ICIC Express Letters
Part B: Applications
Volume **11**, Number **7**, July **2020**

pp. **655–660**

DIFFERENTIATING BETWEEN HIGH-RESOLUTION AUDIO WITH IMPULSE SIGNALS

Yuki Fukuda and Shunsuke Ishimitsu

Graduate School of Information Sciences
Hiroshima City University
3-4-1, Ozuka-Higashi, Asaminami-Ku, Hiroshima-Shi, Hiroshima 731-3194, Japan
y-fukuda@hfce.info.hiroshima-cu.ac.jp; ishimitu@hiroshima-cu.ac.jp

Received January 2020; accepted April 2020

Abstract. *In recent years, high-resolution (Hi-Res) audio formats with higher sampling frequencies and quantization bits than the compact disc (CD) format have received considerable attention in the audio industry. Accordingly, many studies have been conducted for the identification and discernment between Hi-Res and non-Hi-Res (e.g., CD) audio. However, most studies use music sources as the test signals rather than non-music sources (e.g., colored noise). Furthermore, there are different considerations about this possibility of discernment. In other words, some studies consider that people can differentiate between these audios, whereas others deny or do not clarify this. In this study, we considered the identification and discernment between Hi-Res and non-Hi-Res audio without a music source by using AB test and multiple stimuli with hidden reference and anchor tests. Consequently, the probability of discernment was suggested.*
Keywords: Audio signal differentiation, High-resolution audio, Impulse signal

1. **Introduction.** In recent years, high-resolution (Hi-Res) audio formats – with higher sampling frequencies and quantization bits than the compact disc (CD) format – have become increasingly popular on the Internet. Consequently, many studies have examined whether people can differentiate between Hi-Res audio and other formats.

Nishiguchi [1] recorded original music sources with frequencies over 21 kHz and compared them with normal sources. They reported that some people could identify and differentiate between the stimuli sources. However, they did not obtain any significant result for Hi-Res audio formats with four music college students, a violinist, and a recording engineer as their subjects.

Yamamoto et al. [2] compared the CD format with a Hi-Res format music source with 27 participants. The participants could differentiate between these sources with an accuracy of 57%. In addition, nine participants who could differentiate between these sources – including seven people with a musical background (e.g., playing musical instruments, and regularly listening to Hi-Res music) – could also differentiate between higher and lower quantization bit numbers more accurately than others.

Suguro and Miura [3] presented two types of musical sound data with different quantization bit numbers, and the same sampling frequency was simultaneously played in each ear. They reported that sound image localization occurred on the side with a larger quantization bit number.

These previous studies indicate three major problems when using music data to identify Hi-Res audio, as listed below.

- If researchers use restrictive types of music data, there is a possibility that the results will differ when another type of music data is used.

DOI: 10.24507/icicelb.11.07.655

- The results may depend on the participants' listening backgrounds. In other words, the researcher has to carefully collect the audio data.
- A previous report highlights a significant difference between the physical and emotional responses of participants who play musical instruments and those who do not when listening to music [4]. Hence, we consider that a musical background affects the subjective evaluation when listening to music.

Our study not only draws on previous studies in using music data for experiments, but also attempts to identify and differentiate between Hi-Res and non-Hi-Res audio without a music source.

2. **Signal Selection.** For the experiments, we selected impulse signals without a music source based on the following conditions.

- Signals have a flat frequency response.
- Since stimuli are not music signals, the musical background of participants is irrelevant to the subjective evaluation.
- The loads on the participants are exceedingly small because of the short duration.

We created three impulse signals as shown in Figure 1.

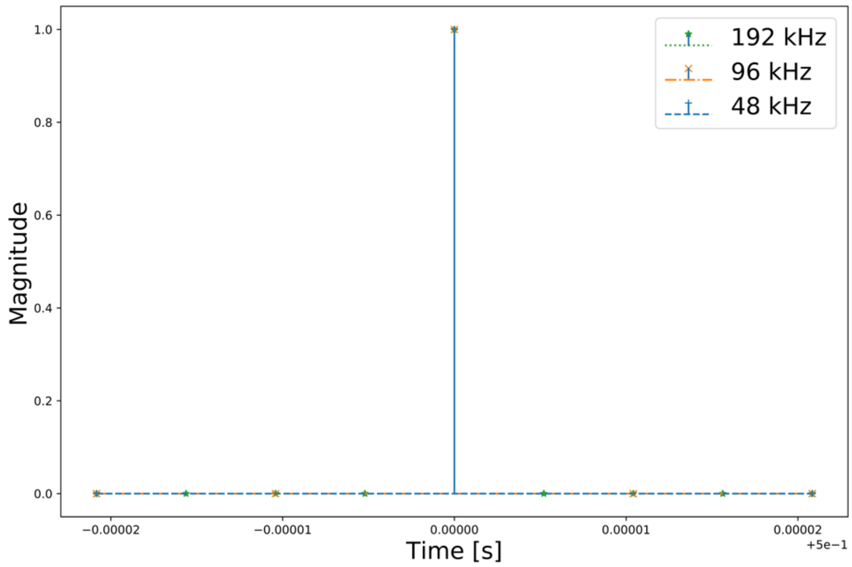

Figure 1. Impulse signals for each sampling frequency

In this study, we define the Hi-Res impulses with the sampling frequencies of 96 kHz and 192 kHz according to the Recording Industry Association of America (RIAA) guideline [5].

The impulse response of the finite impulse response (FIR) filter to prevent aliasing noise [6] in the playing system's digital-to-analog converter (DAC) is shown in Figure 2. The actual impulse signal is similar to the waveform in Figure 2.

3. **Experiments and Results.** To consider the possibility of differentiating between the Hi-Res and non-Hi-Res signals without music data, we conducted two experiments with the impulse signals. The experimental environment is shown in Table 1.

The schematic of the experiment system is shown in Figure 3. The participant controls the computer while hearing the stimuli through the headphones and loudspeakers.

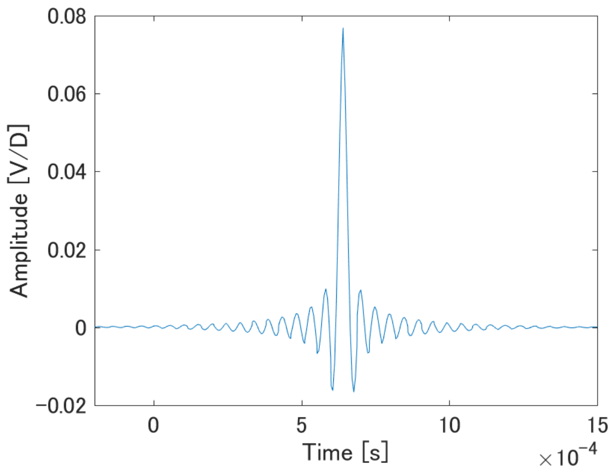

FIGURE 2. Impulse response of the playing system's DAC (averaged 100 times)

TABLE 1. Experiment environment

Experiment room	Anechoic chamber
USB-DAC	FOSTEX HP-A4BL
Loudspeakers	ECLIPSE TD-M1
Headphones	SONY MDR-Z7
Software	Self-made (MATLAB)

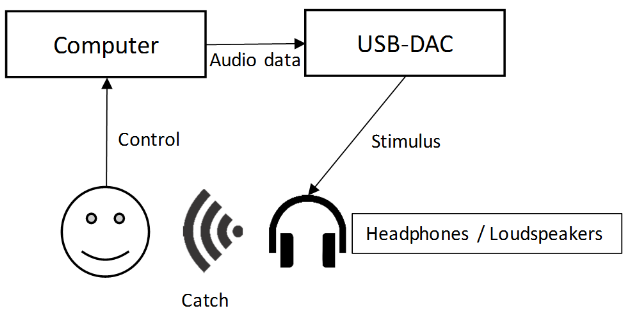

FIGURE 3. Sound playing system

3.1. **AB test.** First, we adopted the AB test, in which the participants judge if the two played impulses are the same [7], to study whether they can differentiate between each of the impulse signals.

First, the volume of the playing system was controlled by the participant to decrease his/her load. Then, the participants answered whether the two impulses, played simultaneously, were the same.

Each participant answered 96 questions in a presentation environment, including a half ratio of the same frequency combination. Furthermore, the tests were conducted in two presentation environments: with loudspeakers and with a pair of headphones. Therefore, the total number of questions answered by each participant was 192. After the test, the participants rested for 5 min. The graphical user interface (GUI) of this test is shown in Figure 4, and the experimental conditions are summarized in Table 2.

Table 3 summarizes the results of the AB test. The answers to this test were evaluated with a binomial test (significance level $p = 0.05$). These results show that there are significant differences between all combinations of the sampling frequencies.

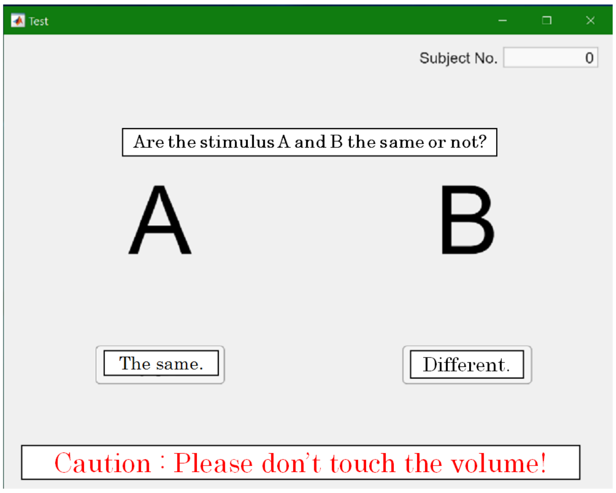

FIGURE 4. The GUI for the double-blinded test

TABLE 2. Experimental conditions for the AB test

The number of participants	9
Gender ratio	Men: 5; Women: 4
Age	22.4 ± 1.64
Presentation environment	Headphones/Loudspeakers
Total number of attempts per participant	192

TABLE 3. Results of the AB test

Combination	p-value	Significant difference ($p < 0.05$)
48 kHz – 96 kHz	< 0.0001	○
48 kHz – 192 kHz	< 0.0001	○
96 kHz – 192 kHz	0.0432	○

Moreover, there is a significant difference in the differentiation ratio between the two presentation environments for the 96 kHz – 192 kHz combination. In other words, there is a possibility of differentiating between the loudspeakers and the pair of headphones. However, the number of samples in the 96 kHz – 192 kHz combination (statistical power $1 - \beta < 0.8$) is not adequate. Therefore, it is necessary to collect more samples for a more accurate consideration.

3.2. **MUSHRA.** Based on the results of AB test, we employed a multiple stimuli with hidden reference and anchor (MUSHRA) [8] test – defined in ITU-R BS.1534-3 – to verify how different the participants' impressions were when listening to impulses. However, as an improvement over MUSHRA test, we ruled that the participants do not have to give full points to stimuli that are the same as the reference to statistically evaluate each point, including the reference. Then, the points were analyzed.

First, as in the AB test (Section 3.1), the participant adjusts the volume of the system. Next, the participant listens to the 192 kHz impulse – the reference signal – repeatedly till he/she can memorize it. Then, the participant gives points between 0-100 to the three stimuli; the reference is 100. While the participant is giving points, he/she can listen to each stimulus as many times as necessary.

The experimental conditions for MUSHRA test are listed in Table 4 and its GUI is shown in Figure 5.

Table 4. Experimental conditions for MUSHRA test

The number of participants	10
Gender ratio	Men: 5; Women: 5
Age	22.5 ± 1.5
Presentation environment	Headphones/Loudspeakers

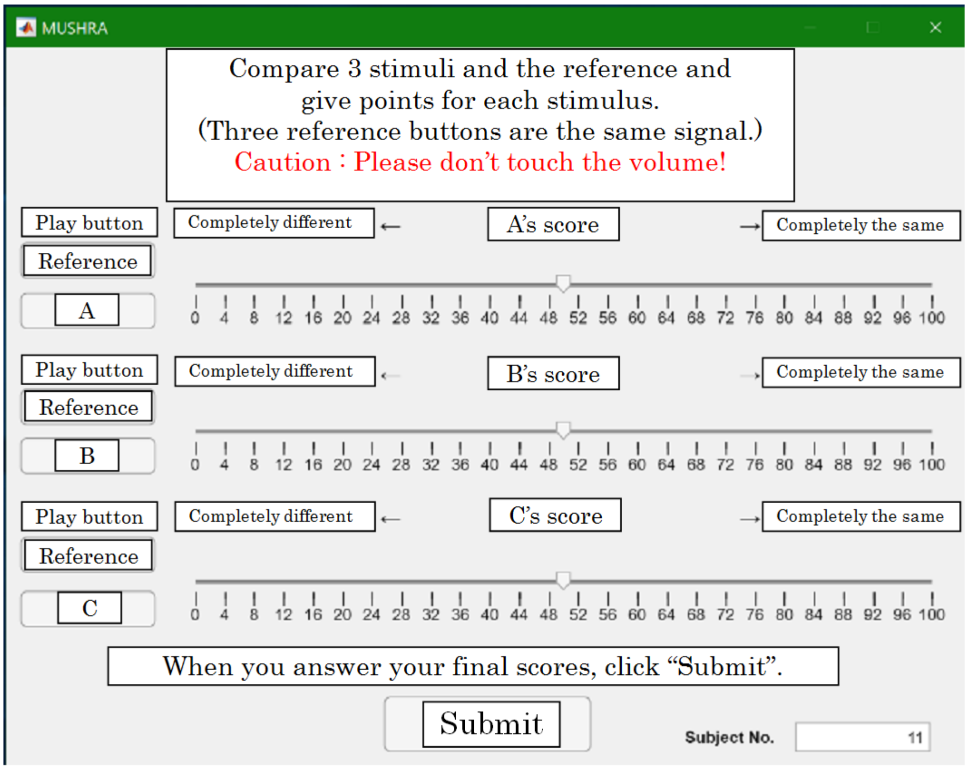

Figure 5. The GUI for the MUSHRA

Table 5. Results of MUSHRA test (2-way ANOVA)

Factor	Degree of freedom	F-value	p-value	Main effect ($p < 0.05$)
Presentation environment and sampling frequency	2	0.02	0.9675	×
Sampling frequency	2	82.92	< 0.0001	◯
Presentation environment	1	3.13	0.0827	×

Table 5 shows the results of MUSHRA test using two-way ANOVA. The presentation environment has no significant effect or interaction with sampling frequency. In other words, the points of each stimulus are different for a sampling frequency.

Considering the main effect of the sampling frequency, multiple comparisons of the points for each stimulus are carried out (Steel-Dwass test). Figure 6 shows the distribution of each impulse. The results of multiple comparisons are shown in Table 6. This table shows significant differences ($p < 0.05$) between all combinations of sampling frequencies.

4. **Conclusions.** This study considered two statistical methods to examine the identification and discernment of Hi-Res and non-Hi-Res audio without a music source.

First, we conducted the AB test for three impulses (48 kHz, 96 kHz, and 192 kHz sampling frequencies). Consequently, we suggest the possibility that people can differentiate between Hi-Res and non-Hi-Res audio, both in loudspeakers and headphones. However,

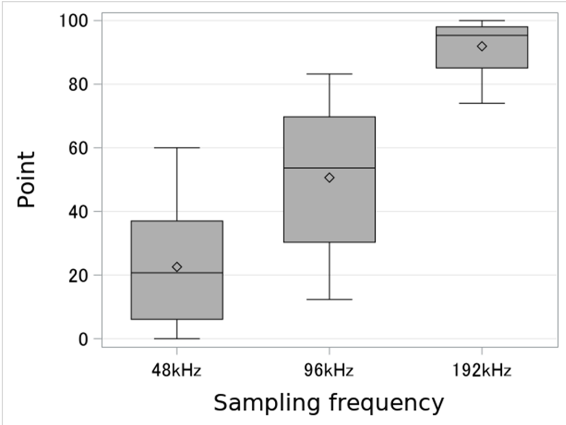

Figure 6. Distribution of points for each sampling frequency

Table 6. Results of multiple comparisons (Steel-Dwass test)

Combination	p-value	Significant difference ($p < 0.05$)
48 kHz – 96 kHz	0.0011	◯
48 kHz – 192 kHz	< 0.0001	◯
96 kHz – 192 kHz	< 0.0001	◯

we must conduct more accurate tests (e.g., ABX test [7]) for a more detailed understanding.

Second, we conducted MUSHRA test for each impulse and the significant differences were shown for all combinations of sampling frequencies for multiple comparisons.

The results of these two tests suggest that people can appreciably differentiate between Hi-Res and non-Hi-Res audio without a music source.

In the future, we will further consider the differentiation between Hi-Res and non-Hi-Res audio. For instance, apart from impulses, we will also conduct similar tests with noise signals. As the amplitudes of noise signals vary with time, we believe that people can identify and discern Hi-Res and non-Hi-Res audio using two factors, sampling frequency and quantization bits.

REFERENCES

[1] T. Nishiguchi, *A Study on Human Hearing of High-Resolution Audio*, Ph.D. Thesis, The University of Electro-Communications, 2009.

[2] R. Yamamoto, N. Kanetada and M. Mizumachi, Evaluation of sound quality of high-resolution audio, *The Japanese Journal of the Institute of Industrial Applications Engineers*, vol.1, no.2, pp.52-57, 2013.

[3] A. Suguro and M. Miura, A study of the possibility of discriminating sound sources which have different quantization accuracy because of interaural localization, *The Technical Committee on Musical Acoustics in the Acoustical Society of Japan on May, 2018*, MA2018-05, pp.1-6, 2018.

[4] 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 experiences, *The 76th Annual Convention of the Japanese Psychological Association*, 2012.

[5] RIAA, *High-Resolution Audio Initiative Gets Major Boost with New "Hi-Res MUSIC" Logo and Branding Materials for Digital Retailers*, https://www.riaa.com/high-resolution-audio-initiative-gets-major-boost-with-new-hi-res-music-logo-and-branding-materials-for-digital-retailers/, 2015.

[6] W. Kaster, *Oversampling Interpolating DACs*, https://www.analog.com/media/cn/training-seminars/tutorials/MT-017.pdf, 2015.

[7] R. E. Pastore, C. J. Friedman and K. J. Baffuto, Comparative evaluation of the AX and two ABX procedures, *The Journal of the Acoustical Society of America*, vol.60, 1976.

[8] ITU, *Recommendation ITU-R BS.1534-3 – Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems*, 2015.