

# Real-Sense Media Representation Technology using Multiple Devices Synchronization

Jae-Kwan Yun, Jong-Hyun Jang, Kwang-Ro Park, Dong-Won Han

Meta-Verse Technology Research Team, IT Convergence Technology Research Laboratory  
Electronics and Telecommunications Research Institute,  
138 Gajeong-no, Yuseong-gu, Daejeon, 305-700, South Korea  
{jkyun, jangjh, krpark, dwhan}@etri.re.kr

**Abstract.** Recently, the requirements for the real-sense media representation are increasing rapidly. Until now, most people mainly used the method SMSD on one device, but we need more than one device to play multiple audio/videos and multiple sensory effects of the real-sense media. Therefore, multi-track media synchronization and effect device synchronization algorithms are very important part of the real-sense media representation. Therefore, in this paper, we suggested concept of the real-sense media playback, sensory effect metadata scheme, real-sense media playback system architecture, and synchronization algorithm of multiple audio/video devices and multiple sensory effect devices.

**Keywords:** SMMD, SEM, Media/Device Synchronization, Real-Sense Media.

## 1 Introduction

Until now, most people have used the method SMSD (Single Media Single Device) which plays one media, including audio and video, on one playback device such as TV or PC [1]. But, as the development of the media playback technology, media running method is changing to the SMMD (Single Media Multiple Devices) which plays new type of media that includes not only previous type of audio, video, and text but also sensory effects(metadata information) to give users real-sense effects [3].

This SMMD service is very effective when we are making a movie for an experience room or an exhibition center, and a real-sense broadcasting service like stocks, cooking, quiz program which plays multiple devices to give users a lot of information at the same time [2].

Therefore, in this paper, we introduced the method for making a real-sense media that combines multiple audio/videos and sensory effects, and the method for playing this media with user peripheral audio/video playback devices and effect devices. Also, we explained the algorithms for the multiple audio/video synchronization and the multiple effect device synchronization. This paper is organized as follows. Chapter 2 discusses the concepts of the real-sense media playback, the real-sense media system architecture and SEM(Sensory Effect Metadata) scheme, chapter 3 explains

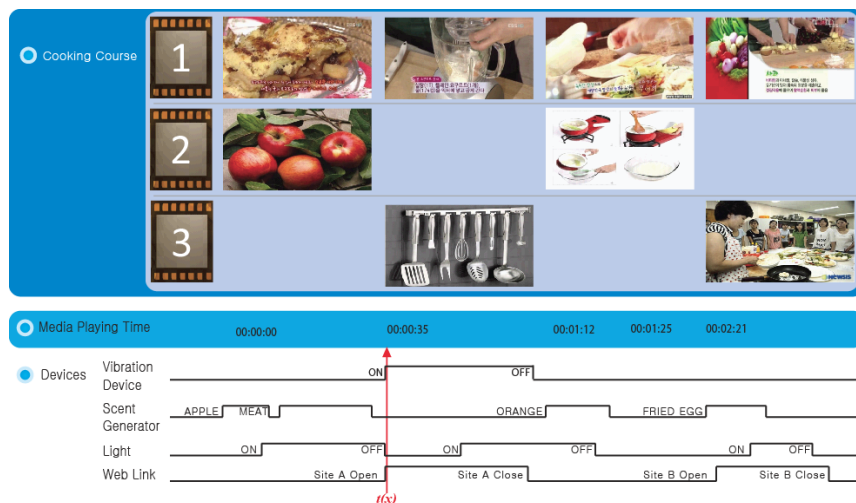
media/device synchronization and synchronization algorithms and chapter 4 shows the evaluation test results. The conclusion of this paper is given in chapter 5.

## 2 Real-Sense Media Playback

In this chapter, we explained concepts of the real-sense media playback, the real-sense media playback system architecture, and the sensory effect metadata scheme.

### 2.1 Concepts of the Real-Sense Media Playback

Basically, in the real-sense media service system, we aimed that one media which contains multiple audio/video tracks and sensory effects can be played with multiple audio/video devices and effect devices. Fig. 1 shows contents of the real-sense media of lecturing how to cook with 3 audio/video tracks and 4 sensory effects.



**Fig. 1.** Concept of the Real-Sense Media

Track number 1 becomes main track media and it contains audio/video of front screen of cooking lecture, track number 2 contains the scene of foodstuffs, and track number 3 contains the scene of cooking tools, and each track is stored in one real-sense media file. Sensory effects (vibration, scent, light, web link) are the real-sense effect related with each track, and also it is edited with the timeline of the main track media.

The real-sense media put together is transmitted to the home server from the contents server. And then, home server analyzes real-sense media, plays main track media on the playback devices like TV connected to the home server. The home server sends track number 2 and 3 to the user's around audio/video devices like laptop or other TV.

Each sensory effect for real-sense presentation is analyzed within the home server, translated to a control information for device control, and each control information

activates effect devices like light, scent generator, PC as the timeline of the main track media. As shown in fig. 1, to activate 3 devices (vibration ON, light OFF, web link A OPEN) exactly at the time of  $t(x)$ , the device execution time and network delay must be considered.

## 2.2 Real-Sense Media Playback System Architecture

To play real-sense media within the home server, we inserted each audio/video into one MPEG-4 media file. First of all, sensory effects that activate playback devices in each scene must be generated in the contents server with metadata authoring tool. And, based on the main track media, multi-track media must have synchronization information within a media. we normally used SL(Sync Layer) information of MPEG-4 file. Transport manager transmits real-sense media to the home server, and main controller sends this real-sense media to media parser. In the media parser, real-sense media is separated into multi-track audio/videos and SEM. The main track media in the separated audio/video tracks is played in the hardware decoder. The other tracks are sent individually to the audio/video playback devices, and sync manager is keep trying to synchronize each track with the main track media.

The SEM is parsed in the media parser and sent to the metadata analyzer. The metadata analyzer converts SEM into proper device information and device mapper makes device control message for each sensory effect. At this time, sync manager also synchronizes effect devices with the timeline of the main track media. Fig. 2 describes the system architecture for the real-sense media playback.

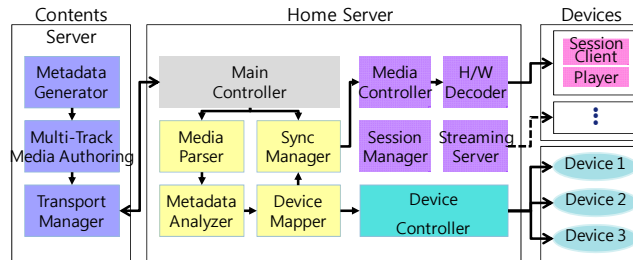
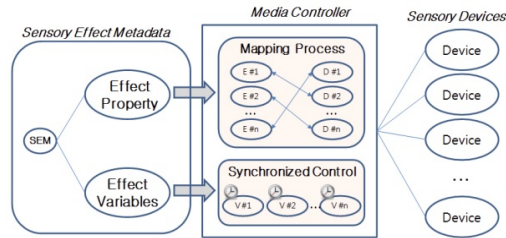


Fig. 2. System Architecture for the Real-Sense Media Playback

## 2.3 Sensory Effect Metadata Scheme

The SEM is the metadata for representing real-sense effects, and it contains effect types and effect variables by using XML [4]. The SEM is consisted of two main parts those are effect property and effect variables. Effect property contains definition of each sensory effect applied to the real-sense media. By analyzing effect property, the home server can map each sensory effect to a proper sensory device in the user's environment before processing real-sense media playback. Effect variables contains the control variables for sensory effect synchronized with the media stream. Fig. 3 shows the process of the SEM.



**Fig. 3.** Process of the Sensory Effect Metadata

Sensory effect metadata should be represented by the SEDL(Sensory Effect Metadata Description Language). Currently, The ISO/IEC MPEG is currently standardizing the SEDL [5]. Fig. 4 shows EBNF(Extended Backus-Naur Form) of SEDL.

```

SEM ::= [DescriptionMetadata] (Declarations|GroupOfEffects|Effect|ReferenceEffect)+
Declarations ::= (GroupOfEffects|Effect|Parameter)+
GroupOfEffects ::= timestamp EffectDefinition EffectDefinition(EffectDefinition)*
Effect ::= timestamp EffectDefinition
EffectDefinition ::= [activate][duration][fade-in][fade-out][alt][priority][intensity][position][adaptability]

```

**Fig. 4.** Sensory Effect Metadata Description Language

*SEM* is the root element and the *DescriptionMetadata* provides information about the SEM itself. The *Declarations* element is used to define a set of elements. The *Parameter* may be used to define common settings used by several sensory effects similar to variables in programming languages. A *GroupOfEffects* defines at least two effects play several devices at once. The *EffectDefinition* comprises all the information pertaining to a single sensory effect.

An *EffectDefinition* may have several optional attributes which are defined as follows: *activate* describes whether the effect shall be activated; *duration* describes how long the effect shall be activated; *fade-in* and *fade-out* provide means for fading in/out effects; *alt* describes an alternative effect identified by a URI; *priority* describes the priority of effects with respect to other effects in the same group of effects; *intensity* indicates the strength of the effect; *position* describes the position from where the effect is expected to be received from the user's perspective; *adaptability* enable the description of the preferred type of adaptation of the corresponding effect with a given upper and lower bound.

### 3 Media/Device Synchronization

In this paper, we synchronized several types of devices with real-sense media to give users real-sense effects. This effect device can be an active device which has its

own CPU, memory, operating system and activated by itself. On the contrary, passive device only receives control messages from the home server, controls and returns the feedback messages. There are two method to give real-sense effects to users, one method is representing effects by directly controlling such devices like electronic fan, scent generator, digital light related with the scene. Another method is 3 dimensional display using left and right movies, or 360 degree dome shape media display by playing multiple audio/videos with synchronized way. In this chapter, we explain audio/video synchronization and effect device synchronization algorithm.

### 3.1 Workflow of the Multi-track Media Synchronization

The multi-track media can be played several audio/video devices that user possesses and sub media is capable of being played in a user's laptop or PC. Multi-track is transmitted to the home server or other audio/video devices by using RTP(Real-Time Transfer Protocol), sends media control information with RTCP protocol. Home server becomes master, and it sends or receives synchronization messages continuously from slaves. Fig. 5 shows the workflow for the multi-track media synchronization.

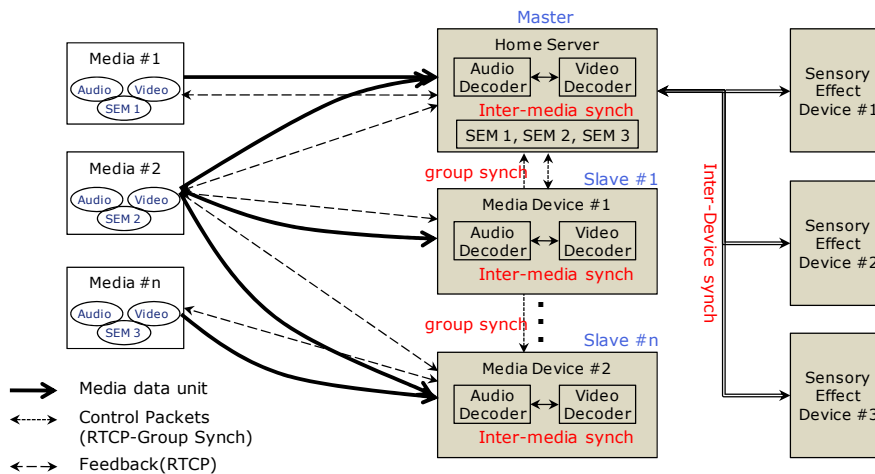


Fig. 5. Workflow of the Multi-track Media Synchronization

To play multi-tracks in the real-sense media, the service flow like fig. 6 is required. SEI(Server External Interface) is the interface for browsing the real-sense media list in the contents server and controlling start, pause, stop of media playback. DM(Device Management) is the interface for controlling the connection between the home server and players installed in each audio/video device, it sends VCR control message generated by the SEI and checks session maintenance of each audio/video device.

PII(Player Internal Interface) do a roll of media playback controlling, it runs and quits player process. DS(Device Synchronization) makes synchronization between each player installed in audio/video device, it do the OCR(Object Clock Reference)

mapping for synchronization. RTSP, RTP/RTCP is protocol for the real-time media transferring. MS(Media Synchronization) do inter-media synchronization by the OCR of the audio track in the player.

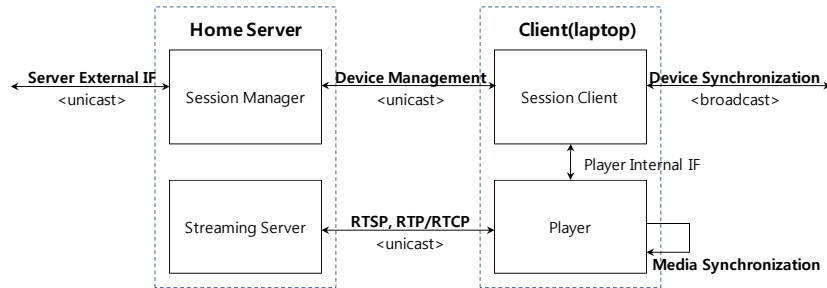


Fig. 6. Audio/Video Synchronization Architecture

### 3.2 Multi-track media Synchronization Algorithm

After choosing real-sense media and initializing players, the home server sends 'Play' message to the session manager. The home server becomes master, opens UDP socket, and broadcasts the master's play time for audio/video device synchronization.

Player records master's play time in a memory mapping file and session client reads this time. After make it media start time  $M(t)$ , register current system time  $C(t)$  by timer event, and broadcasts periodically current time every one second over a UDP socket. Current media time  $MC(t)$  is calculated by the formula (1), fig. 7 shoes audio/video synchronization algorithm.

$$MC(t) = M(t) + C(t) - P(t) \quad (1)$$

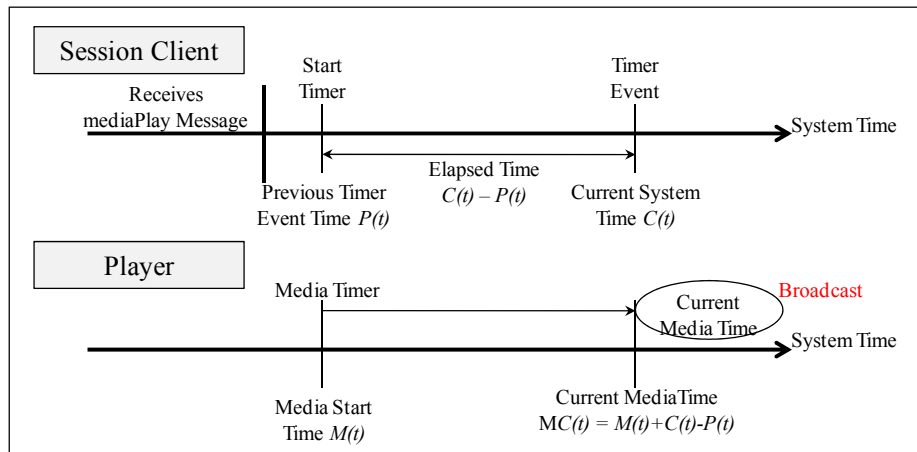
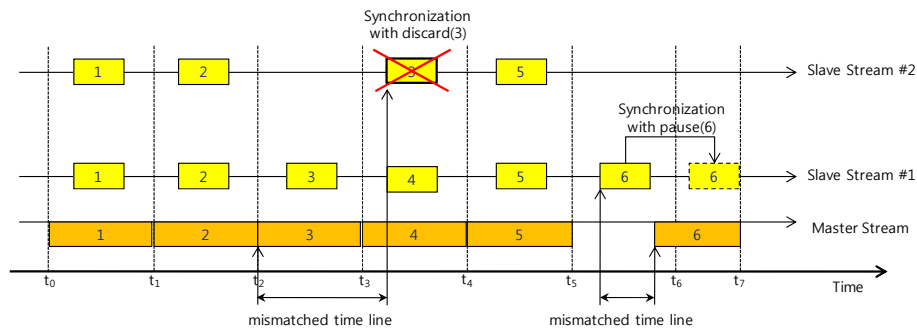


Fig. 7. Audio/Video Synchronization Algorithm

In this case, like fig. 8, there are occasions that stream of the master device does not synchronized with the stream of slaves. Number 3 of the slave stream #2 was discarded because it did not reached the time range  $[t_2, t_3]$  that master stream must be played. Number 6 of the slave stream #1, it paused and replayed to make a synchronization with the master stream, because master stream did not arrived just in



its playing time. By using this way, the home server can synchronize with other devices by performing pause and discard MDU(media device unit)s.

**Fig. 8.** Synchronization with Pausing MDUs and Discarding MDUs

### 3.3 Effect Device Synchronization

To synchronize effect devices efficiently, there need two time table in real-sense media service system. First of all, the sensory effect time table is used to maintain the SEM when real-sense media is transmitted to the home server. In the sensory effect time table, each effects are classified into effect property, start time, duration, and effect variables. Fig. 9 (a) shows the SEM example defined 9 effects during the time  $[t_0, t_6]$  about the effect type Light, Wind, Scent, Heat.

With an authoring tool which makes an expression of the SEM, the author may edit effects about the scene regardless of the effect devices activated. In this case, the home server maintains types and status of effect device connected with the home server. After analyzing the SEM in the parser, mapping process is done to choose best device to represent that effect. In other words,  $H_1$  which represents heat effect, can be mapped to effect device heater, effect variable 30% is converted to heater's capable control message heater level 1(if heater had level 3, effect variable 100% could be heater level 3). There are some devices that they could not uses duration so they only be controlled by start, stop control message. Fig. 9 (b) shows when there are only that kind devices, 9 sensory effects can be mapped to 16 control command stored in the effect device controller time table. When real-sense media is played, control commands are transmitted to the effect devices by the media play time.

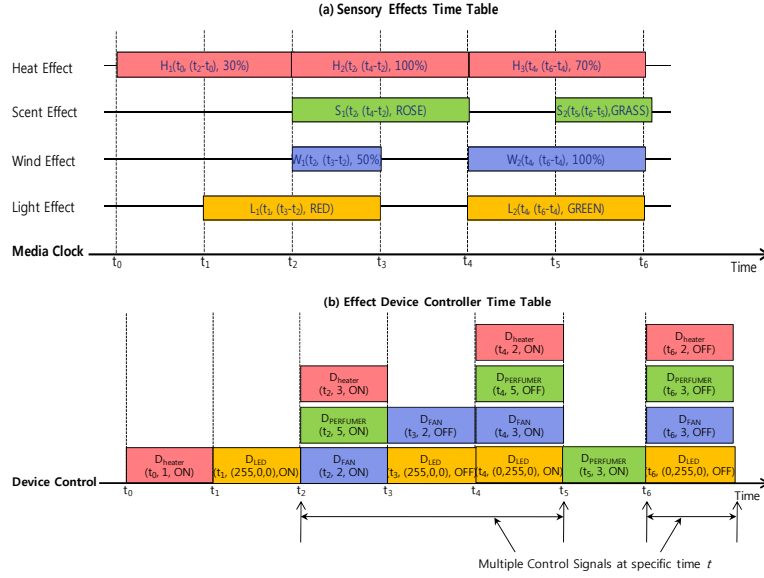


Fig. 9. Sensory Effect Time Table and Effect Device Controller Time Table

### 3.4 Effect Device Synchronization Algorithm

To give real-sense effect to users, same as audio/video synchronization, effect device must be activated at the exact time of the scene to be played. Therefore, we need algorithm which calculates device activation time  $D(t)$ . In the home server, event occurs according to the time line of the main track media.

$$D(t) = MC(t) - \delta(t) - N(t) \quad (2)$$

$$\delta(t) = \frac{1}{n} \sum_{i=1}^n \delta_i(t), N(t) = \frac{1}{k} \sum_{k=1}^k N_k(t)$$

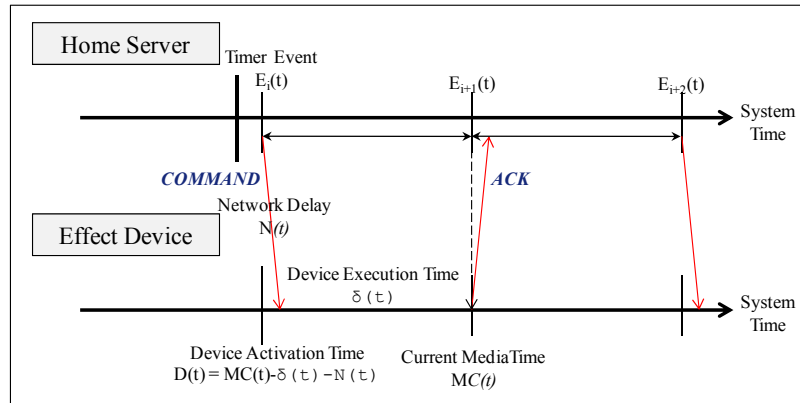


Fig. 10. Effect Device Synchronization algorithm



According to the formula (2), when the sync manager sends control command to the device controller at  $D(t)$  time, the command is analyzed and decomposed into control command type, interface, control value, and start time.

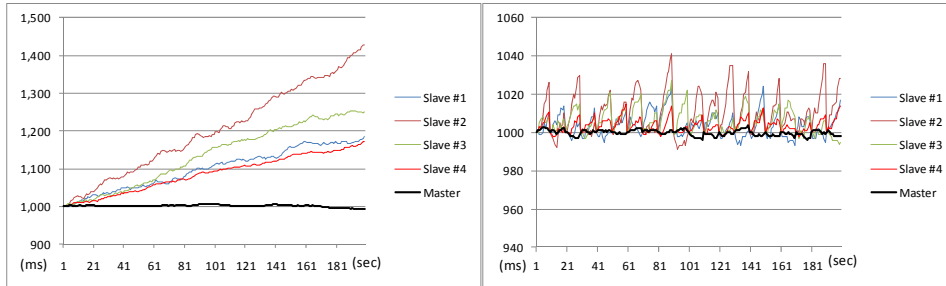
## 4 Evaluation Test

This synchronization algorithms we proposed are tested mainly in an experience room or an exhibition room. In these kind of room we used multi-track media as to display 3 dimensional media, and there are various kind of effect devices that are capable of making real-sense representation. Currently presented media is using 2 main tracks, 4 sub-tracks and 14 kinds of effect devices. 2 main tracks use beam projects for displaying right and left images, 2 sub-tracks in 4 are displayed in UMPC(Ultra Mobile PC) while users are moving, and the other 2 sub-tracks are displayed in PCs embedded in the motion chair. Table 1 shows the information of the tested video, audio.

**Table 1.** Video and Audio Information of the Test Media

|       |                 |               |       |                 |                |
|-------|-----------------|---------------|-------|-----------------|----------------|
| Video | Codec           | MPEG-4(H.264) | Audio | Codec           | AAC            |
|       | Width,Height    | 720 x 480     |       | Bit per sample  | 16bit/sample   |
|       | Frame rate      | 29.97         |       | Channels        | 2 channels     |
|       | Average bitrate | 2.54Mbits/sec |       | Average bitrate | 1411 Kbits/sec |

The test methods for the multi-track media are as follows. First of all, we displayed each track without carrying out synchronization algorithm. We compared time gap between home server and each client whenever home server broadcasts synchronization time in every 10 seconds. As shown in fig. 11, the error can be reduced to 10% compared with not using synchronization algorithm.



**Fig. 11.** Evaluation Results for the multi-track media synchronization

The SMP 8635 hardware video/audio decoder(Revision C, Production ChipMRUA library version is *mrua\_SMP8634\_2.8.2.0*) currently used in the home server is based on the mips chip operated by embedded linux. The decoder's performance is not so good because it was developed only for the purpose of playing audio/video. As shown

in Fig. 11 (a), there was 400 milliseconds time difference between the home server and slave #1, slave #2 when media was played over 200 seconds because slave media was played in a normal PC. But, if synchronization algorithm was applied, the time difference occurred 40ms compare with the case of not using synchronization algorithm.

**Table 2.** Devices for the Evaluation Test

| Effect Property  | Device           | Interface | $\delta D(t)$ | $\delta N(t)$ | Number |
|------------------|------------------|-----------|---------------|---------------|--------|
| Heat Effect      | Heater           | RS-232    | 550ms         | 5ms           | 3      |
| Wind Effect      | Fan              | RS-232    | 500ms         | 10ms          | 2      |
| Scent Effect     | Scent Generator  | RS-232    | 1s            | 12ms          | 2      |
| Light Effect     | Dimmer           | RS-485    | 800ms         | 7ms           | 1      |
|                  | Flash            | USB       | 70ms          | 28ms          | 2      |
|                  | Color Light(LED) | RS-485    | 120ms         | 7ms           | 4      |
| Vibration Effect | Vibration        | RS-485    | 150ms         | 7ms           | 4      |
|                  | Motion           | RS-485    | 100ms         | 7ms           | 1      |
| Shade Effect     | Curtain          | RS-485    | 5.3s          | 10ms          | 1      |
| Other Effect     | Water Sprayer    | RS-485    | 200ms         | 7ms           | 2      |
|                  | Air Zet          | RS-485    | 200ms         | 7ms           | 2      |
|                  | Tickler          | RS-485    | 200ms         | 7ms           | 1      |
|                  | Phone            | Ethernet  | 7s            | 1ms           | 1      |
|                  | Web Browser      | Ethernet  | 1.2s          | 1ms           | 2      |

We used 7 kind and 14 effect devices to represent real-sense effect as shown in table 2. We used electronic appliances that have MCU(Micro Controller Unit) to represent real-sense effect like "Heat", "Wind", "Shade". "Light" effect devices are categorized to "Dimmer" supporting fade in and fade out, and "Flash" to show an effect of lightning, spark, and "LED" supporting RGB color to express real colors.

To show a "Vibration" effect, there are "Vibration" chairs that users feel oscillation from the seat and backrest, also this vibration chair is capable of giving a "Motion" effect like swing, up/down, shaking, forward/backward, left/right turn. And, there are "Other" effect like "Water Sprayer", "Air Zet", "Tickler". To interact with PC, there are "Web Browser" effect that opens related web pages in a specific scene of a movie. The "Phone" effect rings a real world phone or a cell phone when there is a scene of phone calls.

We recorded the time of  $\delta D(t)$  since a control signal was delivered to a device, executed and reached its effect to a person in two meters. And, we measured the time  $\delta N(t)$  that MCU of each device sent a control signal and received a response message. Fig. 12 shows the result with synchronization algorithm and without it. We checked the maximum time difference from media time and device execution time, and it also contains the re-transmit time when control signal did not work correctly. After using synchronization algorithm, devices can be controlled within the time error 25ms.

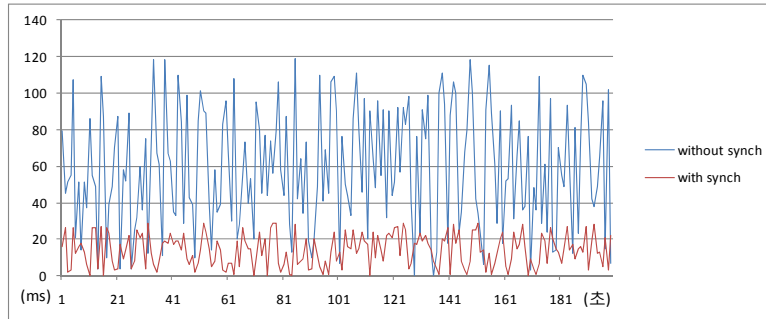


Fig. 12. Evaluation Results for the Effect Device Synchronization Algorithm

## 5 Conclusion

Recently, requirements for media that gives users real-sense effects like real-sense movie, real-sense broadcast, real-sense news, real-sense education is increasing. But, previous playback method like SMSD cannot provide users with the real-sense effect, because one device cannot have all capability of effect representing. Therefore, in this paper, we explained the concepts and architecture of the real-sense media service system, and suggested the playback algorithms for operating multiple audio/video devices with multiple effect devices together. As shown in evaluation test and field tests, people feel awkward when synchronization algorithms did not work. Therefore in the future, these synchronization algorithms can do an important role when real-sense broadcasting service like IPTV will be widely spreaded. The future research issue is a study for the QOS of the multi-track media and the effect device control.

**Acknowledgments.** This work was supported by the IT R&D program of MKE/KEIT, [2007-S010-03, Development of Ubiquitous Home Media Service System base on SMMD]"

## 6 References

- [1] Choi, B.S., Joo, S.H., Lee H.R. and Park, K.R., "Metadata structure for Media and Device Interlocking, and the Method for Mapping It in the Media Format," *Advances in Information Sciences and Services*, 2007, pp186~190.
- [2] Segui, F.B., Cebollada, J.C.G. and Mauri, J. L., "Multiple Group and inter-stream synchronization techniques: A comparative study," *Information Systems*, 2009, pp.108~131.
- [3] Yun, J.K., Shin, H.S., Lee, H.R. and Park K.R., "Development of the Multiple Devices Control Technology for Ubiquitous Home Media Service System," *International Conference on Ubiquitous Information Technology & Applications*, 2007, pp.404~408.
- [4] Timmerer, C., Hasegawa, S., (eds.) "Working Draft of ISO/IEC 23005 Sensory Information," *SO/IEC JTC 1/SC 29/WG 11/N10475*, Lausanne, Switzerland, February 2009.
- [5] Waltl, M., Timmerer, C. and Hellwagner, H., "A Test-Bed for Quality of Multimedia Experience Evaluation of Sensory Effects", *Proceedings of the First International Workshop on Quality of Multimedia Experience (QoMEX 2009)*, July 29-31, 2009.