

PAPER

Detection threshold for distortions due to jitter on digital audio

Kaoru Ashihara^{1,*}, Shogo Kiryu¹, Nobuo Koizumi², Akira Nishimura²,
Juro Ohga³, Masaki Sawaguchi⁴ and Shokichiro Yoshikawa⁵

¹National Institute of Advanced Industrial Science and Technology, AIST,
Tsukuba Central 6, Tsukuba, 305-8566 Japan

²Tokyo University of Information Sciences,
1200-2, Yato-cho, Wakaba-ku, Chiba, 265-8501 Japan

³Shibaura Institute of Technology,
3-9-14, Shibaura, Minato-ku, Tokyo, 108-8548 Japan

⁴Japan Broadcasting Corporation,
2-2-1, Jinnan, Shibuya-ku, Tokyo, 150-8001 Japan

⁵Media Laboratory S,
3-1-10, Mure, Mitaka, 181-0002 Japan

(Received 12 April 2004, Accepted for publication 16 July 2004)

Abstract: Detection threshold for distortions due to time jitter was measured in a 2 alternative forced choice paradigm with switching sounds. Music signals with random jitter were simulated on the digital domain. The size of jitter was arbitrary controlled so that the detection threshold could be estimated. Professional audio engineers, sound engineers, audio critics and semi-professional musicians participated as listeners. The listeners were allowed to use their own listening environments and their favorite sound materials. It was shown that the detection threshold for random jitter was several hundreds ns for well-trained listeners under their preferable listening conditions.

Keywords: Distortion, Jitter, High-resolution audio, Sampling

PACS number: 43.38.Md, 43.58.Ry [DOI: 10.1250/ast.26.50]

1. INTRODUCTION

The sampling period of linear PCM data has to be strictly constant during a digital to analogue conversion in order to reproduce the sounds as they were recorded. If this sampling period is quivering, the reproduced waveform will be different from its original form and therefore the sounds are distorted. Although there are little scientific evidences, it is presumed that this phenomenon, so-called, time jitter depends on a modulation of the power supply, the cable length, circuit implementation details in products [1].

It is pointed that the contents in the next-generation audio that contain more high frequency components than conventional CDs may be more susceptible to jitter, as the jitter-introduced distortion is directly proportional to the frequency of the signal [1,2]. In the case of linear PCM data, amount of distortions can be estimated from a slope of the waveform and a size of jitter. If the distortions are

smaller than the quantization noise, jitter does not induce any degradation of sound quality. Because the maximum slope in a waveform of a high frequency sound is larger than that of a low frequency one, the same amount of jitter results in the more distortions in the higher frequency sounds.

Accordingly, the maximum acceptable random jitter can be defined as the jitter that results in a 1-quantizer level error when the signal is of the highest recordable frequency [1,3]. When the highest recordable frequency is 20 kHz and the quantization bit number is 16, the maximum acceptable size of random jitter is 121.4 ps. In other words, jitter has to be smaller than 121.4 ps in order to reproduce a 20 kHz tone with a 16-bit resolution [3].

Even in the case of music sounds, the maximum acceptable size of random jitter can be estimated if the maximum slope in the sound waveform is known. By analyzing various signals in a lot of CDs, it was shown that jitter has to be as small as several hundreds ps in some cases to preserve a 16-bit resolution [3].

However, these estimations are not practical because

*e-mail: ashihara-k@aist.go.jp

such small distortions are supposed to be masked by the background noise and the internal noise of the reproduction system in the real environment. It is much more practical to evaluate jitter based on its detection threshold. Benjamin and Gannon [1] made an attempt to measure thresholds of audibility for sinusoidal jitter on program materials. Their study seems to have a few problems, however. In their study, a special arrangement was made to the reproduction system in order to add jitter at the digital interface. It is not known if such an arrangement might change sound quality and affect the results. Secondly, they employed a self-administered threshold evaluation in which the listeners determined their thresholds at their discretion. Their results might contain errors due to the cognitive factor.

Ashihara and Kiryu [3] measured thresholds for random jitter on music signals. All listeners in that study were graduate or undergraduate students and the most of them were not well-trained listeners. As Benjamin and Gannon reported, detection thresholds for jitter may depend on how well the listeners are trained.

In the studies of Benjamin and Gannon and Ashihara and Kiryu, the listening conditions and the materials used were fixed for all listeners except that the volume was adjusted for each listener. Although a fixed listening condition and fixed sound materials are preferable for inter-subject comparison and retests, no fixed condition may satisfy all listeners' preferences. The optimal listening condition is difficult to define because that tastes for listening conditions vary among listeners. In the present study, most equipment, environment, and sound materials were not decided and provided by the examiners but by each listener. Because a software simulator of jitter proposed by Ashihara and Kiryu [2–4] was used, no special modification of the reproduction system was needed except that a notebook PC with a mouse had to be used in the experiment. Listeners in the present study were audio professionals or semi-professionals who were supposed to be well-trained listeners. Thresholds were estimated based on the discrimination scores of the listeners in a 2 alternative forced choice paradigm.

2. METHOD

2.1. Listening Conditions and Listeners

In order to measure detection thresholds for jitter, it is necessary to control the size of jitter. To control the size of the actual jitter during reproduction of music sounds, a special digital-to-analogue converter (DAC) or special alternation of DAC is needed. Choice of the DAC, however, is a very crucial part in configuration of the listening condition. To avoid using a special DAC, we had proposed a software simulator of jitter [2–4]. By using this simulator, we could predict distortions due to any given jitter for any given PCM data. We could, therefore,

quantitatively control the amount of jitter.

Experiments were carried out in the listening booth or studio that each listener had offered. The examiner only brought there a personal computer with a digital audio interface and a mouse and each listener provided his or her favorite DAC, amplifiers and loudspeakers.

The sound materials were also selected by each listener. Materials that had been repeatedly heard by the listener could be used.

A total of 23 audio professionals or semi-professionals participated as the listeners. They were audio engineers, audio critics, sound engineers, and musicians. Some of them were volunteers and the others were paid for their participation. All of them were willing to participate in the experiments.

2.2. Simulation of Jitter

Simulation of jitter consisted of interpolation, shifting samples and an anti-aliasing filtering. Figure 1 shows a schematic illustration of an artificial time jitter on PCM data. Open circles represent the original samples and filled triangles represent the artificially jittered samples. A dashed line was a distorted waveform obtained by interpolation of the jittered samples. Vertical grids represent the ideal sampling positions. By re-sampling the distorted waveform at the ideal sampling period, PCM data

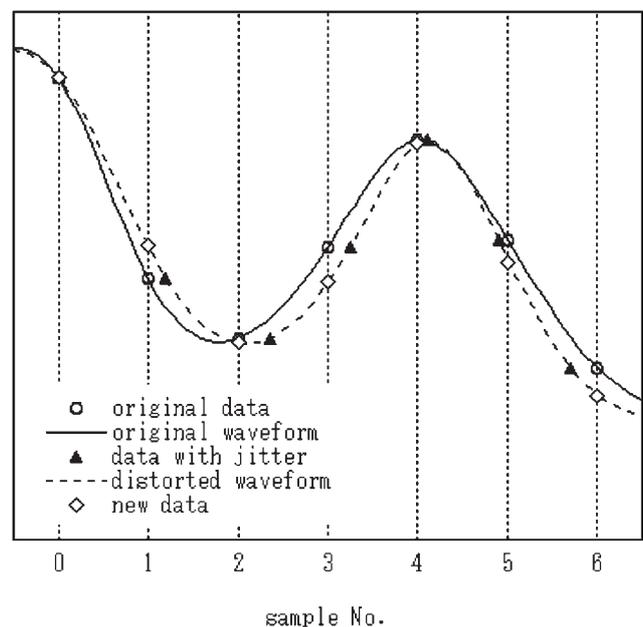


Fig. 1 PCM data before and after addition of artificial jitter are shown. A solid line, open circles, and filled triangles represent the original waveform, samples without distortions and the artificially jittered samples, respectively. Vertical grids represent the ideal sampling positions. A dashed line and open diamonds represent a distorted waveform and PCM data with distortions, respectively.

containing distortions were obtained (open diamonds). As addition of jitter might introduce components above the Nyquist frequency, an anti-aliasing process was needed [2]. The PCM data finally obtained shared the same format with the original PCM data so that the data with and without distortions could be reproduced in the same manner without using any special alternation of equipment.

It was confirmed that the distortions obtained by this method using sinusoidal signals exactly agreed with what were predicted by the phase modulation theory [2].

Each listener had presented the raw PCM data to the examiners in advance. Their distorted versions were prepared by adding artificial random jitter of various amounts. Although most of these materials were music sounds recorded on commercial package media, there were also recordings from radio dramas.

All data were filed on a notebook PC. This notebook PC was brought to the place where the listener had provided a DAC, amplifiers, loudspeakers, a PC monitor and whatever he or she had wanted, headphones for instance. A diagram of the reproduction system is shown in Fig. 2. No modification except for existence of the notebook PC with a mouse, was given to the listener's own listening environment.

2.3. Procedure

Discrimination tests were performed in a 2 alternative forced choice paradigm. Each run consisted of a full reproduction of a material. The duration of sound materials ranged from about 2 min. to 4 min. During a run, two versions of the same material were simultaneously reproduced. They were a reference version without jitter and an artificially jittered version. Only one of them could be presented acoustically in the foreground while the other

was reproduced in the background (muted). The foreground and the background could be switched by a listener at any moment in a run.

A PC monitor was placed in front of a listener. On its screen, three buttons were displayed. Each button was labeled either 'A,' 'B,' or 'X,' respectively. Button X always made reference version audible. Either button A or button B did the same work as button X. Only the other button made the jittered version audible. The instruction given to the listener was as follows.

'There are 3 versions (A, B, and X) of each material. Version X is a reference version. Version A and version B are comparison versions. The reference version is the sound without distortion. One of the comparison versions is identical to the reference while the other version contains distortions. Although only one of them can be heard at one time, you can switch the versions by clicking the buttons on the screen. During a reproduction, you can switch versions as often as you like to compare the sounds. When the reproduction is completed, please judge which of version A or B was identical to version X.'

To prevent clicks from occurring when the versions interchanged, they were smoothly cross-faded in 100 ms. Switching between versions was processed in the digital domain without using electrical switches. A session started with relatively easy condition containing jitter as large as several μ s. A listener had to make at most 9 judgments for the same condition. Because it was a 2 alternative forced choice paradigm, only when the listener scored 75% correct or better, he or she could succeed to the next condition where the size of the artificial jitter was to be reduced by half. When 3 wrong answers before 7 correct answers counted, the session was terminated without succeeding to the next condition. When the listener felt the task was very easy, the run could be terminated before the reproduction completed to save time.

3. RESULT AND DISCUSSION

Table 1 shows the number of listeners who could discriminate between sounds with and without jitter at each

Table 1 Size (r.m.s.) of random jitter added to materials and the number of listeners who could discriminate between sounds with and without jitter.

Size of random jitter (r.m.s.)	Number of listeners who discriminated sounds
2 μ s	23
1 μ s	11
500 ns	6
250 ns	none

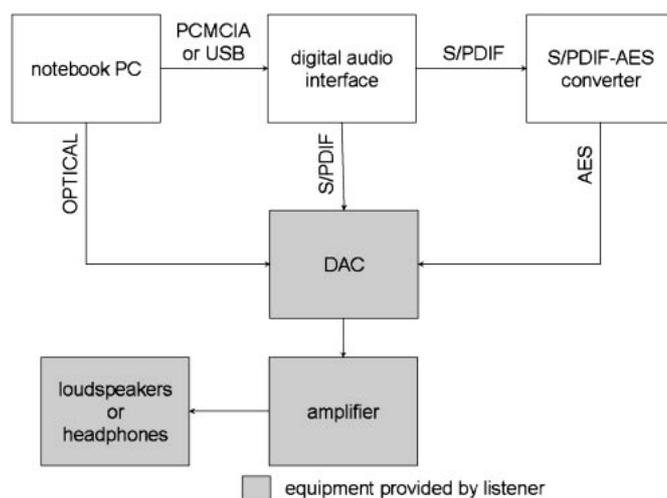


Fig. 2 A block diagram of the reproduction system is shown. A DAC, amplifiers, and loudspeakers or headphones were selected by the listener.

jitter size. When the jitter size was $2\ \mu\text{s}$ (r.m.s.), all listeners scored more than 75% correct. About 25% of the listeners detected jitter when its size was 500 ns. When it was 250 ns, however, no listener could discriminate the sounds. The thresholds for random jitter added to program materials seem to be several hundreds ns. Ashihara and Kiryu [3] measured detection thresholds for random jitter in a fixed listening condition. 14 listeners without special training participated. Some listeners could detect jitter of 1,152 ns but no one detected jitter of 576 ns in their study. Even though they used a fixed condition and non-professional listeners, their result is comparable to the present result. Tomizawa *et al.* [5] measured detection threshold for artificial jitter under the headphone listening condition. In their study, thresholds ranged from several hundreds ns to several μs . They argued that the detectability of jitter would depend on the contents of materials.

Benjamin and Gannon [1] reported that the thresholds for jitter on program materials ranged from 30 ns to 300 ns. The threshold values in their study were a little bit smaller than those in the present study. It may be attributed to several differences in the methods.

In the study of Benjamin and Gannon, they connected the output of the CD player to a distribution amplifier with two outputs. The first output was connected to a DAC via an AB comparator box. The second output from the distribution amplifier was connected to a jitter modulator that could add jitter by using a function generator as the jitter source. The output from the jitter modulator was connected to the DAC via the AB comparator box. Two signals to be compared were, therefore, from the same CD player but had different pathways. This might result in a slight change of sound quality. In the present study, two versions of the material were reproduced completely in the same manner without using special equipment.

Secondly, Benjamin and Gannon used sinusoidal jitter instead of random jitter. They, furthermore, selected the jitter frequency that might result in detectable distortions based on observed spectra of the signals. Jitter frequency they used ranged between 1,530 Hz and 1,850 Hz. Loudspeakers were used in the most cases in the present study while Benjamin and Gannon used headphones.

Another difference which is supposed to be most important is that in Benjamin and Gannon's study, the listeners were allowed to decide their thresholds at their own discretion. The listeners were asked to adjust the jitter level until they decided that their thresholds were reached. The reliability of their own decision was not verified. This self-administered threshold estimation might allow underestimation of the threshold values. In the present study, thresholds were determined objectively based on the scores in a 2 alternative forced choice paradigm where the listeners could not determine their thresholds at their

discretion.

It can be concluded that detection threshold for random jitter added to program materials is several hundreds ns even for well-trained listeners under their preferable listening conditions. According to Benjamin and Gannon, sinusoidal jitter as small as 30 ns (r.m.s.) might be detectable under a certain condition. Considering these results, the maximum acceptable size of jitter would be the order of ns.

In some contents of conventional CDs, it had been observed that jitter had to be as small as several hundreds ps to preserve the resolution of 16 bits [3]. This is way below the detection threshold values. Nishimura and Koizumi [6,7] made attempts to measure actual jitter of various DA systems during reproduction of music signals. They could not detect any jitter larger than 3 ns in their measurements.

So far, actual jitter in consumer products seems to be too small to be detected at least for reproduction of music signals. It is not clear, however, if detection thresholds obtained in the present study would really represent the limit of auditory resolution or it would be limited by resolution of equipment. Distortions due to very small jitter may be smaller than distortions due to non-linear characteristics of loudspeakers. Ashihara and Kiryu [8] evaluated linearity of loudspeaker and headphones. According to their observation, headphones seem to be more preferable to produce sufficient sound pressure at the ear drums with smaller distortions than loudspeakers.

4. CONCLUSION

In order to determine the maximum acceptable size of jitter on music signals, detection thresholds for artificial random jitter were measured in a 2 alternative forced choice procedure. Audio professionals and semi-professionals participated in the experiments. They were allowed to use their own listening environments and their favorite sound materials. The results indicate that the threshold for random jitter on program materials is several hundreds ns for well-trained listeners under their preferable listening conditions. The threshold values seem to be sufficiently larger than the jitter actually observed in various consumer products.

ACKNOWLEDGEMENTS

This study is a part of the joint research of the technical committee of high definition audio. We thank all who were willing to spare their time and participate in the experiments. Special thanks are given to Mr. Kazumasa Takahashi, Mr. Seigen Ono, Accuphase Laboratory INC. and Japan Broadcasting Corporation.

REFERENCES

- [1] E. Benjamin and B. Gannon, "Theoretical and audible effects of jitter on digital audio quality," *Preprint of the 105th AES Convention*, #4826 (1998).
- [2] K. Ashihara and S. Kiryu, "Simulation of sound degradation due to time jitter on digital audio," *J. Acoust. Soc. Jpn. (J)*, **58**, 232–238 (2002).
- [3] K. Ashihara and S. Kiryu, "The maximum permissible size and detection threshold of time jitter on digital audio," *J. Acoust. Soc. Jpn. (J)*, **59**, 241–249 (2003).
- [4] S. Kiryu and K. Ashihara, "A jitter simulator on digital data," *Preprint of the 110th AES Convention*, #5390 (2001).
- [5] T. Tomizawa, H. Ohtake and J. Ohga, "Effect of jitter for listening by a few musical signals," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 703–704 (2003).
- [6] A. Nishimura and N. Koizumi, "Measurement of sampling jitter using a musical signals," *Preprint of the 114th AES Convention*, #5797 (2003).
- [7] A. Nishimura and N. Koizumi, "Measurement and analysis of sampling jitter in digital audio products," *Proc. ICA 2004*, IV, pp. 2547–2550 (2004).
- [8] K. Ashihara and S. Kiryu, "Linearity evaluation of loudspeakers and headphones," *J. Acoust. Soc. Jpn. (J)*, **56**, 713–720 (2000).