAC-LC/2ch, 96 kbps
Downloading to mobile phone at medium quality	MP4	H.264/AVC, 320×240 15 fps, 192 kbps	AAC-LC/2ch, 64 kbps
Downloading to mobile phone at low quality	MP4	H.264/AVC, 176×144 15 fps, 64 kbps	AAC-LC/1ch, 32 kbps
Real-time transcoding at high quality	TTS	H.264/AVC, 640×480 30 fps, 3 Mbps	Pass through
Real-time transcoding at low quality	TTS	H.264/AVC, 320×240 30 fps, 1 Mbps	Pass through

Table 6. Specific Stream Transcoding Results

Input stream	Result
Transcode of stream containing two or more audio ESs (When set to audio pass through)	OK
Transcode of stream containing two or more audio ESs (When set to audio transcode)	OK
Multiple program stream/channel selection transcode	OK
Multiple program broadcast stream/single program broadcast stream switching test No codec change during switching	OK
Multiple program broadcasting stream/single program broadcasting stream switching test Codec change (MPEG2 ↔ H.264/AVC) during switching	NG

4-2 Evaluating the transcoding speed

Time required for the new transcoder to transcode a file was evaluated. **Table 7** shows the results of measuring the transcoding speed. The transcoding speed for each combination of patterns was evaluated based on the ratio to the real-time transcoding speed for input content. The result showed that the transcoding speed remained virtually at 1 when audio data was transcoded, while the speed was 2 or more when audio data was passed through without transcoding. To expand its applications, the audio transcoding function is necessary. From the usability point of view, the transcoder should have a higher transcoding speed than real-time speed. A future challenge is how to increase the overall transcoding speed including the audio conversion.

Table 7. Transcoding Speed Evaluation Results

Input stream	Output parameter	Transcoding speed (ratio)
H.264/AVC, 1440×1080i AAC-LC stereo	Video: H.264/AVC, 320 × 240, 30 fps, 384 kbps Audio: AAC-LC/2 ch, 96 kbps	1.10
	Video: H.264/AVC, 320 × 240, 30 fps, 384 kbps Audio: Pass through	2.21
	Video: H.264/AVC, 640 × 480, 30 fps, 3 Mbps Audio: Pass through	2.25
MPEG2, 1440×1080i AAC-LC stereo	Video: MPEG2, 1440 × 1080i, 30 fps, 6 Mbps Audio: Pass through	3.39
	Video: H.264/AVC, 1440 × 1080i, 30 fps, 6 Mbps Audio: Pass through	2.46
	Video: H.264/AVC, 320 × 240, 30 fps, 384 kbps Audio: AAC-LC/2 ch, 96 kbps	1.03

* Transcoding speed (ratio) =
(Real time for content)/(time required for transcoding)

5. Solutions to Problems Characteristic of Mobile Devices

In the course of evaluating the newly developed transcoder, we discovered several problems characteristic of mobile devices. This chapter discusses these problems and possible solutions.

5-1 Solution to the lip-synching problem

When a broadcast stream is transmitted, the video and audio streams are multiplexed in a format called Transport Stream (TS) (**Fig. 6**). The TS format contains a time stamp (PTS) that indicates the display timing for both the video and the audio. A receiving device controls the video and audio display timings according to the time stamps to maintain synchronization between the video and audio (lip-synching process).

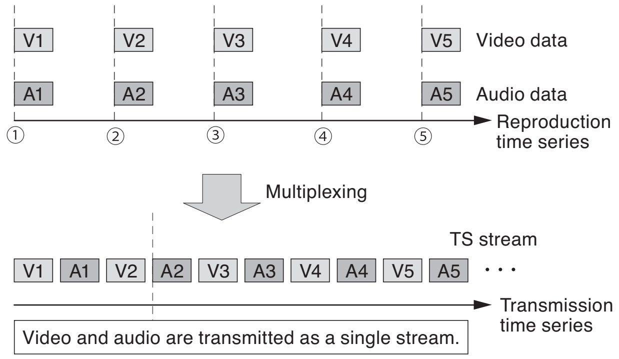

Fig. 6. Multiplexing by Transport Stream

On the other hand, for the MP4 file format used in mobile phones and other mobile devices, both the video and audio display timings are expressed as the relative time from the first data.

If a broadcast stream is taken out at a certain point in time and transcoded, the video and audio may start at different timings. If video and audio with a lag in the start time are directly stored in the MP4 file format, a display timing control scheme that relies on the relative time from the first data cannot maintain the lip-sync (**Fig. 7**).

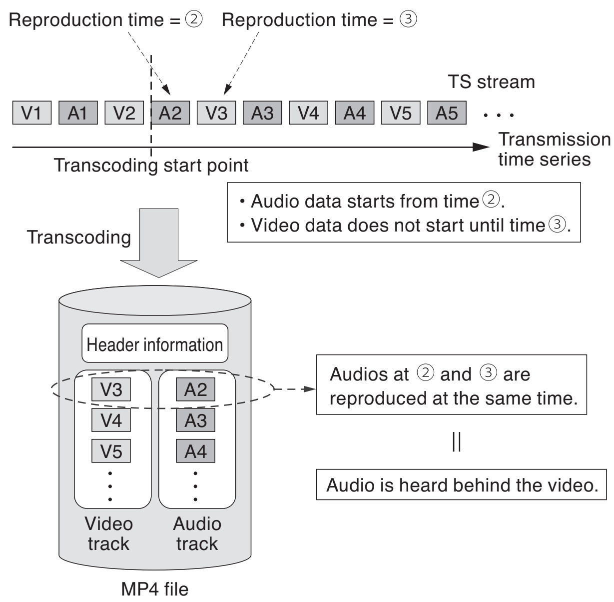

Fig. 7. Lip Sync Loss Resulting from Packaging to an MP4 File

To solve the lip-synching problem, the MP4 file format standard specifies a feature called "Edit List." This feature sets the offset to the start timings of the video and audio to adjust the video/audio synchronization. At the initial transcoder development stage, we planned to introduce the feature called Edit List. However, our survey revealed that some mobile devices did not support this feature. In fact, an initial prototype transcoder produced a maximum synchronization time lag of around a few hundred milliseconds between the video and audio. Although this problem was basically attributed to mobile devices that did not conform to the MP4 file format standard, we decided to modify our transcoder so that it would not be affected by the specifications of the mobile device. Based on this design concept, we replaced the Edit List feature with a new feature that uses the video and audio PTS in the input stream and adjusts the storing way of video and audio data in the MP4 file format. The newly introduced feature maintains the lip sync at a level where mobile terminal users do not notice anything disconcerting even if their terminals do not respond to Edit List.

5-2 Solution to the sample aspect ratio problem

The ratio between the width and height of each pixel is called SAR, and is usually contained in the video codec information. The SAR is processed in the video terminal to display video images on the screen. For example, an original video with an SAR of 4:3 is horizontally enlarged 4/3 times to display it on a screen with the correct aspect ratio. When a video is transcoded to a different resolution, the SAR should be replaced with the proper value. For example, when a video source with an SAR of 1:1 and an aspect ratio of 16:9 is transcoded to 320 x 240 pixels, the SAR should be changed to 4:3 as shown in **Fig. 8**. This system allows each display terminal to enlarge each pixel 4/3 times in the horizontal direction with respect to the vertical direction by referring to the SAR of the video source, thereby displaying the video with the correct aspect ratio.

When this system was tested on a mobile phone that did not respond to SAR, the mobile phone displayed the video with the same aspect ratio as the video source and distorted the images on the screen when the SAR of the input video was not equal to the aspect ratio of the output resolution.

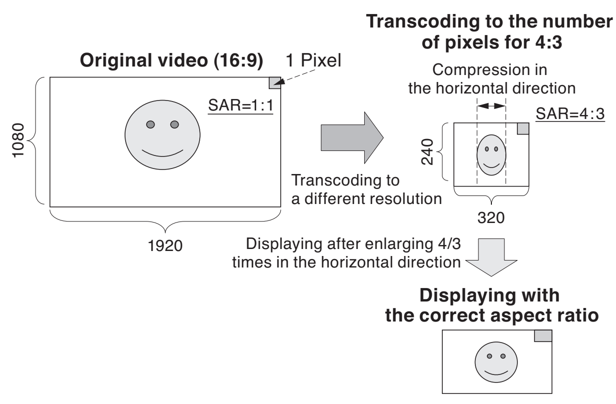

Fig. 8. Displaying Video after Transcoding to a Different Resolution

To rectify the above shortcoming, the final prototype transcoder was modified so that it compares the SAR of the input video with the aspect ratio of the output resolution before transcoding the video. If they are unequal, the transcoder letterboxes the video by adding a black band at the top and bottom of the image as shown in **Fig. 9**. Finally, the prototype transcoder enables a mobile phone to display videos with the correct aspect ratio.

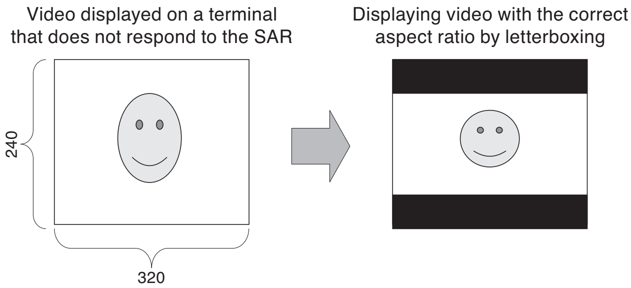

Fig. 9. Adjusting the Sample Aspect Ratio by Letterboxing

6. Conclusion

We have developed a dongle type USB transcoder. This device adds a transcoding function to the conventional video recorder or receiver when connected externally. By using the function of this transcoder, it is possible to convert stored video content to the MP4 file format and download it to a mobile terminal or to transcode video content on a real-time basis to make it available to watch on a mobile terminal.

Evaluation of the prototyped transcoder posed some problems to be solved such as the inability to transcode a specific stream and the transcoding speed with audio conversion. These problems need to be solved for commercialization of the transcoder along with data encryption and copyright protection for content downloaded to mobile devices.

7. Acknowledgements

This work is supported in part by Sumitomo Electric Networks, Inc.

Technical Term

*1 Content: News and other pieces of information, as well as music, movies, comics, animation, games, and various other creations, including books and web pages, offered by media sources.

*2 Transcode/Transcoding: A technique/device for converting video/audio data from a particular format or coding parameter to another.

*3 Chaku-uta: A service for converting music content to the audio file coded in MP3, AAC, or other suitable format as a ringtone for mobile phones.

*4 Chaku-mero: Another service to provide a ringtone for mobile phones.

*5 USB (universal serial bus): A universal bus standard that was established for connecting peripheral devices to personal computers. Recently, application of this standard has been expanded to cover connection between various types of electronic devices.

*6 USB bus power: A method of supplying power to USB devices, which is covered by the USB standard. Devices conforming to this standard can operate without requiring an external power supply by AC adapter.

*7 MicroSD: A nonvolatile memory device specified by the SD Association. MicroSD is a trademark or registered trademark of SD-3C LLC.

*8 DDR3 SDRAM (Double-data-rate3 synchronous random access memory): Volatile semiconductor memory specified by the DRAM standard.

*9 PROM (programmable ROM): A device that functions normally as ROM but can write data when accessed by a special procedure.

*10 ROM (read-only memory): Nonvolatile memory medium used for read-only purposes.

*11 Boot loader: A small program that is loaded when a system is started. The boot loader loads an operating system or other large program into memory and starts the program.

*12 Firmware: A software program that is embedded in a hardware device to control it. Firmware is software functionally closer to hardware than the average software.

*13 RAM (random access memory): Volatile semiconductor memory that can read and write data as needed.

*14 IP (Internet protocol): A standard for the packet-based communication on the Internet service.

*15 Storage: A device such as a hard disk for storing data in a computer. The storage refers particularly to a medium having larger storage capacity than RAM. Stored data is not erased (nonvolatile) even when the power supply is cut.

*16 MP4: A standard specified by ISO/IEC 14496-14 as a file format for video and audio data storage.

*17 TTS (time-stamped transport stream): A format in which 4-byte timestamp information is prefixed to an MPEG2 system TS packet.

*18 Multi program: A system for transmitting two or more channels (programs) in a single frequency band. This system is used for terrestrial and BS digital broadcasting services. A maximum of three standard definition (SD) channels can be multiplexed with a single high-definition (HD) channel band.

*19 SAR (sample aspect ratio): The ratio between the width and height of a digital image pixel.

*20 Letterbox: A method of adjusting the image size by placing a black band at the top and bottom of the image when the image and display screen have different aspect ratios.

References

(1) "Information technology – Coding of audio-visual objects – Part 12," ISO base media file format, ISO/IEC 14496-12, 2004-02-01

(2) "Information technology – Coding of audio-visual objects – Part 14," MP4 file format, ISO/IEC 14496-14, 2003-11-15

(3) "Information technology – Coding of audio-visual objects – Part 10," Advanced Video Coding, ISO/IEC 14496-10, 2004-10-01

Contributors (The lead author is indicated by an asterisk (*).)

T. YAMASHITA*

- Information & Communications Laboratories
  He is engaged in the research and development of video products.

T. TAKASHIMA

- Manager, Information & Communications Laboratories

T. KUSUMOTO

- Assistant Manager, Sumitomo Electric Networks, Inc.