Audio Engineering Society

Convention Paper 8101

Presented at the 128th Convention
2010 May 22–25 London, UK

The papers at this Convention have been selected on the basis of a submitted abstract and extended précis that have been peer reviewed by at least two qualified anonymous reviewers. This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Sampling rate discrimination: 44.1 kHz vs. 88.2 kHz

Amandine Pras[1], Catherine Guastavino[1]

[1] Centre for Interdisciplinary Research in Music Media and Technology,
Multimodal Interaction Laboratory, McGill University, Montréal, Québec, H3A 1EA, Canada
amandine.pras@mcgill.ca, catherine.guastavino@mcgill.ca

ABSTRACT

It is currently common practice for sound engineers to record digital music using high-resolution formats, and then down sample the files to 44.1kHz for commercial release. This study aims at investigating whether listeners can perceive differences between musical files recorded at 44.1kHz and 88.2kHz with the same analog chain and type of AD-converter. Sixteen expert listeners were asked to compare 3 versions (44.1kHz, 88.2kHz and the 88.2kHz version down-sampled to 44.1kHz) of 5 musical excerpts in a blind ABX task. Overall, participants were able to discriminate between files recorded at 88.2kHz and their 44.1kHz down-sampled version. Furthermore, for the orchestral excerpt, they were able to discriminate between files recorded at 88.2kHz and files recorded at 44.1kHz.

1. INTRODUCTION

In 1982, Sony and Philips defined the CD standard with a sample rate of 44.1 kHz. Since then, 'high-resolution' formats, defined by Rumsey [9] as digital formats with a sample rate beyond the CD standard of 44.1 kHz, have been introduced in the market without commercial success. Thus, sound engineers tend to record digital music at very high sample rates and then down-sample the files to 44.1 kHz for commercial release. However, the down-sampling process introduces measurable artifacts [3]. Therefore, it is necessary to question why sound engineers use high sample rates for recording when the final delivery format is in 44.1 kHz.

Sample rate refers to the number of samples per second extracted from the original signal. In order to reconstruct a signal, the sample rate must be at least twice the frequency of the signal being sampled [8]. According to this theorem and limits of human hearing commonly known to be 20 kHz, the CD standard of 44.1 kHz is high enough to encode the audible content of a signal. However, several theories support the practice of recording at very high sample rates.

First, Stuart [10] claimed that some people could hear above 20 kHz, possibly up to 25 kHz. Indeed, in a study conducted by Nishiguchi & Hamasaki [6], one out of 36 participants significantly detected differences between sound with and without frequencies above 20 kHz.

A second theory relates to technological limitations of analog-to-digital converters. To avoid spectral aliasing of frequencies that are too high to be encoded, the first step of analog-to-digital conversion is low-pass filtering. The slope of this anti-aliasing filter could affect the high frequency content of the signal, which may introduce audible artifacts [10].

A third theory refers to the temporal resolution implied by the sample rate [11]. While listening with two ears, humans can discriminate time differences of 2 µs or less [7]. Percussionists can play sounds with transients lasting only a few µs; hall reverberation may include reflections only a few µs apart. The temporal difference between two samples in 44.1 kHz is 22.7 µs, *i.e.* may not be precise enough.

Few studies have been conducted to investigate the perceptual differences between high-resolution and 44.1 kHz or 48 kHz. Meyer & Moran [5] compared Super Audio CD playback and a loop through a digital device in 44.1 kHz in an ABX comparison test but failed to observe significant differences. Yoshikawa et al [12] found that three participants out of 11 could discriminate between musical excerpts in 96 kHz and their down-sampled version in 48 kHz. However, these audible differences could be attributed to the down sampling algorithm and not to the difference of sample rate. Laugier [2] observed a better spatial reproduction and high frequency restitution while listening to high-resolution files recorded at 192 kHz / 24 bits compared to files recorded at 48 kHz / 16 bits. However, since different bit-depths and equipment were used, the perceived differences cannot be attributed to the differences in sample rate alone.

To date, we do not know whether people can perceive differences between musical files recorded at 44.1 kHz and files recorded at higher sample rates. This question is critical to determine if high-resolution audio is economically viable. Furthermore, we aim to determine in which context people are more sensitive to sample rate differences, so that sound engineers can best choose the recording format as a function of the instrument(s) recorded and the acoustics of the room.

In this article, we hypothesize that expert listeners can discriminate musical files recorded at 44.1 kHz and 88.2 kHz. To test this hypothesis, we recorded five different musical excerpts, each presented in three different formats: 44.1 kHz, 88.2 kHz and the 88.2 kHz version down-sampled to 44.1 kHz. Except for the sampling rates, the exact same audio gear and settings were used for recording and playback. Overall, participants were able to discriminate between files recorded at 88.2 kHz and their 88.2 kHz to 44.1 kHz down-sampled version. Furthermore, for the orchestral excerpt, they were able to discriminate between files recorded at 88.2 kHz and files recorded at 44.1 kHz (p = .01).

2. METHODS

2.1. Participants

Sixteen expert listeners, fifteen males and one female[1], with a mean age of 30 (SD = 7.1), took part in the study and received CDN$20 per hour for their participation. All participants reported having studio experience in sound engineering for an average of 8 years (SD = 5.6). Six reported working as professional sound engineers in Montreal and ten were Sound Recording students at McGill University. All participants except one had musical training (15 years on average, SD = 5.5).

2.2. Musical excerpts

We recorded five musical excerpts corresponding to different instruments and hall acoustics, namely Orchestra, Cymbals, Classical Guitar, Voice and Violin (see details of the musical excerpts in Table 1). All musicians except the percussionist were performance students at Université du Québec à Montréal and McGill University.

All musical excerpts were captured with the exact same analog chain, consisting of a non-coincident pair of omnidirectional MKH 8020 microphones (Sennheiser, QC, Canada) and HV-3D preamplifiers (Millennia, CA, USA). The two microphones were separated by 30 cm (12 in), slightly angled (see an example from the cymbal recording in Figure 1).

1 The first author (AP) participated in the study.

Excerpt	Composer/Piece	Performer	Location	Room characteristics	Recording distance
Orchestra	Anton Bruchner Symphony NO. 6	McGill Symphony Orchestra directed by Alexis Hauser	Pollack Hall	Medium concert hall (600 seats) made of wood	50 cm (20 in) above the orchestra director
Cymbals	Improvisation	Mark Nelson	CIRMMT Immersive Presence Lab	Small dry room	50 cm (20 in) from the higher cymbal
Classical Guitar	Johann Kaspar Mertz An Malvina	Michel Salvail	CIRMMT Critical Listening Lab	Small lively room made of wood	50 cm (20 in) from the guitar's soundhole
Voice	Libby Larsen A man can love two women	Margaret Rood	Tanna Schulich Hall	Small concert hall (200 seats) made of wood	150 cm (60 in) from the mouth
Violin	Improvisation	Sonia Coppey	Tanna Schulich Hall	Small concert hall (200 seats) made of wood	150 cm (60 in) from the violin

Table 1. Details of the five musical excerpts used in the study

We chose these microphones for their frequency response, ranging from 10 Hz to 60 kHz. According to Nyquist theorem, the maximum possible frequency to be digitally converted at 88.2 kHz is 44.1 kHz. Therefore, the frequency response of the Sennheiser MKH 8020 microphones does not limit the sound quality when recording at a sample rate of 88.2 kHz.

We split the analog signal from the two outputs of the preamplifier, *i.e.* Left and Right, to four channels, *i.e.* Left and Right twice, that were digitally converted at 24 bits, both at 44.1 kHz and 88.2 kHz, using two Micstasy analog-digital converters (RME, Germany). We used the 744T portable audio recorder (Sound Devices, WI, USA) to record the digital signal at 44.1 kHz, and Logic Studio software in a MacBook Pro (Apple, CA, USA) to record the digital signal at 88.2 kHz. The entire recording chain is detailed in Figure 2.

We isolated five short excerpts from our recordings, corresponding to musical phrases of five to eight seconds, both at 44.1 kHz and 88.2 kHz. No sound processing was applied, except for a fade-in and a fade-out in Pyramix 6 software (Merging Technologies, Switzerland). We made sure that the selected files at 44.1 kHz and 88.2 kHz had the exact same fades (in and out) and length. Then, we down-sampled the 88.2 kHz files to 44.1 kHz. We chose Pyramix to down-sample the files, this software being commonly used by sound engineers who record acoustic music in high-resolution formats. Furthermore, the down-sampling algorithm in Pyramix does not provide any settings that could possibly introduce bias.



Figure 1 Top view of cymbal recording in the Immersive Presence Laboratory of CIRMMT

In summary, five musical excerpts were available in three versions: 44.1 kHz, 88.2 kHz and the 88.2 kHz version down-sampled to 44.1 kHz. The experiment consisted of five blocks corresponding to the five musical excerpts. Each block consisted of 12 trials each, *i.e.* all possible pairwise combinations of the three different versions, each presented four times (twice in each of the two presentation orders).

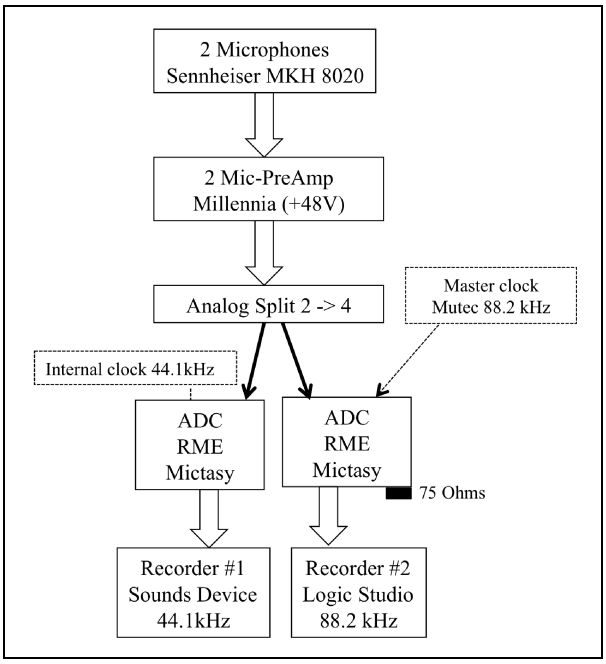


Figure 2. Recording diagram

2.3. Procedure

Participants were asked to perform a double blind ABX task. For each trial, the excerpt was presented with three versions, namely A, B and the reference X. A and B always differ. X is always either the same as A or the same as B. The participant's task is to indicate whether X = A or X = B. To nullify order effects, the order of presentation across trials and blocks was randomized.

Participants had to listen to all three versions presented in a trial at least once, and could then repeat each version as many times as desired or switch between versions while playing before making their decision. If they were unsure, they were asked to pick a version arbitrarily. Before the experimental session, we demonstrated the graphical interface with four practice trials. Listeners were free to adjust the sound level and their position if needed. The duration of the experiment ranged between two and four hours per participant, including a break between each block of trials.

The experiment took place in the Critical Listening Laboratory of the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT, Montréal, QC, Canada). This ITU standard room provides high quality controlled listening conditions. Stimuli were presented through an RME Fireface 800 digital-analog converter, a Grace m906 monitor controller (Grace Design, CO, USA), a Classé CA-5200 stereo amplifier (Classé Audio, QC, Canada) and B&W 802D loudspeakers (Bowers & Wilkin, West Sussex, England). Although the RME Fireface 800 may not be considered a high-end digital-analog converter, we used it as it was the only converter that allowed us to switch sample rates between 44.1 kHz and 88.2 kHz in a reasonable amount of time. To avoid clipping, we adjusted delays in our user interface, programmed in Max/MSP/Jitter 5 (Cycling '74, CA, USA), resulting in 730 ms between each version. B&W 802D loudspeakers have a frequency response ranging from 27 Hz to 33 kHz, thus allowing high-resolution audio formats to be reproduced in good conditions regarding the high frequency content.

2.4. Post-study questionnaire

After the listening task, participants were invited to fill out a questionnaire. The first part concerned demographical information, studio experience and musical training. Then, expert listeners were asked to rate the difficulty of the listening task on a scale of 0 to 10, as well as to describe the perceptual differences between the different versions. Finally, we asked which sample rate(s) they commonly use while recording and why.

3. RESULTS

3.1. Overall discrimination

Cumulative binomial tests on the number of correct responses were conducted for each participant, collapsing over all comparison pairs and all musical excerpts. At this individual level, three expert listeners out of 16 obtained significant results, $p < .05$, 2-tailed. However, they significantly selected the wrong answer, suggesting that they could hear

differences between A and B but picked the wrong one (*e.g.* A = X when in fact B = X). Subsequently, we will present the results of these three participants separately. The remaining 13 participants did not perform above chance level, either at the individual or group level, $p > .05$, 2-tailed, when collapsing over all format comparison pairs and all musical excerpts.

We applied detection theory to take into consideration the false alarm rate [1][4]. This analysis confirmed our findings, *i.e.* whenever the binomial test was significant, the corresponding |d'| was greater than 1 and 95% confidence interval did not include 0.

To further test our research hypotheses, performance results were analyzed as a function of format discrimination.

3.2. Format discrimination

We conducted binomial tests on the number of correct responses for each format comparison collapsing over all 13 participants and all musical excerpts. Significant results were observed for the comparison between files recorded at 88.2 kHz and their down-sampled 44.1 kHz version, $p = .04$, 1-tailed[2]. A tendency was observed for the comparison between files recorded at 88.2 kHz and 44.1 kHz, $p = .1$. No significant result were observed for the comparison between files recorded at 44.1 kHz and files down-sampled to 44.1 kHz, $p = .2$.

The same tests were conducted for the three participants who significantly picked the wrong answer. Significant results were observed for the comparison between files recorded at 88.2 kHz and their down-sampled 44.1 kHz version, as well as for the comparison between files recorded at 44.1 kHz and files down-sampled to 44.1 kHz, $p = .02$, $p < .001$, respectively. However, no significant results were observed for the comparison between files recorded at 88.2 kHz and 44.1 kHz, $p = .15$.

2 1-tailed binomial test were used to test our directional research hypothesis.

3.3. Discrimination by musical excerpts

Figure 3 represents the percentage of times the 13 remaining participants selected the correct answer for each format comparison and musical excerpt. Using the binomial test, performances over 63 % indicate that expert listeners could discriminate between the two versions and picked the correct answer. Performances ranging between 37 and 63 % are not significant ($p > .05$), suggesting that listeners could not discriminate between the two versions.

Regarding the comparison between files recorded at 88.2 kHz and 44.1 kHz, significant results were observed for the Orchestra excerpt only, $p = .02$. Regarding the comparison between files recorded at 88.2 kHz and their down-sampled 44.1 kHz version, significant results were observed for the Classical Guitar and the Voice excerpts, $p = .004$, $p = .04$, respectively. Regarding the comparison between files recorded at 44.1 kHz and files down-sampled to 44.1 kHz, no significant result was observed for any musical excerpt.

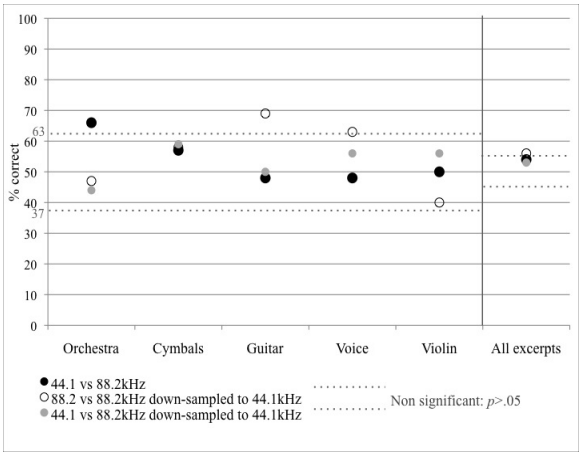


Figure 3 Discrimination results for the 13 remaining participants (n = 149 for Orchestra, n = 150 for Cymbals, n = 156 for Classical Guitar, Voice and Violin, N=767 for all excerpts)

Figure 4 presents the percentage of times the three participants who significantly picked the wrong answer selected the correct answer for each format comparison and the musical excerpt. Using the binomial test, performances under 17 % indicate that listeners could discriminate between the two versions but picked the wrong answer. Performances ranging between 17 and 83 % are not significant ($p > .05$),

suggesting that listeners could not discriminate between the two versions.

It should be noted that significance levels depend on the number of observations, hence the different dotted lines in figures 3 and 4 (number of observations mentioned in the captions).

Regarding the comparison between files recorded at 88.2 kHz and 44.1 kHz, significant results were observed for the Violin excerpt only, $p = .006$. Regarding the comparison between files recorded at 88.2 kHz and their down-sampled 44.1 kHz version, no significant result was observed. Regarding the comparison between files recorded at 44.1 kHz and files down-sampled to 44.1 kHz, significant results were observed for the Classical Guitar and the Violin excerpts, $p = .02$, $p = .006$, respectively.

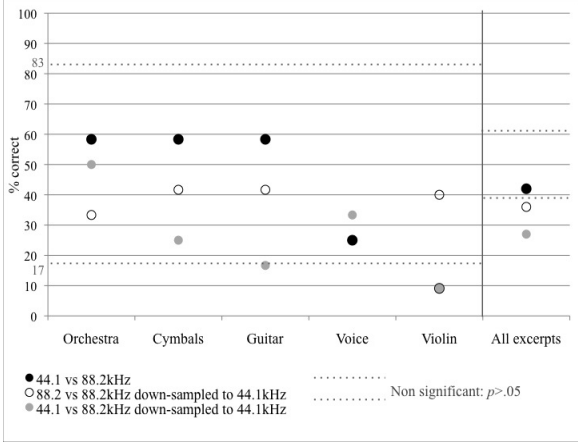


Figure 4 Discrimination results for the three participants who significantly picked up the wrong answer (n = 36 for Orchestra, Cymbals, Classical Guitar and Voice, n = 32 for Violin, N=176)

Although these three participants significantly picked the wrong answer over all comparison formats and musical excerpts, we observed that for the comparison between the Orchestra and Cymbals files recorded at 88.2 kHz and 44.1 kHz, the percentage of correct answers were similar to those of the 13 remaining participants. However, they did not reach statistical significance given the low number of observations. When collapsing over all 16 participants, the results of the comparison between Orchestra files recorded at 88.2 and 44.1 kHz is still significant, $p = .01$.

3.4. Post questionnaire

On a scale from 0 to 10, expert listeners reported that the difficulty level of the task was 9 on average ($SD = 1.1$). They commented that the task was very demanding in terms of concentration and that it was hard to stop doubting about what they heard. Thirteen out of 16 participants described in their own words the perceived differences between the different versions. We extracted a total of 16 phrasings from these verbal descriptions and grouped them into five categories of sound criteria, namely *spatial reproduction* (7 occurrences), *high frequency richness* (7 occ.), *timbre* (5 occ.), *precision* (5 occ.) and *fullness* (2 occ.).

Ten out of 16 participants reported that they are used to working both at 1 fs[3] (*i.e.* 44.1 or 48 kHz) and 2 fs (*i.e.* 88.2 or 96 kHz) in recording studios. Six participants further specified that their choice of sampling rate depends on the format of final delivery. More specifically, three mentioned selecting 2 fs for classical music and 1 fs for pop music due to Digital Signal Processing limitations. Five other participants reported always recording at 1 fs, and the remaining one only recording at 88.2 kHz. Overall, participants justified recording at 1 fs because of storage space (5 occ.) and equipment limitations (4 occ.); while they chose to record at 2 fs to enhance the sensation of space (3 occ.) and to get the highest possible resolution (3 occ.).

4. DISCUSSION

Findings from the listening tests suggest that expert listeners can detect differences between musical excerpts presented at 88.2 kHz and 44.1 kHz. Moreover, the qualitative analysis of verbal descriptors indicates that these differences were perceived in terms of spatial reproduction, high frequency content, timbre and precision. However, the ability to perceive these differences depends on the format comparison and musical excerpt. Listeners could significantly discriminate between files recorded at different sample rates only for the orchestral excerpt, the only recording of a complex scene with different musical instruments playing in a medium concert hall. This finding provides support for theories that high-resolution formats better

3 Frequency sample or sample rate

reproduce the details of transients and room acoustics [10][11].

Furthermore, our findings show that listeners were more sensitive to differences between files recorded at 88.2 kHz and their 44.1 kHz down-sampled version, than to differences between files recorded at different sample rates. As we down-sampled the files through a single software program, further investigation of down-sampling algorithms is required to draw conclusions regarding the impact of down sampling vs. recording at 44.1 kHz. However, our findings question the common practice of recording at high sample rates and later down sampling, as it seems to lower the sound quality more than recording directly at 44.1 kHz. Therefore, sound engineers should consider the format of final delivery and commercial release before choosing the recording sample rate.

While we observed audible differences between sample rates of 88.2 and 44.1 kHz, they remain very subtle and difficult to detect. It is difficult to interpret why three out of 16 participants significantly picked the wrong answer. We verified every step of the data collection and analysis. A possible reason could be that given the difficulty and duration of the listening test, participants doubted so much that they lost confidence and systematically picked the wrong answer.

It should also be noted that all the files used in this study were recorded and presented in 24 bits. Thus, we were not comparing the CD standard (*i.e.* 44.1 kHz, 16 bits) with high-resolution formats but restricted our experiment to sample rate discrimination. This choice was based on the fact that limitations of bit-depth of the CD standard at 16 bits have been identified and documented [10]. Therefore, differences between CD standard and high-resolution audio formats should be easier to detect than the differences observed in this study.

Participants suggested using more excerpts with long reverberation in future experiments. Indeed, we focused here on different instruments and only included one complex auditory scene in a medium hall. For this orchestral excerpt only, participants were able to significantly discriminate between 44.1 kHz and 88.2 kHz. These perceptual differences will be further investigated by varying systematically and independently the complexity of the auditory scene and the acoustics of the room. Furthermore, we plan to replicate this study with professional musicians to quantify the extent to which our ability to hear differences between sample rates depends on expertise. We will also extend this research to preference tests on the file comparisons that provided significant results. Furthermore, the stimuli used for our listening tests were also recorded simultaneously through different analog-digital converters. We are currently investigating the sensitivity of expert listeners to different converters.

5. ACKNOWLEDGMENT

The work reported herein was funded by an FQRSC team grant on the Perception of Audio Quality (P.I.: I. Fujinaga, CG and 4 co-applicants). The user interface for the listening task was programmed in Max/MSP by Guillaume Boutard and Julien Boissinot. The authors would like to thank the musicians who participated in the stimuli recordings, and the expert listeners who offered their listening skills and time for this study. The authors would also like to thank Julien Boissinot, Yves Methot and Harold Kilianski for technical assistance during the recordings and the experiments conducted at the Centre for Interdisciplinary Research in Music Media and Technology (Montréal, QC, Canada), and Maryse Lavoie and Aaron Rosenblum for proofreading.

6. REFERENCES

[1] Boley, J., & Lester, M., "Statistical Analysis of ABX Results Using Signal Detection Theory," presented at the AES127th Convention, New York, NY, USA, 2009 October 9-12.

[2] Laugier, V., "La Haute Résolution Audionumérique," Master thesis at CNSMDP (France). (June 2005) [Not published]

[3] Leonard, B., "The Downsampling Dilemma: Perceptual Issues in Sample Rate Reduction," presented at the AES124th Convention, Amsterdam, The Netherlands, 2008 May 17-20.

[4] Macmillan, N. A., Creelman, C. D., "Detection theory: A user's guide", Cambridge: University Press Cambridge. (1991)

[5] Meyer, E. B., & Moran, D. R., "Audibility of a CD-Standard A/D/A Loop Inserted into High-Resolution Audio Playback", J. Audio Eng. Soc., vol. 55(9), pp. 775-779. (September 2007)

[6] Nishiguchi, T., Hamasaki, K., Iwaki, M., Ando, A., "Perceptual Discrimination between Musical Sounds with and without Very High Frequency Components," presented at the AES115th Convention, NY, USA, 2003 October 10-13.

[7] Nordmark, J. O., "Binaural time discrimination", J. Acoust. Soc. Am., vol. 60(4), pp. 870-880. (October 1976)

[8] Nyquist, H., "Certain Topics In Telegraph Transmission Theory", reprinted from AIEE Trans., vol. 47, pp. 617-644. (April 1928). IEEE Proc., vol. 90(2), pp. 280–305. (2002).

[9] Rumsey, F., "High Resolution Audio", J. Audio Eng. Soc., vol. 55(12), pp. 1161–1167. (December 2007)

[10] Stuart, J., "Coding for high-resolution audio systems", J. Audio Eng. Soc., vol. 52(3), pp. 117-144. (March 2004)

[11] Woszczyk, W., "Physical and perceptual considerations for high-resolution audio," presented at the AES115th Convention, New York, NY, USA, 2003 October 10-13.

[12] Yoshikawa, S., Noge, S., Ohsu, M., Toyama, S., Yanagawa, H., Yamamoto, T., "Sound Quality Evaluation of 96 kHz Sampling Digital Audio, presented at the AES99th Convention, New York, NY, USA 1995, October 6-9.