

# Implementation of a Novel Audio Network Protocol

Jaeho LEE<sup>(1)</sup>, Hyoungjoon JEON<sup>(1)</sup>, Pyungho CHOI<sup>(1)</sup>, Soonchul KWON<sup>(2)</sup>, Seunghyun LEE<sup>(2)</sup>

<sup>(1)</sup> *Department of Plasma Bio Display  
Kwangwoon University*

Seoul, Korea; e-mail: sound@kw.ac.kr, fotojoon@gmail.com, phchoi@hotmail.com

<sup>(2)</sup> *Graduate School of Information and Contents  
Kwangwoon University*

Seoul, Korea; e-mail: {ksc0226, shlee}@kw.ac.kr

(received August 2, 2017; accepted July 13, 2018)

Recently, the rapid advancement of the IT industry has resulted in significant changes in audio-system configurations; particularly, the audio over internet protocol (AoIP) network-based audio-transmission technology has received favourable evaluations in this field. Applying the AoIP in a certain section of the multiple-cable zone is advantageous because the installation cost is lower than that for the existing systems, and the original sound is transmitted without any distortion. The existing AoIP-based technology, however, cannot control the audio-signal characteristics of every device and can only transmit multiple audio signals through a network. In this paper, the proposed Audio Network & Control Hierarchy Over peer-to-peer (Anchor) system enables all audio equipment to send and receive signals via a data network, and the receiving device can mix the signals of different IPs. Accordingly, it was possible to improve the system-application flexibility by simplifying the audio-system configuration. The research results confirmed that the received audio signals from different IPs were received, mixed, and output without errors. It is expected that Anchor will become a standard for audio-network protocols.

**Keywords:** AoIP; DANTE; SR System; audio mixer.

## 1. Introduction

Every object evolves during its adaption to the corresponding environment. Accordingly, audio systems have evolved during their transmission of device signals over audio-shield cables (hereafter “MSC”). Since the invention of the phonograph by Edison (Thomas Alva Edison) in 1877, electric audio devices have evolved in various forms along with the advancement of the corresponding industry. The audio of the industrial field is transmitted through various audio equipment. For the configuration of the general audio system, the MSC was used for signal transmission and component connection until several years ago. But with the introduction of the new audio over internet protocol (AoIP)-using audio-transmission technology, it has been possible to apply the unshielded twisted pair (UTP) cable in the configuration of specific sections of the general audio system. Furthermore, a variety of audio equipment for which this new technology is employed has increased, and this equipment is now spread throughout the market.

In this paper, by ruling out the traditional system for which only the MSC is used, a review of the existing AoIP UTP-cable-using transmission-technology types such as CobraNet and Dante is performed, and an audio-network protocol that is more advanced than the existing AoIP technology is proposed. Section 2 comprises descriptions of the AoIP technology in the market and the audio-system configuration based on it. Section 3 presents an explanation of the proposed system. Section 4 presents the experiment results of the proposed system. Lastly, Sec. 5 comprises the conclusion and a future development plan regarding the proposed AoIP-based audio-system technology (CHO *et al.*, 2017; KIM *et al.*, 2011).

## 2. AoIP-Based audio system

### 2.1. AoIP technology

For the open-system network communications among different types of computers, the layered proto-

Table 1. Comparison of Audio-Network Protocols.

Technology	Date	Transport	Transmission scheme	Control communications	Network capacity	Latency	Sampling rate	Sound Mixing	Sound Control
RAVENNA	2010	Any IP medium	isochronous	IP, RTSP, Bonjour	unlimited	variable [n]	384 kHz and DSD	No	No
EtherSound	2001	Ethernet data link layer	isochronous	Proprietary	64[i]	84–125 $\mu$ s + 1.4 $\mu$ s/node	96 kHz	No	No
AVB (using IEEE 1722 transport)	2011	Enhanced Ethernet	isochronous	IEEE 1722.1	unlimited	2 ms	192 kHz	No	No
CobraNet	1996	Ethernet data link layer	isochronous	Ethernet, SNMP, MIDI	unlimited	1 <sup>1/3</sup> , 2 <sup>2/3</sup> , and 5 <sup>1/3</sup> ms	96 kHz	No	No
Dante	2006	Any IP medium	isochronous	IP, Bonjour	unlimited	84 $\mu$ s or greater [h]	192 kHz	No	No
Proposed System (Anchor)	2017	Ethernet II (DIX Ethernet)	asynchronous	IP, UDP	12ch	10.7 ms	48 kHz	Yes	Yes

col of the International Organization for Standardization (ISO)’s Open Systems Interconnection (OSI) 7 is used. Table 1 is a comparison of the features of the existing AoIP-based-technology types. The bottom line of Table 1 shows the features of the proposed Audio Network & Control Hierarchy Over peer-to-peer (Anchor) system of the present paper. While the characteristics of each AoIP technology applied in the industrial field include a high sampling rate and multichannel transmission, the Anchor-system technology mixes the signals from different IPs and controls audio signals even though it uses the same AoIP technology. The AoIP-using audiotransmission technology, with various features, is available on the market.

Table 2 shows the number of products using the major technology, which have been released to the current market, and it was presented by RH Consulting

at Infocomm in 2016. Hereafter, an analysis of the systemic features of CobraNet and Dante, which are the marketplace product leaders, is presented (KANG *et al.*, 2008; KIM *et al.*, 2011).

Table 2. Quantities of the products using audio-transmission technology released to the market.

Technology	2014	2015	2016
Ravenna	37	43	117
EtcherSound	77	78	107
AVB		11	14
CobraNet	315	351	373
Dante	288	391	706

Figure 1 shows the classifications of the network audio-transmission technology. CobraNet is a com-

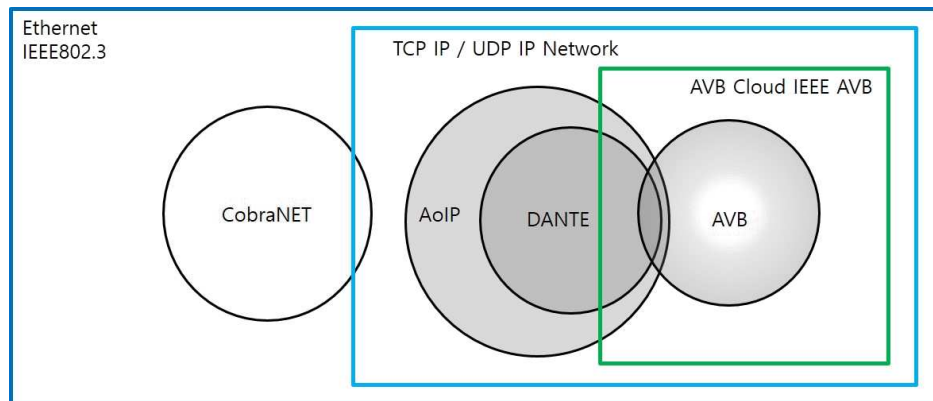


Fig. 1. Network-audio-technology classifications.

munication technology that utilizes the *OSI Layer 2* datalink layer; this is not an internet protocol (IP)-communication technology, although a dynamic IP address is allotted upon its booting. Particularly, CobraNet adopts neither the IEEE1588-standard time-transfer protocol that synchronizes a network precisely, nor the IEEE802.1 internet working standard. Instead, it utilizes its unique bundle package for the network synchronization and communication processes. Dante is a communication technology that transmits packets from the sender to the receiver through multiple networks, and it uses IPs like Transmission Control Protocol (TCP)/IP and User Datagram Protocol (UDP)/IP. On a Gigabit network, Dante can communicate over 512 channels with 48-kHz sampling data, or over 256 channels with 96-kHz sampling data, and on a 100-Mbps network, it communicates over 48 channels with 48-kHz sampling data, or over 24 channels with 96-KHz sampling data (KANG *et al.*, 2008; WEE *et al.*, 2013; MACIEJKO, 2015).

The Audio Video Bridging (AVB) technology has been developed to solve the problems of the connection complexity between audio and video systems and the cost of highly priced cables. AVB technology transmits time-sensitive audio and video data in real time with no data loss by means of the simple local area network (LAN) cable. This new technology has been adopted as an international standard using the name of the specific AVB standard, *IEEE 802.1.AVB*. The applications of AVB technology have been applied for public-address systems, professional audio systems, and automotive networks. The research and development (R&D) projects for the AVB-protocol control and analysis methods are actively conducted at the present time (PARK *et al.*, 2015; WEE *et al.*, 2017; KWON *et al.*, 2015).

### 2.2. Existing AoIP audio-system configuration

The existing AoIP soundsystem configuration is based on an audio mixer that is located in the sound-control room. The audio mixer receives and transmits the signals of the input/output devices at each site through the MSC or the LAN cable. It also determines the destination of the input signal as well as mixing and balancing the sound to output it.

Figure 2 shows the existing AoIP-based audio-system configuration. With the centering of the audio mixer which is installed in the sound-control room, the typical configuration of the system involves the employment of the MSC for the signal transmission and the equipment reception. But recently the LAN cable has been used in cases where the audio signal is transmitted over a long distance or via multiple channels using the AoIP technology. The advantages of the AoIP-using signal transmission are the improvement of the sound degradation and the reduction of the wiring cost compared with the MSC. For the existing AoIP systems like CobraNet and Dante, however, the audio-system configuration basically remains unchanged, with only the partial replacement of the MSC with the LAN cable in a specific section, and this means these systems cannot take advantage of the AoIP technology. The current audio-system configuration is affected by noise-potential factors due to the ground-potential difference according to the distributed equipment installation. In Sec. 3, it is proposed that the Anchor-system signal transmission can be achieved using the LAN cable, thereby reducing the piping and wiring costs, preventing the noise that may occur due to the ground-potential difference, and enabling the sound control regardless of the location (RYU *et al.*, 2012; KIM, 2007; LEE *et al.*, 2017; CHURCH *et al.*, 2010; RUMSEY, 2012).

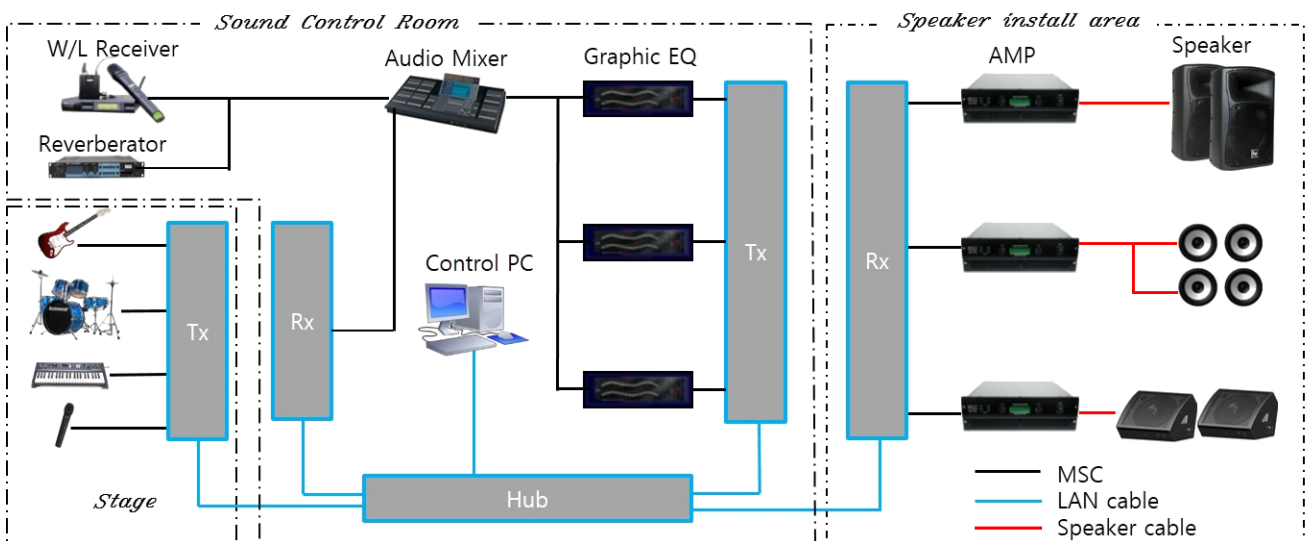


Fig. 2. Diagram of the audio over internet protocol (AoIP)-based audio system.

### 3. Proposed system

#### 3.1. Anchor-based audiosystem configuration

The difference between the Anchor audio-system configuration and the existing AoIP audio-system configuration can be distinguished by the presence of the audio mixer and the full use of the LAN cable in the system for audio-signal transmissions. The Anchor system simplifies the audio-system configuration by allowing all audio devices to send and receive audio signals over the network and by mixing the audio signals from different IPs in the receiving OAC (On-site Audio Center).

Figure 3 presents a diagram of the Anchor-based audio system. The transmitting OAC is connected to the audio source multicasts pulse-code modulation (PCM) data using the UDP protocol, or it transmits the data using the designated IP. The system also comprises an *audio codec* allowing to control the audio signal. As it is different from the existing audio-system configuration, the receiving OAC simplifies the audio system by enabling audio-mixer functions and a number of sound-quality-correction devices including the graphic equalizer (EQ), time delay, and reverberator. The OAC is a network-based device that transmits and receives signals using its connection to the distributed input/output audio equipment. The OAC is a firmware-based system that handles the high-speed transmission and receipt of signals.

Figure 4 shows the concepts of the transmitting and receiving OACs. The transmitting OAC can receive the signals from microphone or line inputs. Fig. 4a shows that the transmitting OAC first receives the input signal from the audio-source device, converts it from analog to digital using the audio codec, transmits the sig-

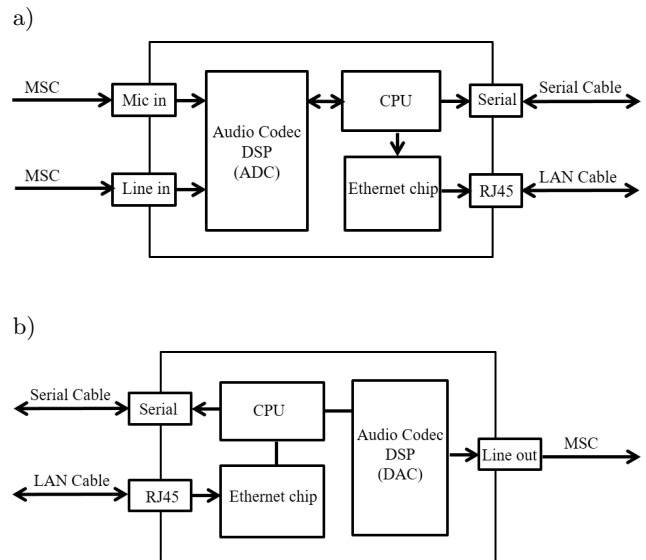


Fig. 4. OAC-block diagram: a) transmitting OAC and b) receiving OAC.

nal to the central processing unit (CPU) by controlling the reverberation level and the audible-frequency level, controls the audio volume and the mute, and lastly it records the IP and port information of the receiving OAC in the electrically erasable programmable read-only memory (EEPROM) of the CPU. The Ethernet chip transmits the audio data that are received from the CPU in the form of the *Ethernet II* (DIX Ethernet) frame at 100 Mbps. Each of the transmitting and receiving OACs consists of a serial communication (RS-232) port that enables the direct-control input/output equipment or the remote control. Figure 4b shows that the signal processing occurs in the reverse order of Fig. 4a.

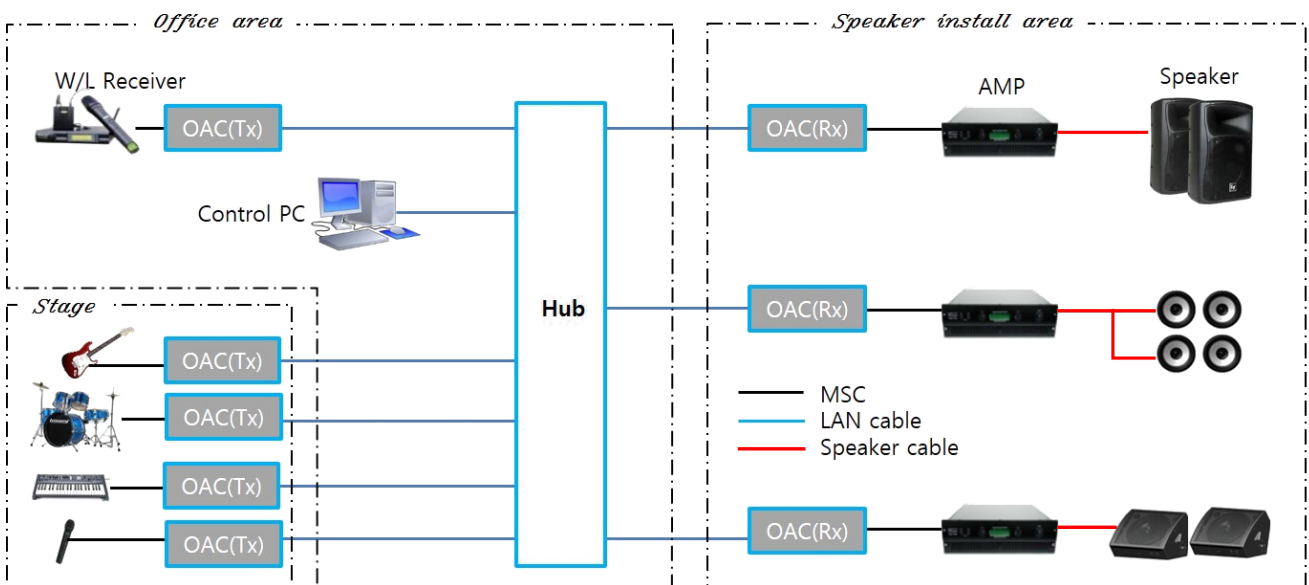


Fig. 3. Diagram of the Anchor-based audio system.

3.2. Audiodata mixing of the receiving OAC

It is generally considered that, for the centering of the audio mixer, an audio source like a microphone or a compact disc (CD) player is the input, whereas the power amplifier (PA) is the output. The receiving OAC connects with the PA to either mix or control the audio signals of different IPs.

Figure 5 is a conceptual diagram of the CPU signal processing of the receiving OAC. The audio data that are received through the Ethernet II are stored at the buffer of the designated port after the callback function is called from the port. The ARM Cortex M4 processor (Arm, U.K.) invokes the Mix function that then mixes the received audio data by referring to the first number of the address of the receiving buffer that is for data storage. The mixed data calls the Play function of the

ARM Cortex M4, and next the Play function transfers the data to the audio codec.

Figure 6 is a conceptual diagram of the mixing of the audio data that are stored in the receiving buffer of the receiving OAC. The audio-data mixing creates the mixed data by adding the hexadecimal data that are stored at the same address of each receiving buffer. This process can be expressed using the programming language, as follows:

```
For (i=0, i<1024, i=i+1)
    Mix buffer_Mixed_data_address(i) =
         $\sum_{k=1}^{12}$  Receiving buffer # k_address(i). (1)
```

The formula of the programming language of (1) is the algorithm that is applied for the addition of the

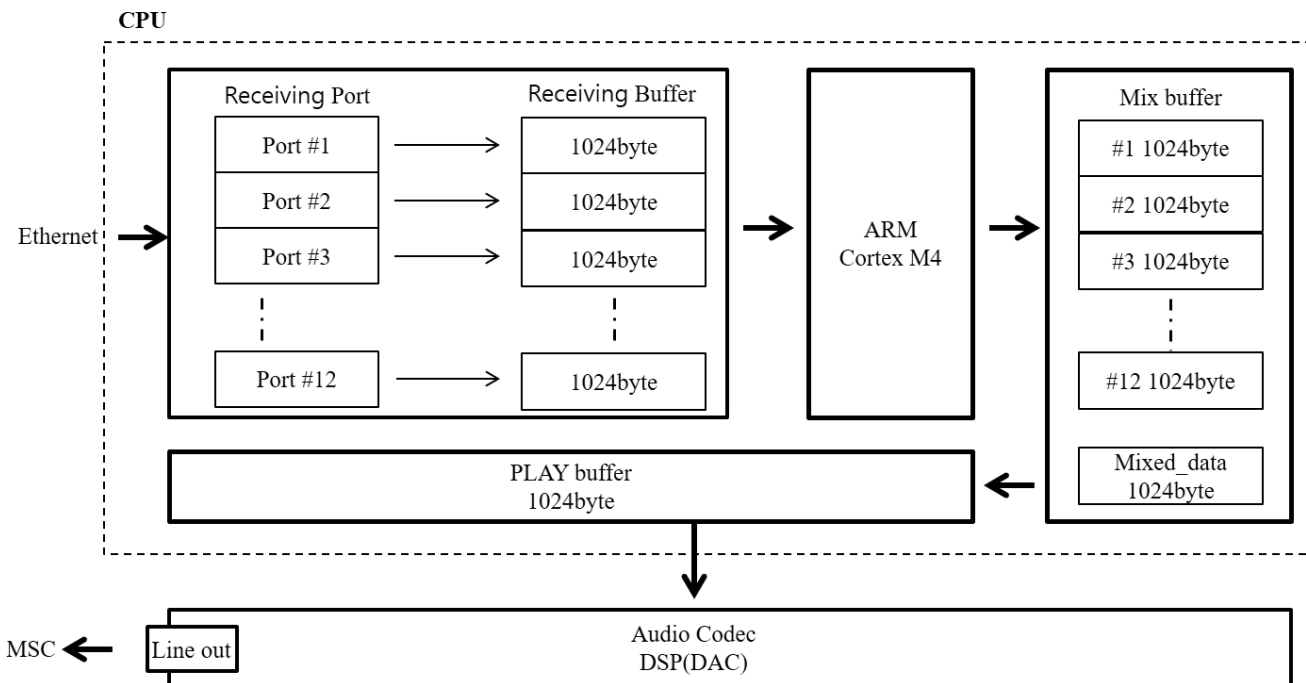


Fig. 5. Diagram of the central processing unit (CPU) data processing of the receiving OAC.

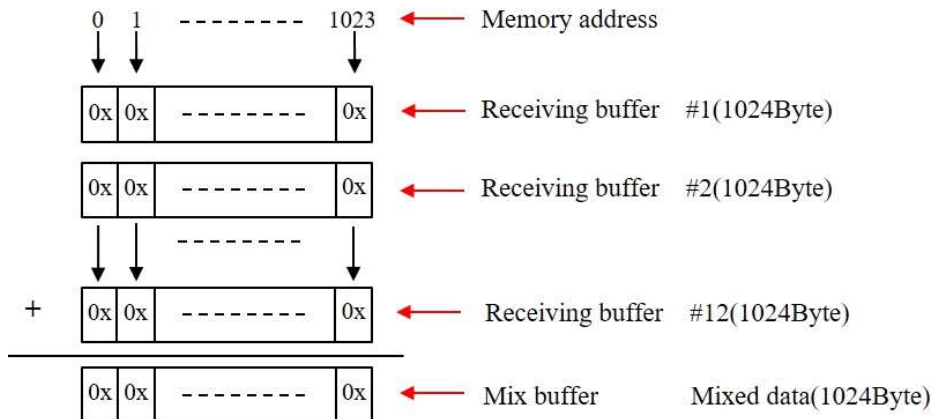


Fig. 6. Receiving-OAC mixing-concept diagram.



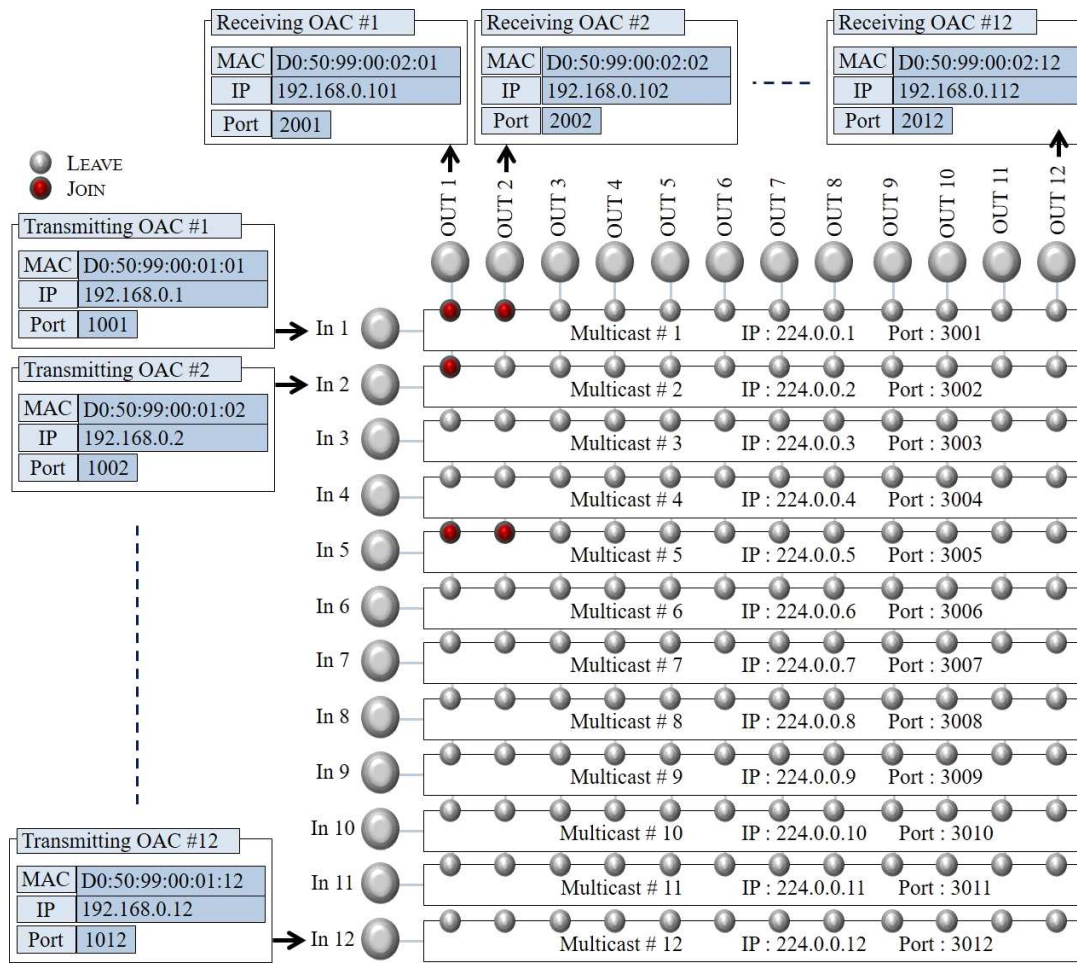


Fig. 7. Receiving-OAC mixing processor.

data of the same address value from Address 0–1023 of the 12 receiving buffers, and the result is stored in the Mix buffer.

### 3.3. OAC operation

The separately placed transmitting and receiving OACs can be operated throughout the network regardless of the location. The Anchor system accesses the CPU by connecting to the IPs of the transmitting and receiving OACs, and then it sets the basic information such as the IP, port, and media access control (MAC) address; subsequently, it controls the audio-codec attributes, including the graphic EQ, time delay, reverberator, signal mute, and volume.

Figure 7 shows a schematic diagram of the receiving OAC receiving multicasted audio data from the transmitting OAC. The transmitting OAC sends the audio data by designating the destination IP as the *multicast IP*, and then the receiving OAC joins the multicast IP and receives the corresponding audio data. Additionally, in the case of joining of two or more multicast IPs in the receiving OAC, the audio data are typically mixed and played.

## 4. Experiment results

The OACs implement the Anchor system by processing audio signals with a small latency, whereby the system receives and mixes multiple channels. Section 4 presents an analysis of the duration of the audio-data transmission by the transmitting OAC, as well as a discussion of the audio-data packet and an analysis of the audio-data processing in the receiving OAC.

### 4.1. Audiodata processing of the transmitting OAC

Up to 1500 bytes of audio data can be transmitted over the Ethernet. The transmitting OAC processes the input audio source according to the unit of 32 bytes using the audio codec. Since the maximum transmission capacity is 1500 bytes, the transmitting-OAC audio codec processes audio data according to the unit of 1024 bytes, which is  $32 \times 32$  bytes. In general, the program-processing analysis is debugged using the *Time* or the *PrintF* function, but in this case, it is impossible to analyze data that are processed rapidly because of the delay caused by the corresponding function processing. Therefore, the oscilloscope performs

the signal-processing time analysis using the high/low signals of the general-purpose input/output (GPIO). The GPIO is able to perform the check without a delay because the operation of the CPU is possible with the maximum clock-frequency value of 168 MHz. The processing times of the transmitting and receiving data can be checked using the GPIO pin (PD3 and PD7) of the ARM Cortex M4.

Figure 8 shows the analysis of the duration for the transmitting OAC to transmit audio-codec-processed audio data. The horizontal grid represents time, while the vertical grid represents voltage (existence or nonexistence of the audio-data transmission). The time unit is represented by 100  $\mu$ s, the voltage unit is represented by 2 V, and the ground is 1 level (0 V). The  $T$  level is the trigger level for the waveform measurement. The yellow waveform represents the data for the measurement of the time in terms of the audio-data transmission, for which the GPIO pin of the transmitting OAC is used. The transmission time can be measured using the change time interval between the high value (3.3 V) that indicates the lack of a signal and the low

value (0 V) that indicates the presence of a signal. Figure 8 (a) shows that 1024 bytes were transmitted within 100  $\mu$ s.

#### 4.2. Audio-data packet

The transmitting OAC transmits the packets in the Ethernet II (DIX Ethernet) frame, and the packets transmitted via the Ethernet are stored in the receiving buffer of the receiving OAC as hexadecimal data. The audio data are transmitted in the format shown in Fig. 9

Figure 9 shows the audio data from the transmitting OAC that are sent through the network and are captured by the receiving OAC. Part (a) is the header part for the data transmission. The header contains the IP address, MAC address, and protocol-type information for the network. Part (b) is the audio-information identifier and comprises the audio-characteristic information including audio format, sampling rate, and stereo. Part (c) is the real audio data that are used in the mixing.

#### 4.3. Audiodata processing of the receiving OAC

The receiving OAC mixes and plays the received data at regular intervals. Figure 10 shows the state of one cycle of the received-data processing in the receiving OAC. The yellow waveform in Fig. 10 is the received-data-processing waveform, and the received waveform is cyan. The horizontal aspect of the grid shows the time in 1-ms intervals. Part (a) of the yellow waveform shows the audio data from the different transmitting OACs, where the processing time is 3.2 ms. Part (b) shows the processing time for the mixing and playing of the received audio data, where the duration is 7.5 ms. Part (c) shows the full process of the receiving, mixing, and playing of the received data, where the duration is 10.7 ms.

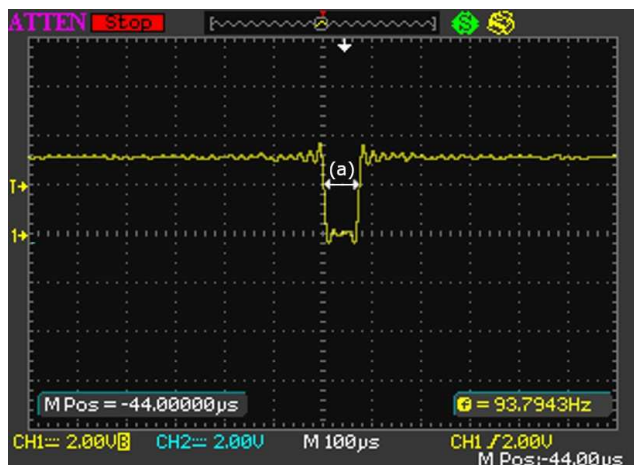


Fig. 8. OAC-transmission time data.

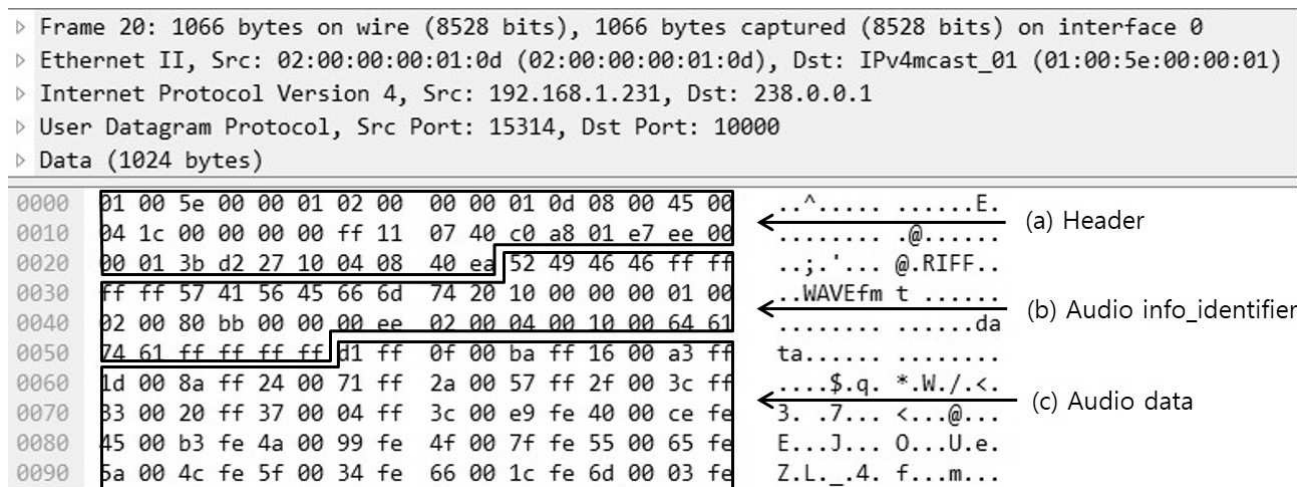


Fig. 9. Transmitting-OAC data format.

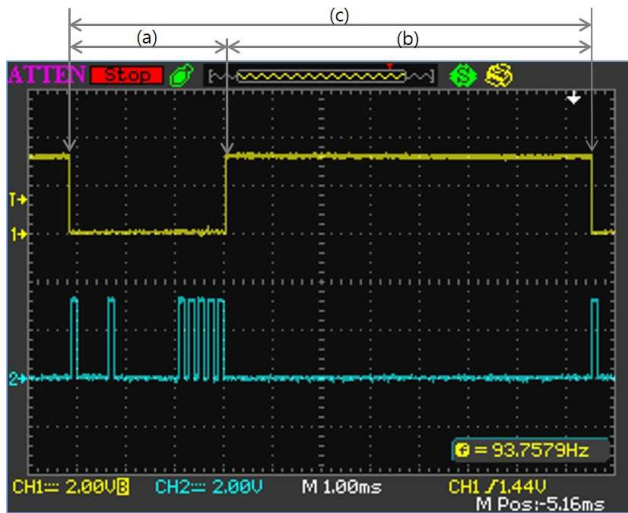


Fig. 10. Processing time of a receiving OAC in one cycle.

Figure 11 shows periodic representations of the full processing in the receiving OAC (yellow waveform) and the mixing processing time (cyan waveform). Part (a) on the yellow waveform is one cycle (10.7 ms) of the processing of the receiving, mixing, and playing, and this is also the case for Fig. 10 (c). The cyan waveform represents the processing of the mixing during one cycle. Since the horizontal grid is 2.5 ms, it can be confirmed that the mixing processing time is 0.5 ms. The sum of the mixing and playing times of the audio data is 7.5 ms, as shown in Fig. 10 (b), which means the playing time is 7 ms.

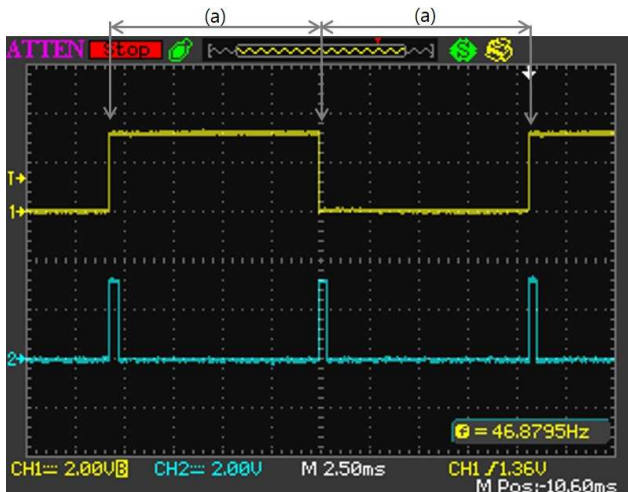


Fig. 11. Mixing processing time of a receiving OAC.

## 5. Conclusion

In this paper, an audio system is configured and tested using the proposed Anchor application that is a novel protocol. The results are as follows: First, since all of the sound equipment are built on the Anchor basis, the system configuration became simplified, whereby the LAN-cable wiring was also simplified, and

all of the systems could be operated on the network basis; second, the input and output audio signals could be controlled remotely; and third, the receiving OAC replaced the audio mixer by mixing multiple audio signals from the different IPs.

The results of this study show that the configuration of the Anchor-based industrial audio system can improve the systemic cost and efficiency. The two major tasks of the proposed Anchor system of this paper are as follows:

- 1) Firstly, it is necessary to expand the Mixing channel (currently  $12 \times 12$ )
- 2) Secondly, it is necessary to shorten the audio-data processing time (10.7 ms)

These tasks will be solved by the improvement of both the network transmission technology and the CPU performance. It is expected that Anchor will become a standard for audio-network protocols.

## Acknowledgment

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-R0992-16-1008) and the Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. 2017-0-00833).

## References

1. CHO J.H., KIM K.W., KIM D.I. (2017), *Development of integrated mixer controller for digital public address* [in Korean], *Journal of Institute of Internet Broadcasting and Communication*, **17**, 1, 19–24.
2. CHURCH S., PIZZI S. (2010), *Audio over IP: building Pro AoIP systems with Livewire*, Focal Press, pp. 113–141.
3. KANG J.A., KIM H.K. (2008), *Implementation of a High-Quality Audio Collaboration System Over IP Networks* [in Korean], *Proceedings of The HCI Society of Korea Conference*, pp. 218–223.
4. KANG M.S., LEE S.W., PARK Y.S. (2008), *Implementation of Local Distribution Audio System Based on AoIP* [in Korean], *Journal of Korean Institute of Maritime Information and Communication Sciences*, **12**, 11, 2165–2170.
5. KIM D.H., WEE J.W. (2013), *A modem design for audio signal transmission* [in Korean], *Korea Electronics Technology Institute*, **34**, 1, 1468–1470.
6. KIM J.S., SONG C.W. (2011), *Development of integrated public address system for intelligent building* [in Korean], *Journal of Korea Intelligent Systems Society*, **21**, 2, 212–217.
7. KIM Y.M. (2007), *A Study on the Audio Transmission Flow Control Using Mixing in real-time communication* [in Korean], *Proceedings of Korean Institute of Communication Sciences*, pp. 997–1000.



8. KWON K.W., PARK K.W., WEE J.H., SONG B.C. (2015), *Design and Implementation of an Embedded Media Storage System for AVB Networks* [in Korean], Proceedings of Korean Institute of Communication Sciences, pp. 779–780.
9. LEE J.H., KWON S.C., LEE S.H. (2017), *Implementation of On-site Audio Center based on AoIP*, International Journal of Advanced Smart Convergence, **6**, 2, 51–58.
10. MACIEJKO W. (2015), *The effect of voice over IP transmission degradations on MAP-EM-GMM speaker verification performance*, Archives of Acoustics, **40**, 3, 407–417.
11. PARK K.W., KIM D.H., KWON K.W., SONG B.C. (2015), *AVB based IEEE 802.1AS performance test* [in Korean], Proceedings of Korean Institute of Communication Sciences, pp. 1342–1343.
12. RUMSEY F. (2012), *Audio networking*, AES: Journal of the Audio Engineering Society, **60**, 4, 282–286.
13. RYU S.H., KIM K.W., KIM H.G. (2012), *Volume normalization algorithm for audio mixer in multi-participant conference system* [in Korean], Proceedings of Institute of Electronics Engineers of Korea, pp. 622–623.
14. WEE J.H., PARK K.W., SONG B.C. (2017), *Implementation of control/management protocol based on audio/video bridging networks* [in Korean], Proceedings of Korean Institute of Communication Sciences, pp. 1066–1067.
15. WEE J.H., PARK Y.S., PARK K.W., SONG B.C., JEUN W.K., (2013), *Audio/video bridging technology for real-time AV transmission* [in Korean], The Journal of The Korea Institute of Communication Sciences, **30**, 6, 69–76.