Audio Engineering Society

Convention Paper 10318

Presented at the 147th Convention
2019 October 16–19, New York, USA



# Discrimination of High-Resolution Audio without Music

Yuki Fukuda[1], and Shunsuke Ishimitsu[1]

[1] *Graduate School of Information Sciences, Hiroshima City University, 3-4-1, Ozuka-Higashi, Asaminami-ku, Hiroshima, 731-3194, JAPAN.*

Correspondence should be addressed to Yuki Fukuda (y-fukuda@hfce.info.hiroshima-cu.ac.jp)

**ABSTRACT**

Nowadays, High-Resolution (Hi-Res) audio format, which has higher sampling frequency (Fs) and quantization bit number than the Compact disc (CD) format is becoming extremely popular. Several of studies have been conducted to clarify whether these two formats can be distinguished. However, most of the studies conducted by the use of music sources to reach a conclusion. In this paper, we will try to bring out the problems due to the use of music sources for the experimental purpose. We will also answer the question related to discrimination between hi-Res and CD formats using sources other than music, such as noise.

## 1 Introduction

In recent years, the audio data named "High-Resolution audio format" (Hi-Res), which has higher sampling frequency and quantization bit number than Compact disc (CD) format, has received a lot of attention [1 - 4]. Many studies have been conducted to distinguish between Hi-Res and non-Hi-Res.

Nishiguchi [2] et al. conducted a Duo-Trio Test for 36 participants to examine if the participants could discriminate between Hi-Res and non-Hi-Res. The study reported that some people of the participants could discriminate between all the stimuli, which has over 21kHz or not. However, no significant results were obtained for a few Hi-res audio formats (e.g. WAV, FLAC, DSD, etc.) from some of the participants (which involved four music college students, a violinist, and a recording engineer).

Mizumachi [3] et al. compared the CD format with the Hi-Res format music source through a test that involved 27 participants. The results showed that the participants could discriminate among these sources with a 57% accuracy. The study results also conducted that there were nine participants who could discriminate between all sources. Seven people with a musical background (such as playing musical instruments or listening to Hi-Res music sources on a daily basis) could differentiate between higher and lower quantization bit numbers than the others.

Suguro [4] et al. presented two different musical sound data that had different quantization bit numbers but the same sampling frequency. These were simultaneously played. The results reported that sound image localization occurred at the side where the source with the larger quantization bit number was played.

However, these previous studies had only used music data for their experiments. We think that these conclusions are not general because these results would be reported only after excluding the effects from the participants' musical background (playing musical instruments, listening to music, etc.) [4, 8, 9].

In this study, we will focus on the fact that all the previous studies have only used music data for their experiments. We consider the possibility of discrimination between Hi-Res and non-Hi-Res formats without any musical sources.

## 2 Definition of High-Resolution Audio

Some organizations have defined Hi-Res [5 - 7] as mentioned below. In this chapter, some of these definitions are shown. We will clarify the existing definition of Hi-Res in this study.

### 2.1 Japan Electronics and Information Technology Industries Association (JEITA)

When defining the Hi-Res audio data, the JEITA has defined the "CD format" as the digital audio data which has 44.1 kHz or 48 kHz sampling frequency and 16 bit quantization [6].

Further, the JEITA defined Hi-Res as the audio data which has higher sampling frequency or quantization bit number than the CD format. The detailed examples to understand this definition is shown in Table 1.

Table 1. Detailed examples of Hi-Res in JEITA [6]

| Fs(kHz) | bits | Hi-Res |
|---|---|---|
| 44.1 | 16 | × |
| 44.1 | 24 | ○ |
| 48 | 16 | × |
| 96 | 16 | ○ |
| 192 | 12 | × |

### 2.2 Recording Industry Association of America (RIAA)

The RIAA has defined Hi-Res as "lossless audio capable of reproducing the full spectrum of sound from recordings that have been mastered from better than CD quality (48 kHz/20 bit or higher) music sources which represent what the artists, producers, and engineers originally intended" [7].

In this study, Hi-Res is defined on similar lines as that of the JEITA.

## 3 Problem Statement

In this chapter, to clarify our motivations in this study, we present five problems that can occur by using music data for identifying Hi-Res audio formats.

- Problem 1: If researchers use restrictive types of music data, then there is a possibility that the results will differ when another type of music data is used.
- Problem 2: The results may depend on the participants' musical background. In other words, the researcher will have to carefully define stimuli [4].
- Problem 3: A previous report states that when listening to music, there is a significant difference in the body responses and emotions between participants who play musical instruments and those who do not [8]. Hence, it can be considered that having a musical background affects subjective evaluation when listening to music.
- Problem 4: Certain types of music adopt different masters to make Hi-Res and non-Hi-Res audio. Hence, these may differ in spite of having the same music title [9].
- Problem 5: Due to the fact that most microphones either have low noise or high bandwidth, there are no musical stimuli which contain high frequency components [10].

Based on these problems, we think that it is necessary to compare Hi-Res to non-Hi-Res without music to gain more general consideration.

## 4 Signal Selection and Quantization

To prepare for the subjective evaluations, we select the signals and quantize them to follow the specification of Linear Pulse Code Modulation (LPCM) [11].

### 4.1 Signal Selection

Instead of music, we select impulse (Gaussian Impulse) and white noise (white Gaussian noise) signals for the examination purpose. These selections are made based on the following conditions:

- The signals have a flat frequency response.
- As the stimuli are not music signals, the musical background of the participants for the subjective evaluation is irrelevant.

- In case of impulse signals, the loads for the participants are very small because of the short time.
- In case of white noise signals, because the magnitude values evolve with time, we believe that they are optimum to consider the effects of quantization bits.
- In case of white noise signals, it is possible to consider the effects of signal length.

In this study, for the evaluation, three impulse signals and three white noise signals are created. These stimuli have different sampling frequencies: 48, 96, and 192 kHz.

White noise signals are made from Mersenne twister random generator.

These stimuli are quantized to satisfy the specification of LPCM [10].

## 4.2 Quantization [10, 11]

The impulse and white noise signals are quantized to convert these values from 64 bit (the length of the mantissa is 52 bit) floating point number $x(t)$ to integer $Y(t)$ using the equation (1)[12].

$$Y(t) = \left\lfloor \frac{x(t)-x_{min}}{x_{max}-x_{min}} (2^n - 1) - 0.5 \right\rfloor - (2^{n-1} - 1) \quad \cdots(1)$$

Then, $t$ means the time, each $x_{max}$ and $x_{min}$ is the maximum and minimum value which the signal $x$ can have, $n$ means the quantization bits and the function $\lfloor \ \rfloor$ means the floor function. Moreover, $x$ have its value inside [-1, 1]. Hence, when the signals are quantized to 16bits, equation (1) means,

$$Y(t) = \left\lfloor \frac{x(t)+1}{2} (2^{16} - 1) - 0.5 \right\rfloor - (2^{15} - 1) \quad \cdots(2)$$

Therefore, $Y(t)$ has its value from minimum $-2^{n-1}$ to maximum $2^{n-1} - 1$.

Thereafter, $Y(t)$ is converted from integer value to discrete decimal value $y(t)$ to play the signal from minimum -1 to maximum 1 given by the equation (3),

$$y(t) = Y(t) \div 2^{n-1} \quad \cdots(3)$$

Figure 1 and Figure 2 are the examples of quantization error in case of 440 Hz tone signal. Figure 3 and 4 show the quantization errors of white noise signals. Hence, the 16 bit LPCM signal has the biggest quantization error in the 16, 24, and 32 bit LPCM signals.

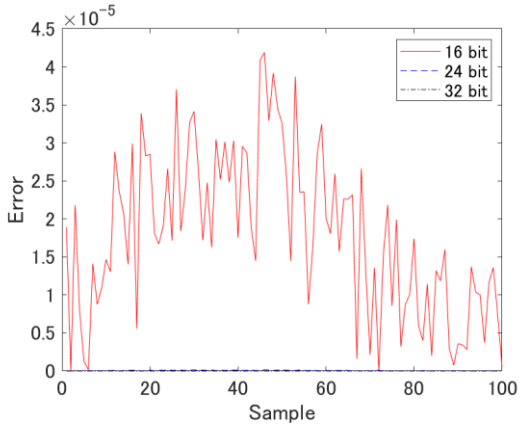


Figure 1. Quantization errors at 440 Hz in each 16, 24, and 32 bit LPCM signal

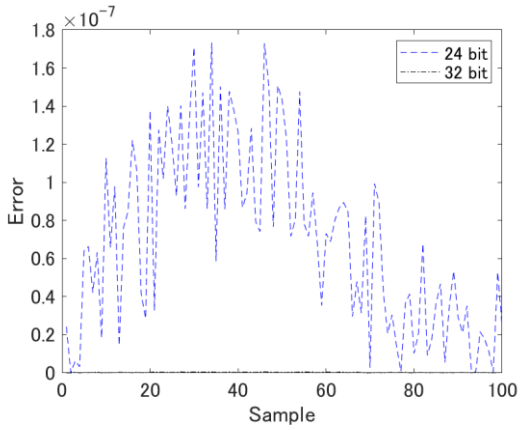


Figure 2. Quantization errors at 440 Hz in each 24 bit and 32 bit LPCM signals

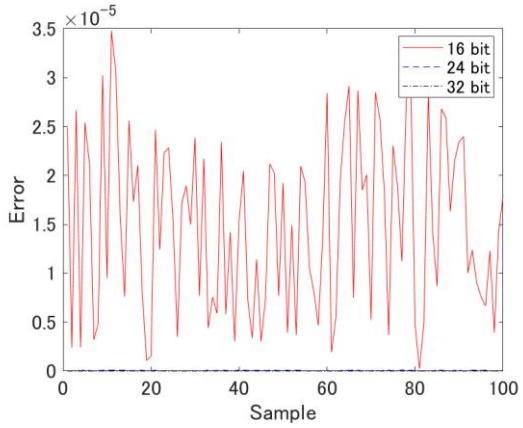


Figure 3. Quantization errors at white noise in each 16, 24, and 32 bit LPCM signal

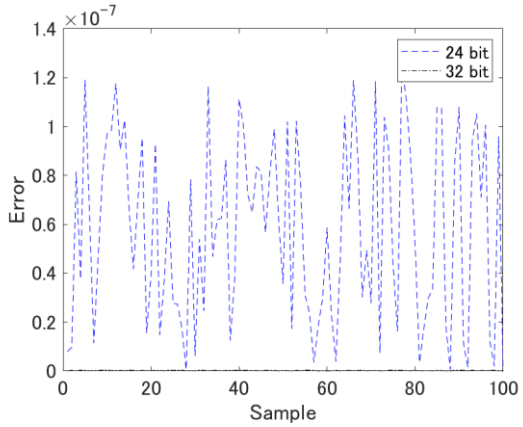


Figure 4. Quantization errors at white noise in each 24 bit and 32 bit LPCM signal

Through these processes, three impulse signals and three white noise signals are made to have magnitude differences within 1 dB.

The impulse response of the finite impulse response (FIR) filter to prevent aliasing noise [13] in the digital-to-analog converter (DAC) used in the playing system is shown in Figure 5. The actual impulse signal played is similar to the waveform shown in Figure 5.

## 5 Experiments

To consider the possibility of discrimination between the Hi-Res and non-Hi-Res without music data, we conduct two experiments with the impulse and white noise signals. The experimental environments are listed below and the schematic of the experimental system is shown in Figure 6. The signals are played in each sampling frequency by using ASIO (Audio Stream Input/Output). So, no lowpass filter is used before playing the stimulus except for the DAC.

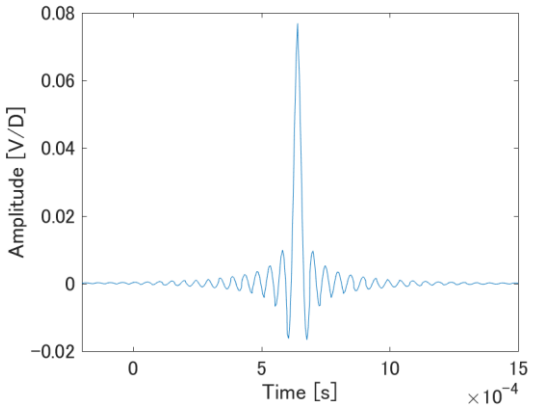


Figure 5. The impulse response of the DAC in the playing system (average is 100 times)

Table 2. Experimental environment

| Experimental chamber | Anechoic chamber |
|---|---|
| USB-DAC | FOSTEX HP-A4BL |
| Headphones | Sennheiser HD-650 |
| Loudspeakers | ECLIPSE TD-M1 |

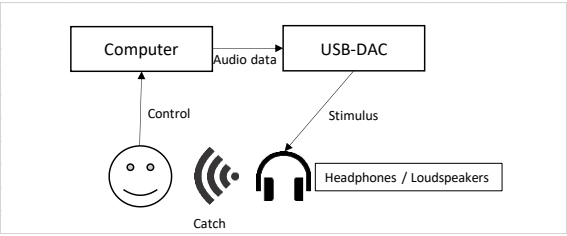


Figure 6. Schematic of the experimental system

### 5.1 ABX Test

First, we adopt an ABX test [14]. In this, the participants will need to identify signal X as either the signal A or B to research the possibility of sampling frequency discrimination among these three sampling frequencies.

First, the sound volume of the playing system is controlled by the participant to decrease the participant's load.

Next, the participants will need to identify signal X as either the signal A or B. Each participant will answer four questions in a combination of sampling frequencies. The three combinations of sampling frequencies provided are 48 kHz vs. 96 kHz, 48 kHz vs. 192 kHz, and 96 kHz vs. 192 kHz.

Furthermore, the tests are conducted in two presentation environments: one with loudspeakers and another one with a pair of headphones. Therefore, the total number of questions answered by each participant is 24. After the tests, the participants get to rest for 5 minutes.

The experimental conditions for the ABX test are listed in Table 3.

This test is adopted for impulse and white noise signals. White noise signals are played for 1 second in these experiments.

Table 3. Experimental conditions for the ABX test

| | |
|---|---|
| The number of participants | 7 |
| Gender ratio (men: women) | 5:2 |
| Age | $22.3 \pm 1.6$ |
| The total number of questions each participant answers | 24 |
| Presentation environment | Headphones /Loudspeakers |

### 5.2 MUSHRA

We use a MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA)[15], defined in ITU-R BS.1534-3, to verify how different the participants' impressions are when they listen to signals. In the ABX test, the participants are not allowed to listen to the stimuli multiple times for each answer. However, the MUSHRA, participants can listen to the stimuli repeatedly until they give the final points. In other words, here, we can consider the discrimination results in detail because we can decrease the participants' capricious behavior. Moreover, we can research how different stimuli are subjectively.

As an improvement over MUSHRA, in this study, we have put across a rule wherein the participants will not have to give full points to the stimulus that is felt to be the same as the reference. This means that all the stimuli including the reference can be analyzed statistically.

The points are analyzed using the two-way analysis of variance (two-way ANOVA) to research the effects obtained from the presentation environments and sampling frequencies.

First, the participant adjusts the sound volume of the playing system, quite similar to how it was mentioned in chapter 5.1.

Next, the participant listens to the 192 kHz sampling frequency signal, which is the reference signal, repeatedly till he or she can memorize it.

Finally, the participant gives points for each stimulus from 0 to 100 (the reference being 100). At this time, when the participant gives points, the participant can listen to each stimulus which has different sampling frequency as many times as necessary. After a listening environment, the participant gives points on the other environment. The assignment order of the sliders to the sampling rates is random permuted.

The experimental conditions for MUSHRA are listed in Table 4.

In this study, the test is adopted for impulse signals. The 6 participants in this experiment are the same as ABX test.

Table 4. The experimental conditions for MUSHRA

| | |
|---|---|
| The number of participants | 7 |
| Gender ratio (men: women) | 6:1 |
| Age | $22.1 \pm 0.98$ |
| Presentation environment | Headphones /Loudspeakers |

## 6 Results

### 6.1 ABX Test

In this section, we report the results of the ABX test for impulse and white noise signals.

The answers to these tests are evaluated by the binomial test (significance value of 0.05).

#### 6.1.1 Impulse Signals

Table 5 displays the number of right answers in this test for 28 questions for each combination of sampling frequency.

Table 5. Number of right answers in impulse signals

| Fs | Headphones | Loudspeakers |
|---|---|---|
| 48 kHz vs. 96 kHz | 21 | 24 |
| 48 kHz vs. 192 kHz | 20 | 23 |
| 96 kHz vs. 192 kHz | 22 | 23 |

Table 6 and Table 7 summarizes the results of the ABX test for each impulse signal.

These results show that there are significant differences between each combination of the sampling frequencies for each presentation environment.

Table 6. Result of the ABX test in the impulse signals with a pair of headphones

| Fs | $p$-value (*:$p < 0.05$) |
|---|---|
| 48 kHz vs. 96 kHz | 0.0019* |
| 48 kHz vs. 192 kHz | 0.0063* |
| 96 kHz vs. 192 kHz | 0.0005* |

It is seen that there are no significant differences between the two presentation environments for each combination of sampling frequency. In other words, it is suggested that the listening environments do not affect the discrimination between Hi-Res and non-Hi-Res in the impulse signals.

Table 7. Result of the ABX test in the impulse signals with the loudspeakers

| Fs | $p$-value (*:$p < 0.05$) |
|---|---|
| 48 kHz vs. 96 kHz | < 0.0001* |
| 48 kHz vs. 192 kHz | < 0.0001* |
| 96 kHz vs. 192 kHz | < 0.0001* |

### 6.1.2 White Noise Signals

Table 8 is the number of right answers in this test for 28 questions for each combination of sampling frequency.

Table 8. Number of right answers in white noise signals

| Fs | Headphones | Loudspeakers |
|---|---|---|
| 48 kHz vs. 96 kHz | 19 | 22 |
| 48 kHz vs. 192 kHz | 17 | 21 |
| 96 kHz vs. 192 kHz | 20 | 21 |

Table 9 and Table 10 summarizes the results of the ABX test for each white noise signal.

Table 9. Result of the ABX test in the white noise signals with a pair of headphones

| Fs | $p$-value (*:$p < 0.05$) |
|---|---|
| 48 kHz vs. 96 kHz | 0.0178* |
| 48 kHz vs. 192 kHz | 0.0925 |
| 96 kHz vs. 192 kHz | 0.0063* |

Table 10. Result of the ABX test in the white noise signals with the loudspeakers

| Fs | $p$-value (*:$p < 0.05$) |
|---|---|
| 48 kHz vs. 96 kHz | 0.0005* |
| 48 kHz vs. 192 kHz | 0.0019* |
| 96 kHz vs. 192 kHz | 0.0019* |

These results show that there are significant differences among most of the combinations of the sampling frequencies for each presentation environment.

Moreover, there is a significant difference between the two presentation environments in 48 kHz vs. 192 kHz ($p < 0.05$). In other words, there is a probability that the listening environment affects the discrimination between Hi-Res and non-Hi-Res in the white noise signals. However, it is necessary to correct more answers for a detailed consideration because there are not enough significant differences for most of the combinations. For instance, we can conduct the ABX test for the 48 kHz vs. 192 kHz combination again. This time with more than 28 questions.

## 6.2 MUSHRA

The answers of MUSHRA-based study were evaluated by the two-way ANOVA and the multiple comparison methods (significance value of 0.05 in each test.).

Figure 7 and Figure 8 are the boxplots of each stimulus which have different sampling frequencies in each presentation environment.

Table 11 is the result of the MUSHRA analyzed using two-way ANOVA. There is a significant main effect due to the sampling frequency Fs ($p < 0.05$). In Table 11, dof denotes degree-of-freedom.

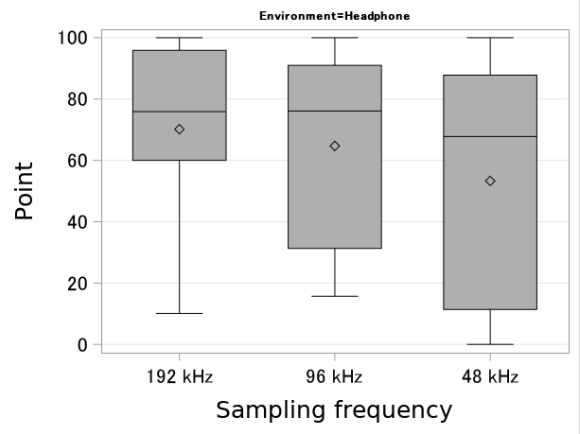


Figure 7. Boxplot of each sampling frequency with a pair of headphones

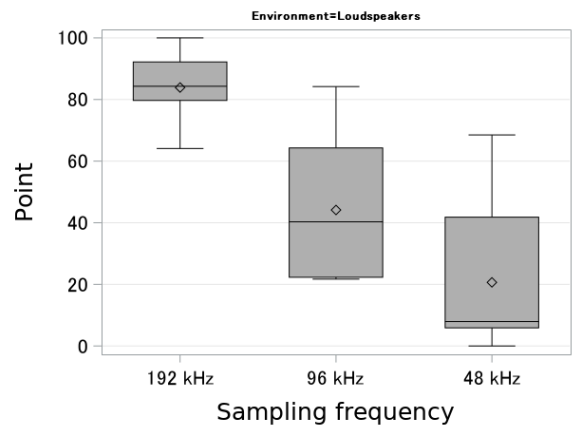


Figure 8. Boxplot of each sampling frequency with the loudspeakers

Table 11. Result of MUSHRA analyzed using two-way ANOVA

| Source | dof | $F$-value | $p$-value (*:$p < 0.05$) |
|---|---|---|---|
| Environment | 1 | 2.36 | 0.1336 |
| Fs | 2 | 7.31 | 0.0022* |
| Environment×Fs | 2 | 2.63 | 0.0862 |

The presentation environment has no effect because the interaction, which occurs in the presentation environment (shown as Environment in Table 9) and due to the sampling frequency, and the main effect of the presentation environment are not acknowledged. In other words, the points of each stimulus are different in the sampling frequency Fs.

Considering the main effect due to the sampling frequency Fs, the points of each stimulus are compared by the Steel-Dwass test.

The results of the Steel-Dwass test is shown in Table 12. From Table 12, it can be noted there is a significant difference ($p < 0.05$) at the 48 kHz vs. 192 kHz comparison point.

Table 12. Result of Steel-Dwass test

| Fs | $t$-value | $p$-value (*:$p < 0.05$) |
|---|---|---|
| 192 kHz vs. 96 kHz | 2.9907 | 0.0868 |
| 192 kHz vs. 48 kHz | 3.9015 | 0.0160* |
| 96 kHz vs. 48 kHz | 2.2424 | 0.2158 |

Going by the results, it is suggested that there is a possibility of a discrimination between Hi-Res and non-Hi-Res audio data.

## 7 Conclusion

In this study, we considered the discrimination between Hi-Res and non-Hi-Res without music in two subjective evaluations based on five problems: music types, participants' musical background, body responses and emotions, masters, and recording characteristics adopted.

We selected and quantized various signals, except for music signals, to satisfy the specifications of LPCM.

The ABX test is also conducted for three impulse signals and three white noise signals (with 48, 96, and 192 kHz sampling frequencies.).

The results of the ABX test suggest that people can discriminate between Hi-Res and non-Hi-Res impulse and white noise signals when using either loudspeakers or a pair of headphones.

In addition to the ABX test, we also conducted a MUSHRA for each impulse signal, and the main effect due to the sampling frequency is shown in the result. Moreover, a significant difference is highlighted in the 48 kHz vs. 192 kHz comparison through the Steel-Dwass test.

From these two experimental results, it can be suggested that people can conveniently discriminate between Hi-Res and non-Hi-Res without music.

We will consider this discrimination between Hi-Res and non-Hi-Res further in the future to gain more general understanding of the discrimination between Hi-Res and non-Hi-Res audio data.

## References


[1] V. R. MELCHIOR, "High Resolution Audio: A History and Perspective", *J. Audio Eng. Soc.,* Vol. 67, No. 5, pp. 246-157, (2019, May).

[2] T. Nishiguchi, "A Study on Human Hearing of High Resolution Audio", The University of Electro-Communications, Ph.D. thesis (2009, in Japanese).

[3] M. Mizumachi, R. Yamamoto, and K. Niyada, "Discussion on subjective characteristics of high resolution audio", AES 142nd Convention, e-Brief No.315 (2017).

[4] A. Suguro and M. Miura, "Quality discrimination on high-resolution audio with difference of quantization accuracy by sound-image localization", Proc. AES Conference on Spatial Reproduction, e-Brief No.74 (2018).

[5] Japan Audio Society, "Definition of Hi-Res Audio (Announced on June 12th 2014)", Japan Audio Society (2018), [online] <https://www.jas-audio.or.jp/english/hi-res-logo-en>, referred on May 21, 2019.

[6] Japan Electronics and Information Technology Industries Association, "The announcement for calling "High-Resolution Audio"", JEITA (2014, in Japanese), [online] <https://home.jeita.or.jp/page_file/20140328095728_rhsiN0Pz8x.pdf>, referred on May 21, 2019.

[7] Record Industry Association of America, "High Resolution Audio Initiative Gets Major Boost with New "Hi-Res MUSIC" Logo and Branding Materials for Digital Retailers", RIAA (2015), [online] <https://www.riaa.com/high-resolution-audio-initiative-gets-major-boost-with-new-hi-res-music-logo-and-branding-materials-for-digital-retailers/>, referred on May 21, 2019.

[8] S. Yasuda, "A Psychological Study of Strong Experiences in Listening to Music based on a Relationship among Strong Experiences, Physical Responses and Emotions from Listener's musical background", Proc. The 76th Annual Convention of the Japanese Psychological Association, 3PMA13 (2012, in Japanese).

[9] ONKYO, "e-onkyo music" , ONKYO (2018, in Japanese), [online]<https://www.e-onkyo.com/music/album/wnr190295511838/ >, referred on May 22, 2019.

[10] T. Nishiguchi and K. Hamasaki, "Differences of Hearing Impressions among Several High Sampling Digital Recording Formats", AES 118th Convention, Paper No. 6469 (2005).

[11] S. P. LIPSHITZ and J. VANDERJKOOY, "Pulse-Code Modulation – An Overview", *J. Audio Eng. Soc.*, Vol. 52, No. 3 (2004, March).

[12] S. Kanai, "Signal Processing Vol. 2" , Hokkaido University (2018, in Japanese), [online]<http://sdmwww.ssi.ist.hokudai.ac.jp/lecture/signal/presen2.pdf>, referred on May 22, 2019.

[13] W. Kaster, "Oversampling Interpolating DACs", Analog Devices Tutorial, MT-017 (2015).

[14] W. A. Munson and M. B. Gardner, "Standardizing Auditory Tests", *The Journal of the Acoustical Society of America*, Vol. 22, pp. 675 (1950).

[15] International Telecommunication Union, *Recommendation ITU-R BS.1534-3 (10/2015): Method for the subjective assessment of intermediate quality level of audio systems* (2015).