Audio Engineering Society

Conference Paper 21

Presented at the AES International Conference on Headphone Technology 2025 August 27-29, Espoo, Finland

This conference paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This conference 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. This paper is available in the AES E-Library (http://www.aes.org/e-lib), 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.

Personalized Timbre Optimization Based on a New Auditory Model for Stereophonic Sound Reproduction via Earphones

Kimio Hamasaki[1,2], Nao Tojo[1], Atsushi Hara[1], Haruto Hirai[1], Shun Saito[1], and Mitsuru Hosoo[1]

[1] final Inc.
[2] ARTSRIDGE LLC

Correspondence should be addressed to Kimio Hamasaki (hamasakik@final-inc.com)

ABSTRACT

This paper proposes a novel personalization method for optimizing the natural timbre of stereophonic sound reproduction via earphones. While conventional headphone personalization techniques mainly focus on spatial sound attributes, our approach introduces an auditory model designed to enhance timbre fidelity. The method utilizes high-resolution 3D scans and individualized head-related transfer functions (HRTFs) to derive a personalized target response curve, combining individualized mid-to-high frequency characteristics with generalized low-frequency targets. Listener-specific weighting coefficients are determined through controlled listening tests, enabling fine-tuning without relying on subjective preference alone. Subjective evaluation employing the semantic differential method demonstrates significant improvements in perceived naturalness and overall sound quality when compared to existing general target response curves, including diffuse field and listener-preferred curves. These results validate the effectiveness of the proposed model to achieve more natural timbral reproduction in stereophonic headphone listening.

1 Introduction

Recent advancements in headphone and earphone personalization have significantly improved the quality of spatial audio reproduction. The fundamental concept of individualizing binaural reproduction originates from the pioneering work of Wightman and Kistler [1], who demonstrated that using listener-specific head-related transfer functions (HRTFs) significantly improves the accuracy of virtual auditory space representation over headphones. The psychophysical foundations of human spatial hearing and the importance of individualized cues were established in foundational studies by Blauert [2], providing the basis for modern binaural audio.

In the late 1990s and early 2000s, the development of public HRTF databases, such as the CIPIC database by Algazi et al. [3,4], enabled systematic investigations into the impact of non-individualized versus individualized HRTFs. These studies confirmed that listener-specific HRTFs provide superior localization accuracy and spatial perception compared to generic or averaged sets [5,6]. More recently, machine learning and model-based approaches have facilitated rapid HRTF personalization using anthropometric features or even photographs [7–9]. Such techniques have led to the commercialization of headphone products and binaural rendering engines offering personalized spatial audio to a broader user base.

Despite these substantial advances, most research and commercial solutions in headphone and earphone personalization have remained focused on optimizing spatial attributes, such as localization, externalization, and envelopment, particularly for immersive binaural sound. In contrast, the equally important aspect of natural timbre reproduction, especially for conventional stereophonic sound, has received considerably less attention [10,11]. In this paper, we define "immersive binaural sound" as binaural audio reproducing three-dimensional spatial sound rendered from immersive content created with multichannel loudspeakers. Conversely, "stereophonic sound" refers to sound produced by two-channel loudspeakers and reproduced through headphones or earphones without any binaural rendering. Since most music content is still consumed in stereophonic formats via earphones and headphones, it is essential to develop personalization methods that enhance the perception of natural timbre in this context.

Subtle inaccuracies in timbral fidelity can undermine musicality and the emotional impact of audio, even when spatial attributes are well rendered [10–12]. Numerous long-term studies have investigated listeners' preferences for sound quality and timbre, leading to the development of listener-preferred target response curves for headphones and earphones [13–15]. These curves can be considered general, non-individualized reference responses for improving audio quality. Traditional timbre personalization approaches typically rely on EQ adjustments according to user preference but do not account for individual auditory physiology or the perceptual mechanisms underlying natural timbre perception. Theile [16] explained auditory system behaviour under natural conditions and proposed a distinction between spatial and timbral perceptual cues, particularly Gestalt perception of timbre.

This paper introduces a novel personalization method designed to achieve natural timbre reproduction in stereophonic sound via earphones. Our approach leverages high-resolution anatomical scanning and individualized HRTF modelling to isolate and optimize the spectral cues most relevant to the perception of natural timbre.

2 New auditory model for personalizing timbre perception

This section describes a newly proposed auditory model designed to achieve natural timbre perception when reproducing stereophonic sound via earphones. While previous models primarily focused on spatial attributes, such as localization or externalization, the proposed model aims to optimize spectral cues responsible for timbre fidelity.

The model builds upon Theile's two-stage concept [16,17], where sound perception is formed through a combination of spatial cues and Gestalt recognition cues, which is key to perceiving timbre. Under natural listening conditions, Theile proposed that the human auditory system integrates directionally dependent spectral modifications to form a cohesive perceptual image of sound. In Theile's framework, the transfer function M represents the frequency-dependent transformation caused by the outer ear's geometry, and its inverse M^{-1}, corresponds to removing this spectral shaping effect, creating a theoretically neutral condition before the source signal reaches the Gestalt determining stage. It is achieved in the case where the distribution of incident sound is entirely at random [17]. The Recommendation ITU-R BS.708 specifies that measurements of studio monitor headphones be conducted in the diffuse sound field to consider the conditions mentioned above [18].

On the other hand, when considering the target response curve for headphones and earphones, there is a concept known as "insertion gain", defined as the difference between open ear loudspeaker response and earphone response at the Eardrum Reference Point (DRP) [19]. There are several well-established target responses at DRP, including the diffuse field response [19,20].

Our model combines Thiel's hearing framework and the insertion gain concept to personalize the target response curve at DRP, to achieve natural timbre perception in listening to stereophonic sound through earphones. In our auditory model, transfer function $H(f)$ from a sound source to Ear Reference Point (ERP), which is defined as the plane at the entrance of the ear canal、is decomposed into two distinct components: $P(f)$, which encodes directional cues essential for spatial perception, and another $T(f)$, which encodes spectral features relevant to timbral perception.

$$H(f) = P(f) \cdot T(f) \tag{1}$$

However, in the case of earphone listening, the directional cues in $P(f)$ are eliminated mainly due to the bypassing of outer ear-related spatial encoding. This degradation leads to an unnatural perception of

timbre. To address this issue, we propose neutralizing the directional component by setting $P(f) = 1$, thereby preserving only the timbral component $T(f)$. The central premise of our model is the introduction of a directional power spectrum averaging process of $H(f)$, representing the average power across multiple directions in the horizontal and vertical planes. In this study, we employed 864 directional measurement points based on a uniformly distributed grid, with a horizontal resolution of 5.0 degrees and a vertical resolution of 10.0 degrees. The resolution of 5-degree horizontal and 10-degree vertical increments was chosen as it provides sufficient directional resolution to capture significant spectral variations in typical human head-related transfer functions (HRTFs), without excessive computational complexity. Where N is the total number of directional HRTF measurements used in the averaging process (in this case, $N = 864$). This concept leads to the equation as

$$H(f) = 1 \cdot T(f) = k \cdot \frac{1}{N} \sum_{i=1}^{N} HRTF_i(f) \tag{2}$$

This approach effectively cancels directional encoding effects while retaining the overall spectral energy distribution, which is essential for timbre perception. Furthermore, to account for individual differences in auditory perception and ear anatomy on timbre, we introduce a personalized weighting coefficient k, which enables the natural timbre perception. The personalized weighting coefficient k represents an individual-specific scaling factor derived from controlled listening tests, but not based on the listener's preferences, allowing fine-tuning of the averaged HRTF response to match subjective perceptions of natural timbre. Recent studies by Olive et al. have highlighted listener variability in preferred headphone frequency responses, emphasizing the importance of personalized spectral tuning [13-15]. These findings support the motivation for individualized weighting coefficients in our model.

The Personalized Timbre Target response Curve (PTTC) at DRP for left-ear earphones for natural timbre perception in listening to stereophonic sound through earphones is defined by

$$PTTC^L(f) = k \cdot HRTF_{av}^L(f) + n \cdot TC_{low}(f) + m \cdot C^L(f) \tag{3}$$

Where, $HRTF_{av}$ represents the listener-specific averaged HRTF in the mid-to-high frequency range (800 Hz – 10 kHz), TC_{low} is a non-listener-specific predefined low-frequency target curve in the low frequency range below 800 Hz, C is a listener-specific compensation curve for the individual ear canal, and coefficients k, n, and m are optimized through listening experiments. The frequency range division at 800 Hz is based on psychoacoustic research indicating that below this frequency, spatial cues have minimal directional dependency, allowing the adoption of a generalized low-frequency target curve (TC_{low}), while frequencies above 800 Hz strongly affect individualized timbre perception, necessitating personalized spectral shaping. $PTTC^R$ is defined as in equation 3, where k, n, m, and TC_{low} are the same as left earphones, but $HRTF_{av}^R$ and C^R are specific for right earphones. The effect of C^L and C^R on the timbre perception is relatively minor compared to the other two functions; therefore, this paper will not provide a detailed explanation of this parameter.

3 Personalization method for natural timbre perception

The proposed method for personalizing the target response curve of earphones to achieve natural timbre perception in stereophonic sound reproduction consists of four main stages.

3.1 High-resolution 3D scan of the listener's head

The first stage involves acquiring a high-resolution 3D scan of the listener's head, including the detailed geometry of the pinna. In the second stage, individualized head-related transfer functions (HRTFs) are calculated from the scan data using the boundary element method, covering 864 sound incidence directions. This two-stage process is commonly used in personalization approaches to achieve accurate reproduction of spatial impressions when listening to immersive binaural sound through headphones or earphones.

3.2 Individual target response curve for personalized natural timbre perception

In the third stage, the process of averaging 864 individual HRTFs for the left and right ears, respectively, has been completed, aiming to remove spatial encoding components within the HRTFs, primarily those shaped by the outer ear and

responsible for directional cues. Individual $HRTF_{av}^{L}$ and $HRTF_{av}^{R}$ retains mainly the spectral information essential for the target response curve for natural timbre perception. Figure 1 shows an example of $HRTF_{av}^{L}$ of 10 subjects. Figure 2 shows $HRTF_{av}^{L}$ and $HRTF_{av}^{R}$ of subject 02, which are similar in both the left and right ears. Figure 3 shows $HRTF_{av}^{L}$ and $HRTF_{av}^{R}$ of subject 01, which are different at over 3 kHz in the left and right ears. Difference between $HRTF_{av}^{L}$ and $HRTF_{av}^{R}$ are kept generating the personalized target response curves for earphones.

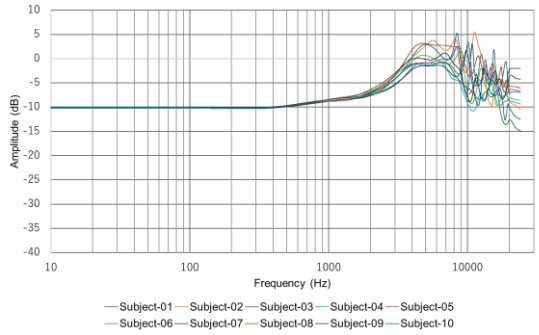


Figure 1. Examples of $HRTF_{av}^{L}$ of 10 subjects.

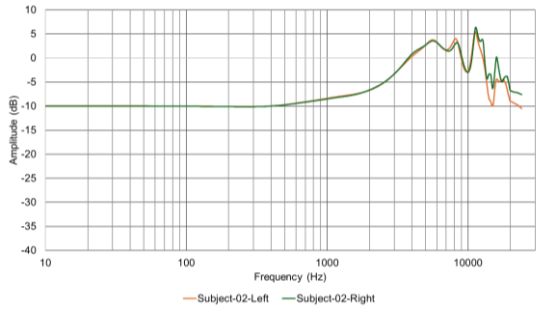


Figure 2. $HRTF_{av}^{L}$ and $HRTF_{av}^{R}$ of Subject 02.

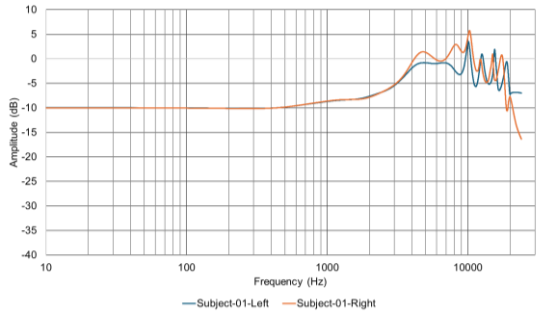


Figure 3. $HRTF_{av}^{L}$ and $HRTF_{av}^{R}$ of Subject 01.

The fourth stage determines two listener-specific weighting coefficients: k (ranging from 0.00 to 1.00, varied in 0.05 steps) , n (expressed in decibels, varied in 1 dB steps), and m (0 to 1) . The coefficient k is applied equally to both listener-specific $HRTF_{av}^{L}$ and $HRTF_{av}^{R}$, and is used to optimize the perceived naturalness of timbre in the mid-to-high frequency range. Figure 4 illustrates the effect of varying k (1.0, 0.8, and 0.5) on $k \cdot HRTF_{av}^{L}$ for Subject 09. In contrast, n serves as a low-frequency gain adjustment applied to a default low-frequency target curve. This dual-coefficient approach enables the separate optimization of spectral cues in the mid-to-high and low-frequency bands, respectively, aiming to individualize natural timbre perception.

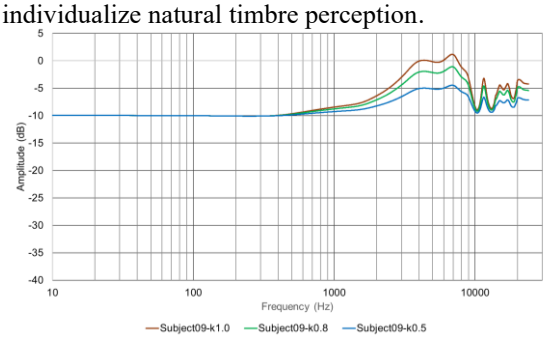


Figure 4. Example of $k \cdot HRTF_{av}^{L}$ of Subject 09 with different k values (1.0, 0.8, and 0.5).

The determination of these coefficients is conducted through controlled listening tests based on the proposed auditory model, rather than relying solely on subjective preference. The listening tests employ four categories of auditory stimuli designed to represent diverse timbral characteristics: (i) dialogue and a cappella vocals, (ii) solo musical instruments, (iii) a four-piece rhythm band without vocals, and (iv) vocals accompanied by a four-piece rhythm band.

Each stimulus is convolved with a listener-specific impulse response, which consists of two components:

- A mid-to-high frequency component (800 Hz to 10 kHz), shaped by $k \cdot HRTF_{av}^{L}$ and $k \cdot HRTF_{av}^{R}$. The upper frequency limit of 10 kHz and lower frequency limit of 800 Hz are determined based on practical considerations for timbre personalization and the characteristics of the listener-specific HRTFs obtained in the first and second stages, as shown in Figure 1.

- A low-frequency component (below 800 Hz), shaped by a default low-frequency target curve designed for the personalization scheme shared across listeners, with its amplitude adjusted by the listener-specific coefficient n (in dB).

The listening test was conducted using a three-alternative forced-choice (3AFC) task, in which participants were asked to select the sound stimulus they perceived as having the most natural timbre among the three options. Each of the three stimuli was processed using different values of the weighting coefficients either k or n. Initially, the differences

among the three values were set to be large (0.3 for k and 3 for n) and were progressively reduced according to a predefined adjustment sequence. For k, the difference was reduced stepwise until it reached a minimum increment of 0.05. Similarly, for n, the difference was reduced until it reached a minimum increment of 1.0 dB. The test procedure was divided into two consecutive stages. First, the optimal value of k was determined by presenting stimuli with varying k values while n was fixed. Once the participant's preferred k value was identified, it was held constant, and the second stage focused on determining n by presenting stimuli with varying n values. Upon completion of this two-step procedure, the individualized optimal values of k and n were obtained for each participant.

The final k and n values are determined as the individualized optimal coefficients upon test completion. The final target response curve for each listener's earphones is computed by combining the k-weighted listener-specific HRTFs for frequencies above 800 Hz with the n-adjusted low-frequency target curve for frequencies below 800 Hz. Figure 5 shows the resulting personalized target response curves for the left earphone of four representative subjects. The distribution of k and n obtained from listening tests conducted on 132 participants is presented in Figures 6 and 7, respectively. The final target response curve obtained through these four stages is implemented as an FIR filter in the earphones' digital signal processing (DSP) pipeline.

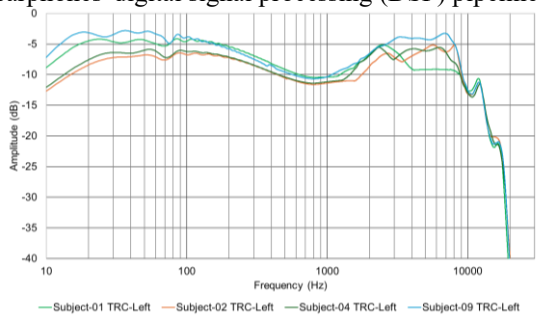


Figure 5. Personalized target response curve for the left earphone of four representative subjects.

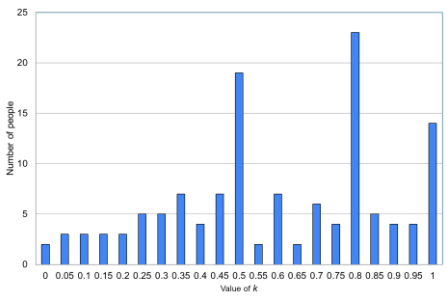


Figure 6. Distribution of coefficient k obtained from listening tests with 132 subjects.

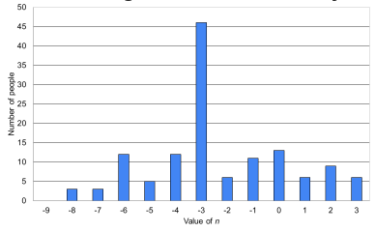


Figure 7. Distribution of coefficient n obtained from listening tests with 132 subjects.

4 Subjective assessment

The objective of this subjective assessment is to investigate how personalized timbre optimization enhances the perception of natural timbre by comparing it against other commonly used general target response curves for earphones available on the market. The Personalized Timbre Target Curve (PTTC), developed in this study, is designed to maximize the perception of natural timbre through individualized optimization.

4.1 Method of subjective evaluation

The Semantic Differential (SD) method [21] was employed for this subjective evaluation. Participants were asked to rate their impressions using a seven-point scale (1-2-3-4-5-6-7) across 36 pairs of bipolar evaluation attributes, as shown in Table 1. Each pair consisted of two opposing descriptors, such as "natural" and "unnatural." For instance, when evaluating naturalness, participants selected a score from 1 to 7, where 1 indicates "very unnatural," 4 is "neutral," and 7 represents "very natural." The seven participants included five students from music universities and two audio professionals.

Balanced	Pleasant	Noise-free	Diffused	Not dull	Bass+
Treble+	Not hissing	Not hard	Not hollow	Open	Clear
Not cracking	Bright	Not feeble	Soft	Not Muddy	Natural
Presence	Clean	Space	Not Jarring	Not too loud	Gentle
Not dry	Thick	Not sharp	Wide band	Transparent	Smooth
Non-fatigue	Rich	Good timbre	Good sound	Enjoyable	Like

Table 1. List of abbreviations of 36 bipolar evaluation attributes used in the SD method.

The subjective evaluation employed four types of auditory stimuli: (i) dialogue and a cappella vocals, (ii) solo musical instruments, (iii) a four-piece rhythm section without vocals, and (iv) vocals accompanied by a four-piece rhythm section. Three target response curves were evaluated: the Personalized Timbre Target Curve (PTTC), proposed in this study and

derived through individualization; the Diffuse Field Target Curve (DFTC), which shares the same low-frequency response (below 800 Hz) as the PTTC [22,23]; and the Listener Preferred Target Curve (LPTC) [15]. Each participant evaluated all four stimuli using TWS earphones for each of the three target response curves. The listening sessions were structured so that participants listened to all four stimuli with one target response curve before proceeding to the remaining two curves. This procedure was repeated weekly for five consecutive weeks. Figure 9 shows the three target response curves for Subject 4.

4.2 Results of subjective evaluation

Figure 8 presents the SD evaluation results, showing the mean scores and 95% confidence intervals for each target response curve across all subjects. Figure 10 summarizes the overall mean scores and 95% confidence intervals, averaged across all subjects and attributes.

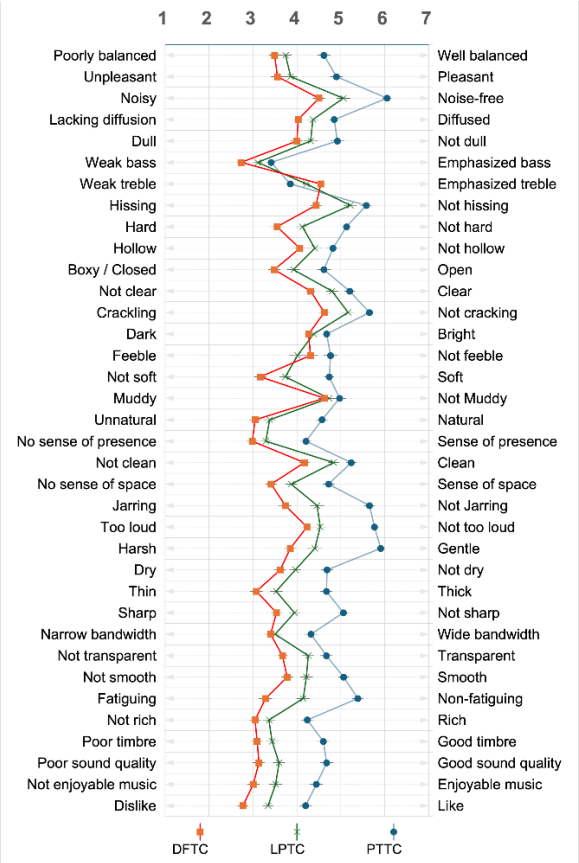


Figure 8. Results of the SD evaluation showing mean subjective rating and 95% confidence intervals for each target response curve.

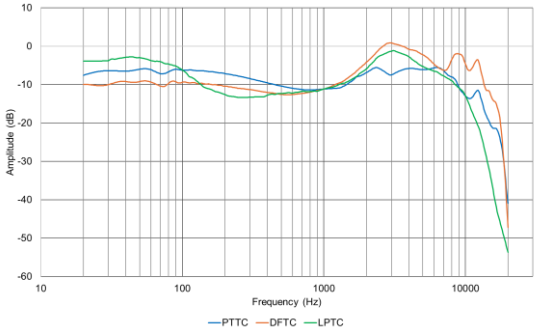


Figure 9. Examples of target response curves for Subject 04 (PTFC, DFTC, and LPTC), applied to the subjective experiment.

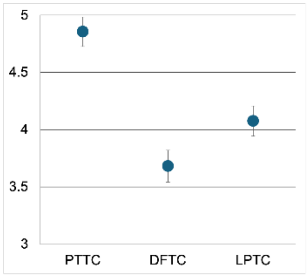


Figure 10. Overall mean subjective ratings and 95% confidence intervals across all subjects and attributes for each target response curve.

Principal Component Analysis (PCA) extracted two principal components, accounting for 50.01% and 9.16% of the total variance, respectively, cumulatively explaining 59.2% of the total variance (Table 2). The first principal component (PC1) was identified as the primary dimension reflecting overall sound quality perception, while the second principal component (PC2) captured specific timbral characteristics. PC1 exhibited high positive loadings for attributes such as "Good timbre" (0.915), "Like" (0.908), "Good sound" (0.899), "Natural" (0.898), "Enjoyable" (0.891), "Rich" (0.878), "Non-fatigue" (0.870), "Pleasant" (0.865), "Not jarring" (0.846), "Wide band" (0.836), "Balanced" (0.827), and "Presence" (0.809), indicating a strong correlation with overall favourability and sound quality. PC2 showed high positive loadings for "Treble+" (0.723), "Diffused" (0.730), "Not dry" (0.627), and "Open" (0.598), and negative loadings for "Gentle" (-0.574), "Soft" (-0.528), and "Not sharp" (-0.612), reflecting brightness, diffuseness, treble emphasis, and perceived harshness. Figure 11 illustrates PCA plots to compare PTTC, LPTC, and DFTC on PC1 (overall sound quality) and PC2 (timbre characteristics).

Component	SS Loadings	% of Variance	Cumulative %
1	18.00	50.01	50.0
2	3.30	9.16	59.2
3	2.41	6.71	65.9
4	1.58	4.38	70.3

Table 2. Summary of principal component analysis results from the SD evaluation.

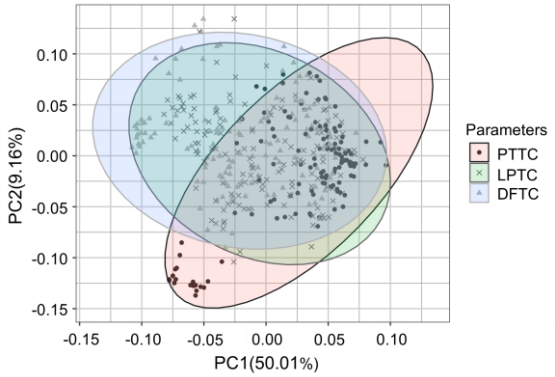


Figure 11. PCA plots comparing PTTC, LPTC, and DFTC along PC1 (overall sound quality) and PC2 (timbre characteristics).

Non-parametric analysis of variance using the Kruskal–Wallis test revealed statistically significant differences among the three target curves for all attributes, except "Bright" ($p = 0.060$) and "Not feeble" ($p = 0.051$), with $p < 0.001$ for most attributes. According to common guidelines, ϵ^2 values above 0.14 are considered large effect sizes. The largest effect sizes (ϵ^2) were observed for "Gentle" ($\epsilon^2 = 0.2072$), "Non-fatigue" ($\epsilon^2 = 0.1909$), "Not jarring" ($\epsilon^2 = 0.1652$), "Not sharp" ($\epsilon^2 = 0.1513$), "Not hard" ($\epsilon^2 = 0.1470$), and "Soft" ($\epsilon^2 = 0.1425$).

Post-hoc pairwise comparisons using the Dwass–Steel–Critchlow–Fligner method further confirmed these differences. PTTC consistently demonstrated statistically significant improvements over both DFTC and LPTC (often $p < 0.001$), receiving higher scores in favourable attributes.

The subjective assessment demonstrates that PTTC consistently yields higher ratings in overall sound quality and listener preference compared to DFTC and LPTC. These findings align with PC1 loadings, reflecting general sound quality. While DFTC and LPTC showed similar ratings in some attributes, LPTC outperformed DFTC in several aspects related to comfort and clarity, such as "Noise-free," "Not hissing," "Not hard," "Soft," "Clean," "Transparent," "Non-fatigue," "Enjoyable," and "Like."

5 Discussions

The subjective evaluation results consistently show that the proposed Personalized Timbre Target Curve (PTTC) offers significant perceptual benefits over conventional diffuse field target curve (DFTC) and listener-preferred target curve (LPTC). The improvements are particularly evident in attributes associated with naturalness, richness, and reduced listening fatigue, as reflected in higher scores across key semantic differential scales. Principal Component Analysis (PCA) further indicates that PTTC optimizes the primary dimension related to overall sound quality while also improving secondary dimensions associated with timbre characteristics. The high degree of statistical significance and large effect sizes support the robustness of the findings. These results confirm the effectiveness of our model, especially for music listening contexts where stereophonic sound reproduction remains dominant. Importantly, our approach achieves personalization without requiring subjective user adjustment, relying instead on physiological features and perceptual modeling, which may enhance user accessibility and consistency in real-world headphone applications.

6 Conclusions

This study introduced a novel auditory model and personalization method aimed at optimizing natural timbre in stereophonic sound reproduction via earphones. By integrating high-resolution anatomical scanning, individualized HRTF modeling, and perceptually guided weighting coefficients, the proposed method effectively compensates for timbral distortions caused by earphone use. Controlled subjective evaluations confirm that the proposed personalized target response curve (PTTC) significantly enhances the perceived naturalness and overall sound quality compared to conventional target curves. The findings highlight the importance of considering timbre-specific personalization in addition to spatial factors, contributing to improved audio experiences in everyday headphone listening.

References

[1] F. L. Wightman and D. J. Kistler, “Headphone simulation of free-field listening. I: Stimulus synthesis,” *J. Acoust. Soc. Am.*, vol. 85, no. 2, pp. 858–867 (1989).

[2] J. Blauert, *Spatial Hearing*: *The Psychophysics of Human Sound Localization,* Revised ed., MIT Press (1997).

[3] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 99–102 (2001).

[4] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," *J. Acoust. Soc. Am.*, vol. 111, no. 5, pp. 2097–2098 (2002).

[5] J. C. Middlebrooks, "Virtual localization improved by scaling non-individualized external-ear transfer functions," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1493–1510 (1999).

[6] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, "Head-related transfer functions of human subjects," *J. Audio Eng. Soc.*, vol. 43, no. 5, pp. 300–321 (1995).

[7] H. Ziegelwanger and P. Majdak, "Modeling the direction-continuous time-of-arrival in head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 135, no. 3, pp. 1278–1293 (2014).

[8] J. Ravindran, Z. Li, and D. Wang, "Personalized binaural audio using machine learning," *IEEE Signal Process. Mag.*, vol. 38, no. 4, pp. 78–89 (2021).

[9] J. Mackenzie, S. Favrot, C. Faller, J. Lee, and B. Shinn-Cunningham, "Personalized binaural rendering: Past, present, and future," *J. Audio Eng. Soc.*, vol. 69, no. 7/8, pp. 540–555 (2021).

[10] T. Lokki, V. Pulkki, M. Karjalainen, and L. Savioja, "Creating audibly convincing virtual acoustics and 3D audio: Past, present, and future," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 18–28 (2011).

[11] F. Rumsey, *Spatial Audio*, Focal Press (2001).

[12] B. C. J. Moore and B. R. Glasberg, "Modeling binaural loudness," *J. Acoust. Soc. Am.*, vol. 121, no. 3, pp. 1604–1612 (2007).

[13] S. E. Olive, T. Welti, and E. McMullin, "Listener Preference for In-Room Loudspeaker and Headphone Target Responses," presented at the AES Convention, New York, USA, Paper 8994, (2013).

[14] S. E. Olive and T. Welti, "Factors that Influence Listeners' Preferred Bass and Treble Balance in Headphones," presented at the *AES Convention,* New York, USA, Paper 9382, (2015).

[15] S. E. Olive, T. Welti, and O. Khonsaripour, "The Preferred Low Frequency Response of In-Ear Headphones," presented at the *AES Conference on Headphone Technology*, Aalborg, Denmark, (2016).

[16] G. Theile, "Equalization of studio monitor headphones," presented at the *AES Conf. on Headphone Technology*, Aalborg, Denmark (2016).

[17] G. Theile, "On the Standardization of the Frequency Response of High-Quality Studio Headphones," *J. Audio Eng. Soc.*, vol. 34, no. 12, pp. 956–969 (1986).

[18] *ITU-R BS.708*, "Determination of the electro-acoustical properties of studio monitor headphones," (1990).

[19] C.J. Struck, S. Temme, "Headphone Response: Target Equalization Trade-offs and Limitations," *AES convention* paper 9374, New York, USA (2015).

[20] C.J. Struck, "Free Plus Diffuese Sound Field Target Earphone Response Derived From Classical Room Acoustics Theory," *AES convention* paper 8993, New York, USA (2013).

[21] C.E. Osgood, G.J. Suci, P.H. Tannenbaum, *The measurement of meaning*, University of Illinois Press, Urbana (1957).

[22] *IEEE Standard 1652-2016*, "for translating head and torso simulator measurements from eardrum to other acoustic reference points." (2016).

[23] *IEC/TS 60318-7*, "Electroacoustics – simulators of human head and ear – Part 7: Head and torso simulator for the measurement of hearing aids." Edition 1.0, (2011).