Evaluation of Sound Quality of High Resolution Audio

Naoto Kanetada^{a,*}, Ryuta Yamamoto^b, Mitsunori Mizumachi

^aKyushu Institute of Technology, 1-1 Sensui-cho, Tobata-ku, Kitakyushu 804-8550, Japan

^bDigifusion Japan Co., Ltd 1-1-68 Futabanosato Higashi-ku Hiroshima, 7320057 Japan

*Corresponding Author: n349406n@tobata.isc.kyutech.ac.jp

Abstract

High resolution audio (HRA), which is recorded in the digital audio format with high sound quality, appears on the audio market. HRA has the quality equal to or better than the standard compact disc (CD), and is distributed as the super audio CD (SACD), DVD-audio, Blue-ray audio, and a data file through the internet. In this paper, sound quality of HRA is investigated in the view point of auditory perception. Perceptual characteristics of HRA have been examined by listening tests as compared with the standard audio CD and the compressed MPEG audio layer-3 (MP3) qualities. The listening tests were carried out by the method of paired comparison, and the participants were asked to discriminate among HRA, CD, and MP3. Jazz music, which was sampled at 192 kHz with 24 bits resolution, was selected out as the original HRA. Both CD and MP3 quality audio files were prepared from the original HRA file by the authors. It is found that the participants prefer the audio format with richer information, that is, HRA is most desirable and CD is better than MP3. The participants were also required to fill in the questionnaire on subjective impression for each audio format. Subjective characteristics of HRA include much presence, high spatial clarity, and wide spaciousness. In addition, perceptual discrimination in quantization distortion is also focused on in between 16 bits and 24 bits resolution. As the result of the listening test with 28 participants, the difference of quantization resolution could be discriminated with the accuracy of 60.3 %. Concerning the perceptual discrimination of HRA, it is suggested that quantization resolution is an important factor as well as wide frequency range.

Keywords: High resolution audio, CD, MP3, subjective quality evaluation, quantization resolution

1. Introduction

In recent years, high resolution audio (HRA), which is sampled at 96 kHz or 192 kHz with 24 bits accuracy, is becoming popular in the audio market. HRA is commercially distributed as the lossless encoded file via the internet, and is also available in the Blu-ray audio disc. Compact disc (CD) and lossy compression such as MPEG Audio Layer-3 (MP3) are the current major audio formats as the storage medium and the data file, respectively. As a memory capacity increases and a wide communication network spreads out, HRA must be increasingly popular. In this paper, perceptual audio quality is investigated by listening tests for HRA, CD, and MP3. This paper aims to define the perceptual characteristics of each audio format. Nishiguchi reported that two listeners out of 13 participants could significantly discriminate the difference between HRA and CD in the formal listening test. It suggests that the high frequency component above 20 kHz, which is the upper limit of the human audible range, influences the perception of sound. Oohashi *et al.* mentioned that an analogue LP record, which includes the high frequency components above 40 kHz, is superior in perceptual sound quality to CD as the result of the paired comparison test. Oohashi *et al.* hypothesized that an analogue LP record has an advantage over CD due to the hypersonic effect. Recently, audio fans and critics also have an interest in perceptual difference among HRA, CD, and MP3. It is necessary to quantitatively evaluate each audio format by listening tests. In this study, HRA files with 192 kHz/24 bits accuracy, which are available on the market, are prepared as the reference. Then, CD equivalent and MP3 with high and low bit rates are carefully produced in a computer. The CD equivalent is sampled at 48 kHz with 16 bits accuracy due to reduce the error in downsampling,

although the sampling frequency of the standard audio CD is standardized at 44.1 kHz. In this paper, the CD equivalent is regarded as CD, because the CD equivalent is band-limited by the low-pass filter with the cut-off frequency of 20 kHz. Concerning the lossy compressed MP3 files, HRA is converted into MP3(H) and MP3(L) with the bit rates of 320 kbps and 128 kbps, respectively.

In addition to comprehensive sound quality, perceptual discrimination in quantization distortion is also investigated in between 16 bits and 24 bits resolution.

2. Evaluation of perceptual sound quality for audio format

2.1 Preparation of audio files

Perceptual sound quality is compared among HRA, CD, MP3(H), and MP3(L). It is reported that the commercial HRA files are frequently generated from the CD quality data, which contains no high frequency components originally, and some of them are intentionally remixed in order to enhance the perceptual difference in between the generated HRA and the original CD. In the selection of the reference in HRA, a variety of HRA sources were prepared, and the frequency analysis was carried out in advance. Finally, three kinds of HRA sources, which are commercially available, are carefully selected out in the genre of classical music, jazz music, and female vocal music. In this study, HRA, CD and MP3 sources were prepared as follows.

2.1.1 HRA

All HRA sources have been sampled at 192kHz with 24 bits accuracy.

2.1.2 CD

In the preparation of CD source, an anti-aliasing low-pass filter has been designed with cut-off frequency of 20 kHz and -120 dB/octave slope. It has been implemented as the 20th-order Butterworth filter using MATLAB. As mentioned above, sampling frequency is set at 48 kHz, that is, one fourth of the sampling frequency of HRA. Quantization resolution has been converted into 16 bits using MATLAB.

2.1.3 MP3

MP3 sources have been generated using LAME encoder 3.99.5 from the CD source prepared in 2.1.2. MP3(H) and MP3(L), of which bit rates are 320 kbps and 128 kbps, are high and common quality formats, respectively.

2.1.4 Preliminary experiment for selecting suitable sound source

Informal listening test was carried out to verify the music sources prepared in four kinds of formats. As the result, the jazz music was the most suitable in discriminating the audio formats, because it was easy for the listeners to discriminate the minute perceptual difference relying on the crisp sound caused by percussion instruments. Therefore, the jazz music has been chosen as the sound stimuli in the listening tests. Figure 1 shows the long-term amplitude spectra for HRA, CD, and MP3(L).

Fig.1. Long-term averaged amplitude spectra for HRA, CD, and MP3(L).

2.2 Experiment setup

Fig.2. shows the composition of the audio equipment. Audio files in the network-attached storage (NAS) were presented to listeners through the loudspeakers. In the listening test, each stimulus was manually given to the subject under the blind condition. Two sets of the loudspeakers and the listening area were arranged at the apex of the equilateral triangle, of which the length of the side was 2.5 m, as shown in Fig. 3. The super tweeters with the narrow directional characteristics were turned toward the center of listening area. Four subjects sat together in the listening area for efficient operation of the listening test. Each subject adjusted the height of the chair to match the

ear level with the height of the super tweeter of 1.2 m. Frequency-dependent sound pressure level at the listening area had sufficient characteristics for the listening test. A-weighted sound pressure level was in between 75 dB and 80 dB at the peak of each stimulus. It is reported that the perception of HRA is often affected by the non-linear distortion caused by audio equipment, especially by loudspeakers. In this experiment, it is confirmed that the effect of the non-linear distortion could be negligible.

Fig.2. Experimental setup for listening test. Sound file stored in NAS was D/A converted and amplified prior to branch into loudspeakers.

Fig.3. Geometrical relationship among two sets of loudspeakers and listening area, where four listeners sat together.

2.3 Procedure

2.3.1 Subject

27 subjects, who were 11 females and 16 males, participated in the listening test. Their ages spread from teens to seventies, and their interests to audio and music varied from uninteresting listeners to active musicians.

2.3.2 Questionnaire

The subjects were asked to fill in questionnaire on age, sex, musical experience, a favorite genre, listening environment, a preferred audio format, and so on.

Fig.4. Sound pressure level at left and right ear positions in listening area in blue and red lines.

2.3.3 Method

Paired comparison was carried out in sound quality between two successive stimuli. In two-alternative forced-choice, the subjects selected the better stimulus for each pair.

2.3.4 Adjectives describing sound quality

The subjects were also asked to describe the reason of the judgment. It is hard for the unfamiliar subjects to describe the slight difference in sound quality. Then, the subjects were allowed to refer the adjectives describing sound quality. The list of the adjectives is presented in Table 1. Concerning the discrimination in frequency characteristics, quantization distortion, and compressed audio formats, the adjectives in Table 1 were prepared in reference to the previous works by Oohashi *et al.* and Kuriyagawa *et al.* Oohashi *et al.* focused on perceptual discrimination between LP records and CDs, and Kuriyagawa *et al.* summarized the adjectives on general sound quality systematically.

2.3.5 Duration of stimuli

Duration of each stimulus was set at 120 s, and the rest of 30 s was inserted in between two successive stimuli. In total, it took 270 s for each paired-stimuli set. There was the rest

of more than 60 s in between the sets. In general, duration of audio stimulus is set in the range between 10 s and 30 s. Oohashi *et al.* mention that the hypersonic effect is caused with temporal delay, when we listen to sounds containing rich high frequency components. The hypersonic effect is considered in this experiment.

Table.1. Pairs of adjectives for sound quality assessment

ID	Pairs of adjectives
1	soft – hard
2	rich – poor
3	pleasant – pungent
4	sleek – dry
5	deep – shallow
6	spacious – narrow
7	atmospheric – monotonous
8	expressive – inexpressive
9	natural – unnatural
10	profound – light
11	smooth – rough
12	wide treble – narrow treble
13	wide bass – narrow bass
14	clear – dull
15	brisk – blurred
16	reverberant – anechoic
17	distinct – indistinct
18	powerful – weakly
19	favorite – disliked

2.3.6 Session

In the paired comparison, two stimuli were selected out among HRA, CD, MP3(H), and MP3(L). Each session consisted of 12 patterns of the permutation of the paired-stimuli sets. Each subject joined two sessions. In other words, each subject evaluated each pair twice.

2.3.7 Duration of experiment

In consideration of the subject's fatigue, experiment time for one set was set for 30minutes. And interval for 15minutes was taken. The total experiment time of the day was set 2~2.5 hours.

2.3.7 Training

In the beginning of the experiment, the subjects participated in the training session, where the point of discrimination was instructed and the subjects could communicate with each other for 30 minutes.

2.4 Results

Experimental results, which are averaged over the 27 subjects, are summarized in Fig. 5. In Fig. 5, both pair-dependent results and the overall results are given with the rate when the subjects answer better to the stimulus with richer information. For example, in the pair of CD with MP3(L), 61.1 % of the subjects preferred CD to MP3(L).

Fig.5. Results of paired comparison in preference of sound quality.

2.5 Discussions

The adjectives given by the subjects as the ground of discrimination are summarized in Table 2 for both the audio formats with richer and poorer information. In Table 2, “○” and “×” indicate superior and inferior evaluation against the alternative of the pair. HRA is evaluated with wide spaciousness, much depth, and much presence compared with CD. CD has advantages in less noisiness, less distortion, and separation of sounds over MP3(H). There is perceptual difference in clearness and distortion between MP3(H) and MP3(L). It is found that the audio formats with rich information yield the subjective impressions of much presence, clear, and wide spaciousness. On the other hand, the formats with poor information are

evaluated with rough and shallow. In Fig. 5, the majority of the subjects prefer the audio formats with richer information in all pairs. However, even the richest HRA is negatively evaluated with annoyance and noisiness as shown in Table 2. It is supposed that the strange feeling caused by HRA might be close to the noisiness for MP3. Further investigation should be done under other experimental conditions.

Table.2. Subjective impression given by the subjects.

	HRA		CD	
HRA-CD	○	high presence stereophonic wide spaciousness natural clear	×	unsatisfactory less reverberant light rough indistinct
	×	indistinct noisy loud	○	coherent clear sharp pure

	CD		MP3(H)	
CD-MP3(H)	○	clear brisk wide spaciousness pleasant natural stereophonic	×	noisy narrow spaciousness shallow indistinct
	×	hard narrow range	○	wide treble distinct preference

	MP3(H)		MP3(L)	
MP3(H)-MP3(L)	○	clear wide spaciousness high presence deep brisk pleasant distinct	×	distortion narrow spaciousness rough shallow poor separation
	×	thin mild rough hard	○	coherent mild wide spaciousness

3. Evaluation test for quantization accuracy

Perceptual discrimination in quantization resolution is investigated in between 24 bits and 16 bits. HRA is physically different from CD concerning both frequency range and quantization resolution. It is necessary to investigate how the quantization resolution influences sound quality perceptually. In this paper, the possibility of the perceptual discrimination between 24 bits and 16 bits accuracy is examined by another listening test, when only the quantization resolution differs with the same frequency range.

3.1 Procedure

3.1.1 Stimuli

The same sound source is used, that is, the jazz music in Section 2. The reference HRA was band-limited using the steep low-pass filter, of which cut-off frequency is 20 kHz, and downsampled into 48 kHz using MATLAB. The, two versions were prepared with 24 bits and 16 bits accuracy.

3.1.2 Subjects

28 subjects including the 24 subjects in Section 2 participated in the listening test.

3.1.3 Duration of stimuli

Paired comparison was carried out in perceptual discrimination of the quantization distortion. Duration of stimulus was set at 30 s with the rest of 3 s in between the stimuli, because it is not necessary to care the hypersonic effect.

3.1.4 Session

There were only two stimuli with 24 bits and 16 bits accuracy. Each session consisted of four paired-stimuli patterns with permutation.

3.1.4 Task

The subjects answered the same or not for each pair

3.2 Result and discussions

Each subject joined four sessions, and then 112 evaluations, which were 28 subjects by 4 sessions, were obtained in total. Overall correct rate was 60.3 %, and the subjects could significantly discriminate the difference of quantization resolution according to the chi-square test. For each subject, there is no correlation in the correct rate between the experiments in Section 2 and Section 3. It suggests that the clues for discrimination must be different in between the audio formats and quantization resolution.

4. Conclusions

In this paper, perceptual evaluation of HRA has been investigated compared with the standard CD and the lossy

compressed MP3. In the discrimination of the audio formats, HRA was superior to CD, where 57.4 % of the subjects preferred HRA to CD. CD had an advantage over MP3(L), and MP3(H) was better than MP3(L). In short, the subjects preferred the audio format with richer information in paired comparison. Subjective impressions for HRA, CD, and MP3, include much presence and separation of sound source, clearness and low distortion, and distortion and less presence, respectively. Concerning the perceptual evaluation of quantization distortion, 28 subjects have discriminate the difference between 24 bits and 16 bits accuracy in the frequency range up to 20 kHz, and 60.3 % of the subjects could correctly discriminate the quantization resolution. Concerning the perceptual discrimination of HRA, it is suggested that the quantization resolution is also an important factor as well as the sampling frequency. In future, statistical significance in discriminating HRA should be examined with a wide variety of music sources.

the Acoustical Society of Japan, 34(9), pp. 501-509, 1977 (in Japanese).

Acknowledgments

This study, which was supported by program under Hiroshima prefecture's subsidies system for the technology development promotion of the next-generation automotive, was conducted as the part of the project, "Development of car audio systems for high resolution audio with online music distribution," by Ooasa Electronics Co.,Ltd and DigiFusion Japan Co.,Ltd. The authors thank Katsuyuki Niyada of Hiroshima Industrial Promotion Organization, for his kind support, to Fujio Iwaki and Kazushi Uemura for their kind advice, to Kenji Suehiro for his assistance in the experiences. The authors also thank the use of experimental equipments in Hiroshima Innovation Center for Biomedical Engineering and Advanced Medicine.

References

- (1) T. Nishiguchi, "A Study on Human Hearing of High Resolution Audio," Ph.D. thesis, 2009 (in Japanese).
- (2) T. Oohashi, E. Nishina, and Y. Huwamoto, "A Study on the difference of the sound quality between LP
- (3) S. Kiryu and K. Ashihara, "Estimation of quality control in high-sampling audio softwares," Technical Report of IEICE, EA99-37, pp. 71-73, 1999 (in Japanese).
- (4) S. Kiryu, K. Ashihara, S. Sato, and S. Yoshikawa, "Next generation high-definition audio and measurement technologies," The Journal of the Acoustical Society of Japan, 56(9), pp. 653-656, 2000 (in Japanese).
- (5) M. Kuriyagawa, H. Yahiro, and S. Kashiwagi, "Seven Attitude in Tone Quality Evaluation," The Journal of