Audio Engineering Society

Conference Paper 23

Presented at the 6th International Conference on Audio for Games
2024 April 27–29, Tokyo, Japan

Precise Construction Method for Virtual Headphones

Haruto Hirai[1], Kazuhiko Kawahara[2], Nao Tojo[3], Atsushi Hara[3], and Kimio Hamasaki[3]

[1]*Graduate School of Design, Kyushu University, Fukuoka, Japan*
[2]*Faculty of Design, Kyushu University, Fukuoka, Japan*
[3]*final Inc., Kanagawa, Japan*

Correspondence should be addressed to Kazuhiko Kawahara (`kazuhiko_kawahara@ieee.org`)

ABSTRACT

This study examined a method of constructing "Virtual Headphones," in which one set of headphones simulates another. Methods for accurately measuring headphone characteristics and designing a filter to construct the virtual headphones were studied. The simulated performance of the designed filter was confirmed through subjective evaluation experiments. The proposed method exhibited physically high simulated performance, but the subjective similarity was not high. The accuracy of simulating high-frequency bands of several kHz and above was necessary for the high subjective simulation performance of the virtual headphones. The need to compensate for differences in individual ear shapes was suggested.

1 Introduction

Many people use headphones to listen to music and game sounds. Target response curves of headphones include the effects of the head-related transfer function lost by headphone listening and the free-field and diffuse-field target curves. Furthermore, some target curves have been subjectively preferred over free- and diffuse-field ones [1].

However, these target curves assume speaker playback. Many people use headphones instead of speakers; there has been an increase in sound sources that assume headphone playback (e.g., binaural sound). When listening to such sources with headphones, target curves that exploit the features of headphones, rather than assuming speaker playback, might be more suitable. Consequently, it is necessary to design an evaluation experiment to compare the sound quality of headphones. However, because headphones need to be physically worn, designing a subjective evaluation experiment that eliminates the influence of tactile fit on listening impressions is challenging.

The "Virtual Headphone" concept has been proposed to enable such experiments by Olive et al. [2]. Virtual headphones simulate the sound quality of another headphone by equalizing the acoustic characteristics of a headphone to the acoustic characteristics of another headphone.

In the study by Olive et al. [2], virtual headphones were constructed by equalizing the amplitude–frequency response. The results of sound quality evaluation experiments among actual and virtual headphones were highly correlated. Consequently, similar results can be obtained from sound quality evaluation experiments using actual headphones through subjective evaluation experiments using virtual headphones. However, in the method in the paper, the filter does not compensate for the frequency band above 10 kHz. Also, there are issues we had inspired, such as the filter design method for constructing the virtual headphones and trying to directly compare the actual and virtual headphones subjectively. Moreover, to what extent do the virtual headphones simulate the sound quality of the actual headphones?

This study examined in detail the method of accurate headphone measurement to design a filter that corrects the full audible frequency range and the design method of filters to construct virtual headphones. Then, virtual headphones were constructed using the designed filter, a one-to-one comparison listening experiment with actual headphones was conducted, and the simulation subjective accuracy of the virtual headphones was verified against the actual headphones.

2 Virtual Headphone Construction

2.1 Headphone Measurement

When measuring headphones, the international standard IEC 60318-4 frequency range for the conventional device has an upper-frequency limit of 10 kHz, making it impossible to accurately measure the full audible frequency range with a measuring device compliant with this standard.

However, the upper limit of human audible frequency is up to 20 kHz. Independent of the decline with age and individual differences, perceived sound quality may also be affected by frequency responses above 10 kHz. Therefore, accurate measurements are necessary for high-frequency bands above 10 kHz where the reliability of measurements with existing couplers is questionable—these should be included in the range of filter design.

Furthermore, at several kHz and above, the amplitude–frequency response changes markedly because of differences in headphone positioning. As filters are designed based on the measured amplitude–frequency response, the size of variations in amplitude–frequency response at the measurement time affects filter accuracy. Therefore, the change in response caused by the position of the headphones can be broadly divided into two categories:

- Large differences caused by different positions
- Small differences caused by the fit with the ear and head, even at the same position

Small changes in response at the same position are also related to differences, such as the shape of the ear simulator or dummy head and the individual's head. Therefore, a method to minimize significant differences in response caused by different positions during measurement is considered.

Therefore, this study aimed to measure the headphone amplitude–frequency response with less variability using the ear simulator GRAS 45CC, which can accurately measure up to 20 kHz and is equipped with a positioning guide. The coupler equipped with this ear simulator complies with IEC 60318-4 up to 10 kHz and has a tolerance of ± 2.2 dB from 10 to 20 kHz. Therefore, accurately measuring headphones over the entire audible frequency is possible. Furthermore, by having scales printed on the plates around the ear and positioning guides to keep the position of the headphones on the ear simulator constant, variations in position can be suppressed when reseating headphones.

2.2 Equalizing Filter Design

This study equalized the amplitude–frequency response in the same way as in the previous study by Olive et al. [2] as a method of constructing virtual headphones. We design an equalizing filter by combining an inverse filter that cancels the amplitude–frequency response of the reference headphones and a simulation filter that simulates the amplitude–frequency response of the target headphones.

The average amplitude–frequency response of multiple measurements of the target headphones was used for the simulation filter because it is best represented by the average of multiple measurements [3].

When designing the inverse filter for the headphones, unintended peaks in the response may be possible after the filter application because (1) there are fine peaks and dips in the high-frequency band and (2) the response of the headphones changes depending on the wearing position. Although slight peaks in the amplitude–frequency response are easily recognized and affect sound quality, dips with high Q values are difficult to recognize and have less impact on sound quality [4]. Therefore, when designing the inverse filter, it is necessary to permit dips but not peaks. Therefore, this study used the method proposed by Bolaños et al. [5] using a frequency-dependent regularization parameter. This method enables control of the filter's frequency range and peaks without the need to determine the appropriate regularization parameter value.

Then, the equalizing filter $H(\omega)$ is expressed as

$$H(\omega) = \frac{C^*(\omega) \cdot A(\omega)}{|C(\omega)|^2 + \hat{\beta}(\omega)} \tag{1}$$

The parameter $\hat{\beta}(\omega)$ is obtained as,

$$\hat{\beta}(\omega) = \alpha(\omega) + \sigma(\omega)^2 \tag{2}$$

and the factors $\alpha(\omega)$ and $\sigma(\omega)$ is defined as

$$\alpha(\omega) = \left(\frac{1}{|W(\omega)|^2} - 1 \right) \tag{3}$$

$$\sigma(\omega) = \begin{cases} |C(\omega)| - |\hat{C}(\omega)|, & \text{if} \quad |\hat{C}(\omega)| \geq |C(\omega)| \\ 0, & \text{if} \quad |\hat{C}(\omega)| < |C(\omega)| \end{cases} \tag{4}$$

where $C(\omega)$ is headphone response, $A(\omega)$ is target response, $W(\omega)$ is unity gain filter, and $\hat{C}(\omega)$ is half-octave smoothed version of $C(\omega)$. The average amplitude–frequency response of multiple measurements is used for the response $C(\omega)$ of the reference headphones, just like the simulation filter. The unity gain filter $W(\omega)$ was designed as a band-pass filter in the frequency domain to have a frequency width of 20 to 20000 Hz, -3 dB at the cutoff frequency, and -60 dB outside the band. First, an inverse filter was designed with $A(\omega) = 1$, and the filter's accuracy was verified. Then, an equalizing filter was designed by setting $A(\omega)$ as the amplitude–frequency response of the simulation filter, with the headphones that best fit the inverse filter as the reference headphones.

3 Measurements and Filter Design Result

3.1 Measurement Method

This research used the five types of headphones presented in Table 1. Prices were obtained from the manufacturer's official or domestic online shop in Japan.

Table 1: Headphone Descriptions

Brand / Model	Design / Type / Driver / Retail Price (JPY)
Audio-Technica ATH-M50x	Circumaural / Closed / Dynamic / 20,900
Beyerdynamic DT990 Edition 2005	Circumaural / Open / Dynamic / 27,500
Final D8000 Pro	Circumaural / Open / Planar Magnetic / 495,000
Sennheiser HD650	Circumaural / Open / Dynamic / 71,500
SONY MDR-CD900ST	Circumaural / Closed / Dynamic / 19,800

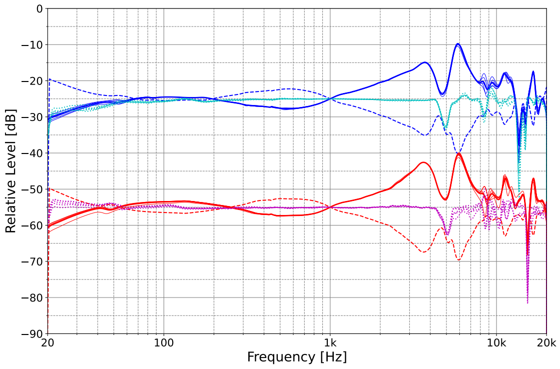
All the headphones are circumaural but include open-back and closed-back designs and different types of drivers, such as dynamic and planar magnetic. In this paper, all headphones are referred to as A∼D, or X, to prevent a direct correlation between headphone models and measurement results.

The measurements were conducted in an anechoic room. The headphones were plugged into the headphone output of the audio interface RME Fireface UC, which was connected to a PC. Then, they were measured on the ear simulator GRAS 45CC configurations with the pinna KB5010, 5011, and the coupler RA0402. For the measurements, final Inc. ' s in-house R&D software was used. The signal for measurement was log-TSP.

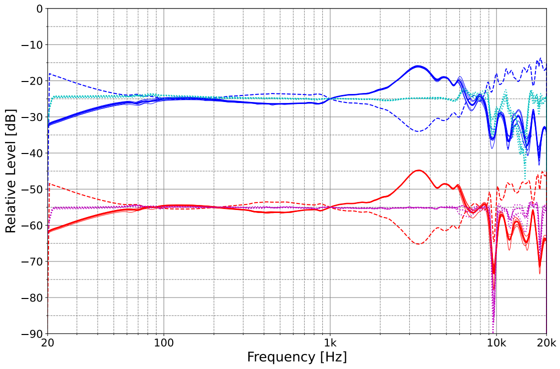
The headphones were measured five times each for the left and right, with the headphones being reseated between measurements. When putting on and taking off the headphones, efforts were made to ensure that the position of the headphones on the ear simulator was the same by visually confirming the memory printed on the side and using the positioning guide of the ear simulator. Filters were designed based on Eq. 1 in the frequency domain and calculated as a linear-phase finite impulse response filter with a sampling frequency of 48 kHz and 2^{16} taps. The calculations were performed in Python without using any special libraries.

3.2 Results of Measurement and Inverse Filtering

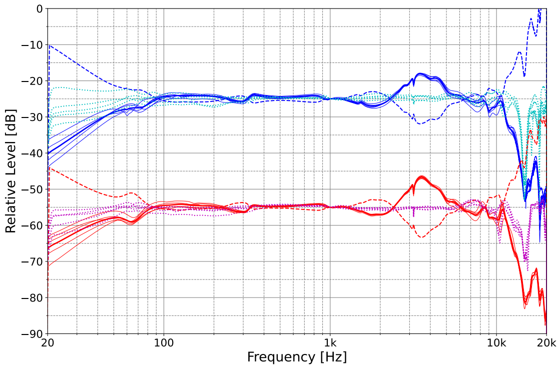
The measured results of each headphone and those using an inverse filter are depicted in Fig. 1. The headphone with the most stable characteristics for inverse filtering was Headphone X. If the peaks and dips of the amplitude–frequency response of the headphones are relatively few and have a smooth response, they are less affected by the regularization parameter, resulting in a better fit for the inverse filter. The dip that occurs in the response when the inverse filter of Headphone X is applied is thought to have little effect when using this inverse filter design method because the Q value is sufficiently high.



(a) Headphone A



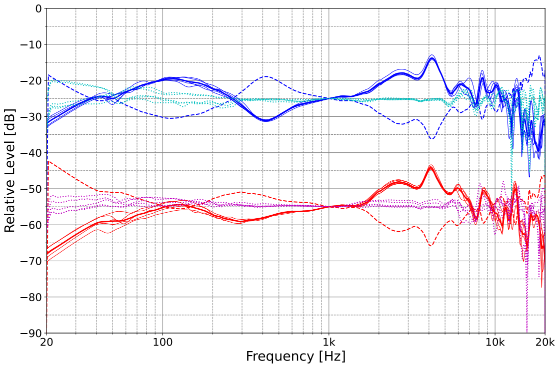
(b) Headphone B



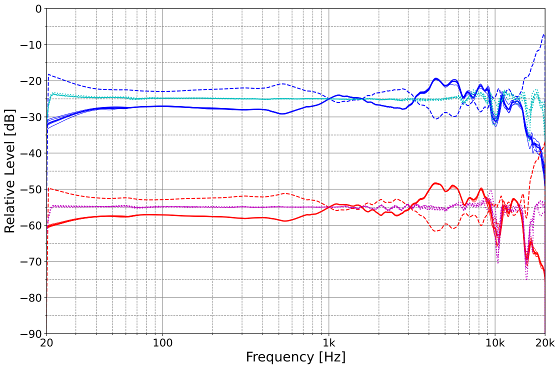
(c) Headphone C

3.3 Results of Equalizing Filter Design

We designed a filter to equalize the amplitude–frequency response of other headphones based on Headphone X, which had the best fit for the inverse filter. The measured results using an equalizing filter and differences between actual and virtual headphones are depicted in Fig. 2.

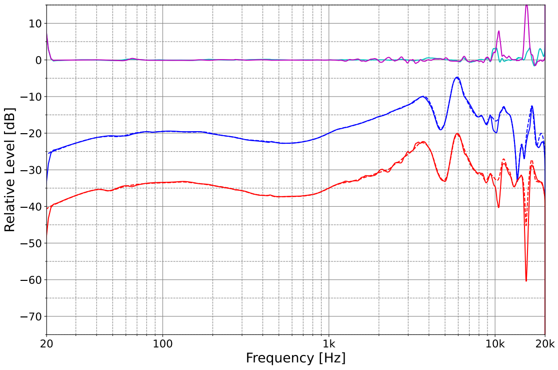


(d) Headphone D

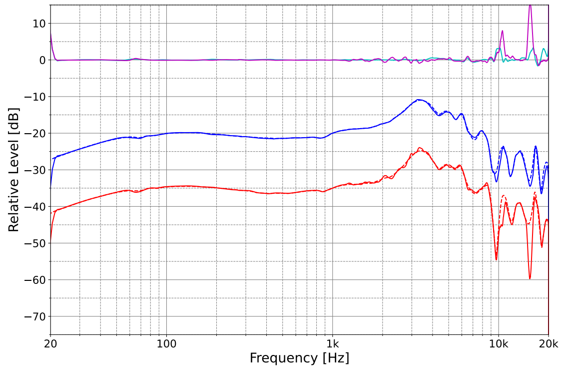


(e) Headphone X

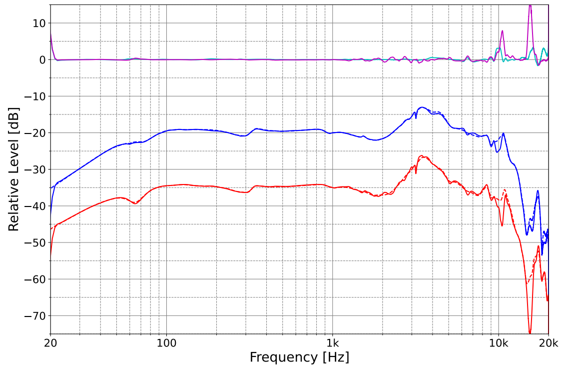
Fig. 1: Amplitude–frequency response (solid), inverse filter (dashed), inverse filtering result (dotted), Left: Blue and Cyan, Right: Red and Magenta



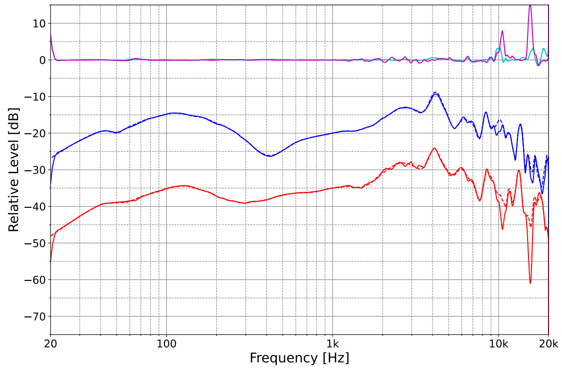
(a) Headphone A



(b) Headphone B



(c) Headphone C



(d) Headphone D

Fig. 2: Amplitude–frequency response of virtual headphones (solid), actual headphones (dashed), error (solid), Left: Blue and Cyan, Right: Red and Magenta

The frequencies at which differences occurred between the actual and virtual headphones matched the frequencies at which dips occurred in the inverse filter and generally matched the measurement results of the actual headphones at other frequencies. We considered that it is possible to design a filter that accurately simulates the response of actual headphones by suppressing variations in response caused by the position of the headphones with the measurement and filter design methods used in this study.

4 Subjective Evaluation

4.1 Experiment Method

The subjective simulation accuracy of the equalizing filter, physically verified in section 3.3, was validated through a subjective evaluation experiment. In the subjective evaluation experiment, a pair of actual and virtual headphones were presented with the same type of stimulus, and the impressions were compared and evaluated.

The headphones to be simulated were of four types, and there were five types of stimuli, with one evaluation for each combination of headphones and stimuli considered one trial. The evaluation of each of the five stimuli for one headphone, a total of five trials, was considered as one block, repeated for all four types of headphones for a total of 20 trials.

The headphones used in the experiment were the same as those in Section 3, and the reference headphones were the same as those in Section 3.3: Headphone X. The experiment was conducted in the anechoic booth. The actual and virtual headphones were plugged into the two headphone outputs of the audio interface RME Fireface UFX II, which was connected to the PC. The entire experiment was conducted according to a program written in Python, and the listening to and answering of stimuli was conducted on the GUI.

The participants in the experiment included 19 Japanese students (14 males and five females) from Kyushu University, aged 22 to 35 years. They were reported to have normal hearing and trained in listening according to the method in Iwamiya et al. [6]. The position of the headphones was determined by the participants (in the most natural position for them) without any special instructions from the experimenter. The brand and model names were hidden (taking care not to affect the headphone mechanism) to avoid the influence of listening impressions inferred from visual information (e.g., brand and model).

Participants were asked to respond to the similarity of the sounds listening through actual and virtual headphones and the differences in spectral balance.

Evaluation of Similarity – Participants were asked to select the most appropriate answer from a five-point scale of “not at all similar (0), not very similar (1), somewhat similar (2), very similar (3), no difference (4). ’’

Evaluation of Differences in Spectral Balance – Participants were asked to answer on an 11-point scale from -5 to +5 for seven equal log-spaced frequency bands. Compared with the actual headphones, if it felt that the virtual headphones had too much energy of a band, participants were instructed to assign a positive value, if too little energy, a negative value, and if exactly the same, zero.

The stimuli in Table 2 included 15 to 20 seconds of music extracted from CDs. Fig. 3 illustrates the average long-term spectra for the stimuli. The stimuli were normalized at -19 LUFS by calculating the loudness according to the method described in ITU-R BS.1770-4. The average playback level was adjusted to a root mean square level of 82 dB (C-weighting).

Table 2: Stimuli Characteristics

Artist/Track/Album/Year	Description
SW4MT: John Williams: London Symphony Orchestra / Main Title Rebel Blockade Runner / Star Wars: A New Hope – Original Motion Picture Soundtrack/1997	Classical with Symphony Orchestra
AL1: Bill Evans Trio / Autumn Leaves [Take 1] / Portrait In Jazz / 1959	Jazz Piano
CTW: Eric Clapton / Change the World / Clapton Chronicles / 1996	Rock Male Vocal
HANA: Hikaru Utada / Hanataba Wo Kimini / Fantôme / 2016	Pops Female Vocal
MJ: HKT48 / Melon Juice / Melon Juice / 2013	Pops Female Vocals

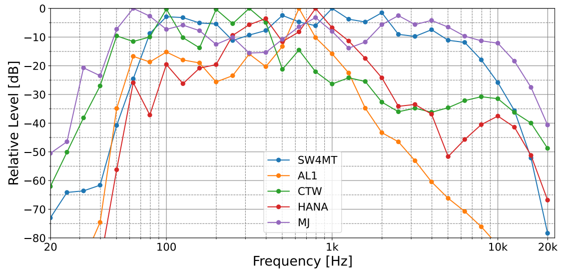


Fig. 3: The average spectra of stimuli

4.2 Results and Analysis

A repeated measures analysis of variance with two factors was conducted, with headphones (4 levels) and stimuli (5 levels) as independent variables within participants and similarity ratings as dependent variables. At a significance level of 5%, the main effect of the headphones ($F(3,54) = 3.159, p < 0.05$), the main effect of the stimuli ($F(4,72) = 6.421, p < 0.01$) were significant, and the interaction ($F(12,216) = 3.516, p < 0.01$) was also significant. On the basis of Tukey’s honestly significant difference test, a significant difference was obtained between headphone B and headphones A and C for stimulus AL1 (B>A, C, both $p < 0.01$), between headphones B and C and headphone D for stimulus HANA (B, C>D, both $p < 0.01$) and between headphone A and headphones B and C (B, C>A, $p < 0.05$, $p < 0.01$, respectively). Furthermore, a significant difference was obtained between stimulus AL1 and stimulus SW4MT for headphone B (AL1>SW4MT, $p < 0.05$) and between stimulus HANA and stimulus AL1 and CTW for headphone D (AL1, CTW>HANA, $p < 0.01$, $p < 0.05$, respectively).

The average similarity ratings are depicted in Fig. 4. The trend that the simulation of a specific headphone is uniformly higher or lower in similarity compared with the simulation of other headphones was not observed. The average similarity was at most 2.42, with the majority less than 2, indicating that the subjective similarity of the virtual headphones to the actual headphones in this experiment was not high. In contrast, there were cases where the similarity varied when the stimulus was different even with the same headphones

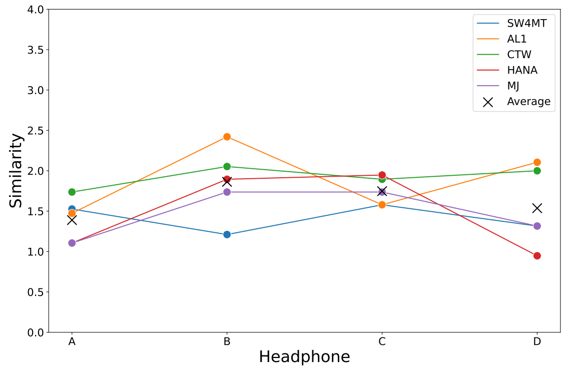


Fig. 4: Mean similarity ratings

or when the headphones were different even with the same stimulus. These results suggested that the actual headphone sound quality can be sufficiently simulated with the virtual headphone configuration used in this study for a specific headphone and sound source combination. Stimulus AL1, which had less energy at a higher frequency, exhibited a higher similarity than the other stimuli in headphones B and D because evaluation using the stimuli—which had more energy at a higher frequency—is affected by the simulation accuracy of the high-frequency bands. The subjective similarity was low when the simulation accuracy of that band was low.

Regarding the evaluation value of spectral balance, the average spectral balance evaluation value of the top five conditions with high subjective similarity and the bottom five conditions are depicted in Fig. 5.

There was no significant difference in the average evaluation value between the top five and bottom five conditions, and there was no significant difference in the perception of the spectral balance difference between the actual and virtual headphones, whether they were relatively highly or lowly evaluated. The difference in spectral balance is not considered an absolute factor in determining the similarity between actual and virtual headphones.

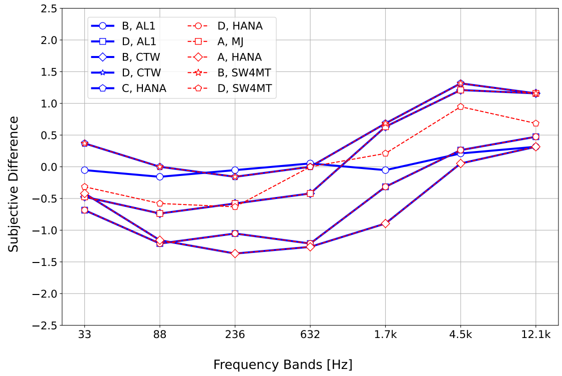


Fig. 5: Mean spectral balance difference ratings (Blue: top five, Red: bottom five)

Table 3: Multiple Regression Analysis Results

	Coefficient	p-value
const	2.6052	$p < 0.01$
33 Hz	0.0064	n.s.
88 Hz	-0.1744	$p < 0.01$
236 Hz	-0.1010	n.s.
632 Hz	-0.1655	$p < 0.01$
1.7 kHz	-0.2092	$p < 0.01$
4.5 kHz	-0.3864	$p < 0.01$
12.1 kHz	-0.0516	n.s.

n.s. : not significant

A multiple regression analysis was conducted to investigate the effect on the band-by-band similarity of the spectral balance. The evaluation value of the similarity was the dependent variable, and the absolute value of the evaluation value of the spectral balance was the independent variable. The Variance Inflation Factor (VIF) was calculated, but multicollinearity between the independent variables was not confirmed (all VIF < 3).

Based on the analysis, the regression coefficients were significant at the 5% significance level in the 88, 632, 1700, and 4500 Hz bands. The regression coefficient of 4.5 kHz was about twice as large as the other significant bands, suggesting that the evaluation value of 4.5 kHz had the most significant impact on similarity. Also, while 12.1 kHz was insignificant, a significant correlation existed between the evaluation value of 4.5 kHz and the evaluation value of 12.1 kHz ($r = 0.63, p < 0.01$). Consequently, the accuracy of high-frequency bands (several kHz and above) had the most significant impact on the similarity between the actual and virtual headphones. Therefore, it is necessary to include bands of 10 kHz and above in the filter correction target to achieve higher simulation accuracy in constructing virtual headphones.

5 Discussion

This study's "Virtual Headphone" construction method resulted in a discrepancy between physical simulated accuracy and subjective similarity. The causes that led to this discrepancy include (1) possibility that the simulation on the ear simulator could not be reproduced during actual wearing and (2) possibility that the sound quality simulation is insufficient with just equalization of amplitude frequency response.

The ear pads of the headphones used in this study were all different, and none were the same size. Therefore, the relationship between the headphones and the ear when worn varies from headphone to headphone. Furthermore, although the pinna of the ear simulator is of average human size, it differs in shape from the ear of each subject. Because these differences interact, the relationship between headphones and ears at the time of measurement could not be reproduced while listening, resulting in a high subjective similarity not being achieved.

Amplitude–frequency responses are one of the primary factors determining sound quality, but there is a possibility that the sound quality of another headphone cannot be reproduced with only equalization of amplitude–frequency response. Other responses, such as phase, transient, and non-linear distortion, may also affect sound quality. However, these characteristics cannot be included when limited to equalization of the amplitude–frequency response. It is unknown how much these responses contribute to the subjective similarity between the actual and virtual headphones. However, ignoring these factors is a reason for lowering the subjective similarity.

This study aimed to examine a method of constructing virtual headphones and to reduce and simplify the number of filter design parameters. Thus, potential complex parameters such as individual correction terms were not introduced. In the future, it will be necessary to introduce additional parameters (e.g., to consider ear shape) for designing filters with higher subjective correction accuracy.

6 Summary

This study examined a method of constructing "Virtual Headphones" that simulate the sound quality of another set of headphones.

Based on the measurement results and equalizing filter design, the virtual headphones can closely simulate the amplitude–frequency response of the actual headphones on the ear simulator. Furthermore, headphones with "few peaks and dips in the amplitude–frequency response, relatively smooth response" are suitable for a reference headphone to simulate another headphone.

Based on the subjective evaluation, when comparing listening through virtual headphones with actual headphones, the sound quality of the virtual and actual headphones was "not unlike, but not exactly the same." Furthermore, the simulation accuracy in high-frequency bands such as several kHz and above has a sufficient effect on subjective similarity.

This study included frequencies above 10 kHz in the filter design, which were not examined in previous studies. The simulation accuracy above 10 kHz also significantly affects the similarity between the virtual and actual headphones. Virtual headphones with no correction above 10 kHz cannot have the same listening impression as listening with actual headphones.

Despite the high physical simulation accuracy, subjective similarity was not very high. Increasing the subjective simulation accuracy of virtual headphones will require introducing parameters that correct individual responses, such as the shape of the ear, in the filter.

Acknowledgments

We thank Philip Pape, Ph.D., from Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript.

References

[1] Olive, S. E., Welti, T., and McMullin, E., "Listener Preferences for Different Headphone Target Response Curves," in *134th AES Convention*, Paper 8867, Audio Engineering Society, 2013.

[2] Olive, S. E., Welti, T., and McMullin, E., "A Virtual Headphone Listening Test Methodology," in *in 51st AES International Conference*, Audio Engineering Society, 2013.

[3] Møller, H., Hammershøi, D., Jensen, C. B., and Sørensen, M. F., "Transfer Characteristics of Headphones Measured on Human Ears," *J. Audio Eng. Soc.*, 43(4), pp. 203–217, 1995.

[4] Bücklein, R., "The Audibility of Frequency Response Irregularities," *J. Audio Eng. Soc.*, 29(3), pp. 126–131, 1981.

[5] Bolaños, J. G., Mäkivirta, A., and Pulkki, V., "Automatic Regularization Parameter for Headphone Transfer Function Inversion," *J. Audio Eng. Soc.*, 64(10), pp. 752–761, 2016.

[6] Iwamiya, S., Nakajima, Y., Ueda, K., Kawahara, K., and Takada, M., "Technical Listening Training: Improvement of sound sensitivity for acoustic engineers and sound designers," *Acoustical Science and Technology*, 24(1), pp. 27–31, 2003.