

# Research on headphone perceptual sound field enlargement technique

Enlarging Headphone Sound Fields for AR/VR

Hao Zhu

Physics Department, Baylor University, hao\_zhu1@baylor.edu

Liu Yang

Physics Department, Baylor University, yang\_liu1@baylor.edu

Haoxuan Yu

School of Building Engineering (ETS Edificación), Universidad Politécnica de Madrid, yuhaoxuan2001@gmail.com

As AR/VR rises, the demand for larger sound fields drives ongoing research and industrial manufactures in headphone soundscapes. In contrast, a systematic review of these studies and products is needed to help the headphone sound field designing in meeting AR/VR requirements. In this work, the principle of perception of the sound field is first explained, and then several headphones are tested and classified to form a review of enlarging sound field design. Utilized by both target frequency response curves and manufactured periods, these headphones are classified into three categories: open-front chamber design, large-closed-front chamber design, and computational virtual surrounding design. Upon our test, the open-front design showed great acoustic superiority over the large-close-front although many studies and products have proposed to implement the large-close-front design. Despite the application of psycho-acoustic, this review did not find computational virtual surrounding design significantly better than the other two. In conclusion, this review featured a combination of open-front design and virtual surrounding algorithms using neural networks.

## INTRODUCTION

Since the invention of headphones by Nathaniel Baldwin in the 1920s, it soon became a trend to use musical reproduction for the superiority of economic and privacy reasons [0]. After the Walkman emerged and gained popularity, headphones diffused worldwide at an incredible velocity for their portable advancement [0-0].

Early headphone developments focused on enhancing the extension of bass and treble due to the poor performance of the transducer [4]. Sennheiser and AKG found that the bass response will be raised and extended deeper with a closed rear chamber attached to the back of the transducer. Therefore, a typical construction of headphones was proposed simultaneously, which contained a treble-preferred transducer, a rear chamber to extend the bass, and a conducting wire [5,6,7]. This design gained great success in the market for good reasons like clear audio quality, low price, and easy-to-repair parts design. Due to the popular music genre, the bass was preferred at the time which the development of headphones was driven by [8,9,10]. John. C. Koss introduced the modern design of headphones in 1958 by enlarging the rear chamber and replacing the coupling foam pad with a closed-covered leather pad [11]. This closed-back design took the world by storm and set the standard for modern headphone design with the consideration of the front chamber such that the headphones could enjoy a sense of space and become an option in the high-end audio market [12,13].

With the evolution of the material and the fabrication of the transducer diagram, bandwidth is no longer a problem of headphone reproduction, and thus a new challenge arises in headphone designing. Despite the bass-heaving design, a natural tone with a clear sound field was targeted in this era with the proponent of the perceptual sound field. This design concept is aimed at simulating the sensing on the physical sound field via headphones with several factors [14,15]. Moller et al. defined the acoustic impedance of acoustic wave pressure over the wave velocity to present the sound resulting that

can be imitated from the circuit [16]. This brings a standard to evaluate the “natural” of a headphone if the acoustic impedance of a headphone is close to the acoustic impedance of the air. Under this standard, closed-back headphones will hardly reproduce a natural sound since the closed design brings a high impedance inside the chamber. To survey the purpose of the new challenge, an open-back design was introduced such that the headphone acoustic impedance is close to the air propagation, enabling the user to experience a natural wide sound field via headphone reproduction, respectively [17,18,19].

Frequency response plays an important role in gaining a more natural sound. Early studies measured the frequency response of free field (FF) in an anechoic chamber that minimized the reflection and soon became the first standard for the headphone monitor [20]. However, the FF frequency response curve did not take room response into consideration and thus would not gain advantages in space sensing. To improve the feeling of space, diffusion field (DF) curve was proposed in the 1980s [21]. DF curve was measured in a room with little absorption such that the observer points always received random reflection and reverberation from all directions. Another target response proposed by JBL is called the room gained curve [22,23]. This curve targeted the relationship between bass gain and room size. Very recent studies suggest that measuring the head-related transfer function (HRTF) to customize the compensation for individual persons [5,24,25]. Such a compensation can have a transparent sound. In some literature, another approach was proposed that measures the transfer function from the external ear to the reference point at the eardrum and compensates for this. However, as pointed out by Frederic L. Wightman and Doris J. Kistler, such compensation reproduced a finer free-field sensation [26].

Recent developments focused on reverberant field reproduction. This development is based on the control theory point of view that when the headphone’s transfer function is identical to a room response, the user will have a similar perception. By acoustic designing and DSP controlling, modern headphones can have a sound field close to a near-field monitor [27,28]. AR/VR is a rising industry. Future headphones will need a much larger sound field to correlate the virtual sound source to the vision. Here, a review of techniques that help to enlarge the headphone sound field is needed for future headphone development [29,30].

## **PERCEPTUAL SOUND FIELD RECONSTRUCTION**

Perceptual sound field reconstruction is critical for designing headphones. The theory of reconstructing a sound field in a perceptual way has been intensively studied and can be divided into two parts: sound localization and space sensing [31]. Sound localization refers to how the human ear system works to identify the direction and distance of a certain sound source. Space sensing presents how the human auditory system feels the size of the acoustic space. In this section, these two concepts are reviewed to evaluate the technique for enlarging the perceptual sound field of headphones [32].

Sound localization contains two main types of psychoacoustic factors: bi-aural cues and mono-aural cues. A simple model of the bi-aural cue can be present as inter-aural cues consisting of inter-aural level difference (ILD) and inter-aural time difference (ITD) [33,34,35]. The ILD describes a sound level difference between left and right ears while the ITD measures the arrival time differences between them. In the language of geometry, these two cues help to reconstruct both amplitude and phase angle that determine a vector pointing at the sound source. Studies have pointed out that a matching pair of ILD-ITD in the frequency domain is the key to obtaining a clear direction of the sound source point. Therefore, a concept to describe how the ILD-ITD pair localizes the sound source is proposed as the Inter-aural Cross-Correlation (ICC) [36,37]. The ICC measures the cross-correlation coefficient between the two ears in both amplitude and phase at the frequency domain and thus can predict the sound imaging, respectively [38].

The mono-aural cue presents a modification from the outer ear and the ear canal. The shape of the outer ear acts as an angle-dependent filter and modifies the frequency response based on the incident angle of the acoustic wave. Thus, this

angle-modified frequency-dependent ILD-ITD pair enabled a high-accuracy sound localization [39]. However, due to the shape differing on each person, the Head-Related Transfer Function (HRTF) was proposed to measure the individual audience's transfer function for a better sound localization reconstruction. The HRTF studies consider the outer ear shape, head size, shoulder width, etc., and thus can reconstruct the angle-dependent filter for the individual audience. Therefore, applications based on HRTF have obtained superiority compared to other mechanisms [40,41,42].

Since reverberant space is the most common space, understanding the reverberant space could extract the key to space sensing [43]. Reverberant space has sound reflection, diffusion, reverberation, etc. with the frequency response called the DF curve. If a sound source were put in a reverberant space, the audience would first hear the directly delivered sound, which is called the direct sound or dry sound [44]. With a hard boundary, the acoustic wave would be reflected and delivered delayed to the audience [45,46]. The delay time is the first impression of the acoustic space size, longer delay time indicates a larger room space in the reverberant field. The decay rate defined by the magnitude ratio of direct sound and early reflection also affects the room sensing. Although irrelevant to the room size, this ratio characterizes the wall's material and helps identify the scenario [47]. Cross-feed is critical for room sensing. Originally cross-feed is applied to the stereo reproduction case in Figure 1A. When two speakers work together, one may have that ear from one side of the head receiving the signal from the other channel. i.e., the left ear hears the sound from the right speaker. In the reverberant field situation like Figure 1B, cross-feed can be used when applying the method of image charge. This method treats the reflection sound as a virtual source mirrored by the boundary and thus, a sound source on the left can create a virtual source beyond the right boundary and obtain the cross-feed to enhance the space characterization.

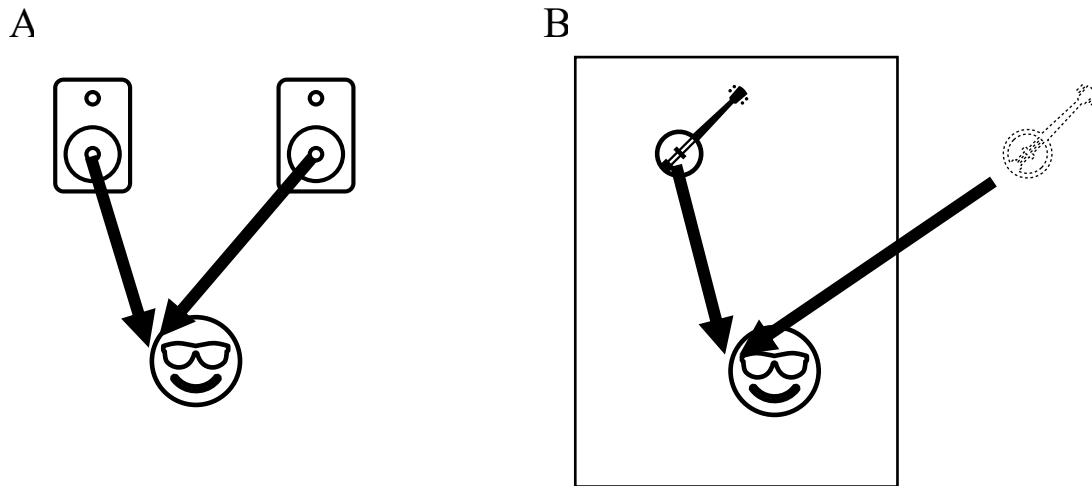


Figure 1. A: cross-feed **FROM STEREO SPEAR SETUP**; B" **CROSS-FEED FROM** image sound

Reverberation is the directionless echo that happens after reflection. To reconstruct the reverberation, one may need the magnitude and the relaxation time. In the spectral studies, the relaxation time is linked to the spatial resolution that the math description follows the form of  $\sin(\omega * t) * \exp(-t/\tau)$  where the  $\omega$  characterizes the excitation frequency and  $\tau$  for the relaxation time. If the on-off test is applied to the acoustic system, the  $\omega$  then becomes  $F_0$  which is the low cut-off frequency of the system [48]. Applying the Fourier transform to the math expression, one may find a Gaussian peak broadening in which the  $\tau$  is exactly full wave half maximum (FWHM) and hence affects the spatial resolution. Like the IACC, there's

an Inter-aural Cross-coherence (IAC) that measures the difference between left and right ears on the perception of the reverberation. Studies have shown that IAC can help to determine the distance between the audience and the sound source in a way that the more similar the left-to-right reverberation is, the closer the sound source. According to the literature, cross-feed can be related to the room's width and image of the sound while the reverberation can be linked to the room's depth [20].

## THE OPEN-FRONT DESIGN

AKG K1000(KK) has gained a lot of commendation due to its open-front design. By canceling the front ear cup, the KK achieved the free-field coupling that enables an extremely large sound stage. An interesting design of KK is the incident angle could be changed up to  $67.2^\circ$  by introducing an axis in the frame [45,46]. This enables a simulation of a near-field loudspeaker situation regardless of the room-sensing application. Figure 2B presents the FR of KK in terms of the incident angle measured by miniDSP's dummy head. According to the point view of the HRTF, changing the incident angle varies the transfer function modification, however, the dummy head measurements at ear-drum-reference did not show obvious change after the HRTF modulation. Instead, distance increment-induced reduction can explain the treble reduction. One should notice

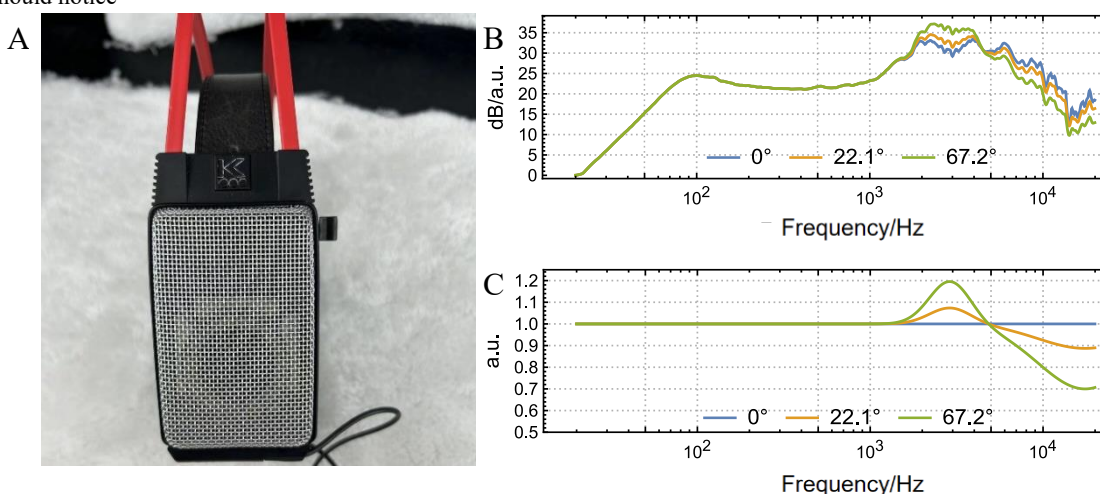


Figure 2. A: AKG K1000; B: frequency response on the open-angle; C: smoothed frequency response normalized to the  $0^\circ$ .

that the 5000 Hz peak indeed came from the dummy head setup while the 3000 Hz peak should be signed as the free-field characterization peak. This indicates that the HRTF modification is not the major factor for the perceptual sound field variety by changing the incident angle. Practices have been done to find the best incident angle and many reports support that the optimal incident angle is about  $22^\circ$  that the driver unit points directly to the outer ear, minimizing the reflection and resonance at the ear canal entrance. One hypothesis is that the human auditory system might be sensitive to the average pressure level and the flow, which requires an in-depth experimental study.

Advanced room simulation can be done by simply adding cross-feed and reverberance as Figure 4. Approximately 20-30% cross-feed and about 1:0.4 for the dry: wet sound ratio (KK max incident angle) will simulate a 4-ince near-field monitor regardless of sound color. Of the best incident angle, about 15% cross-feed with a dry-sound-only setup can reproduce a

boundless sound. Crossfeed also improves the experience of a larger angle since the larger the angle, the wider the horizontal but shorter the lateral perceptual sound field. Crossfeed can elongate the lateral and therefore realize a more natural sound field sensing.

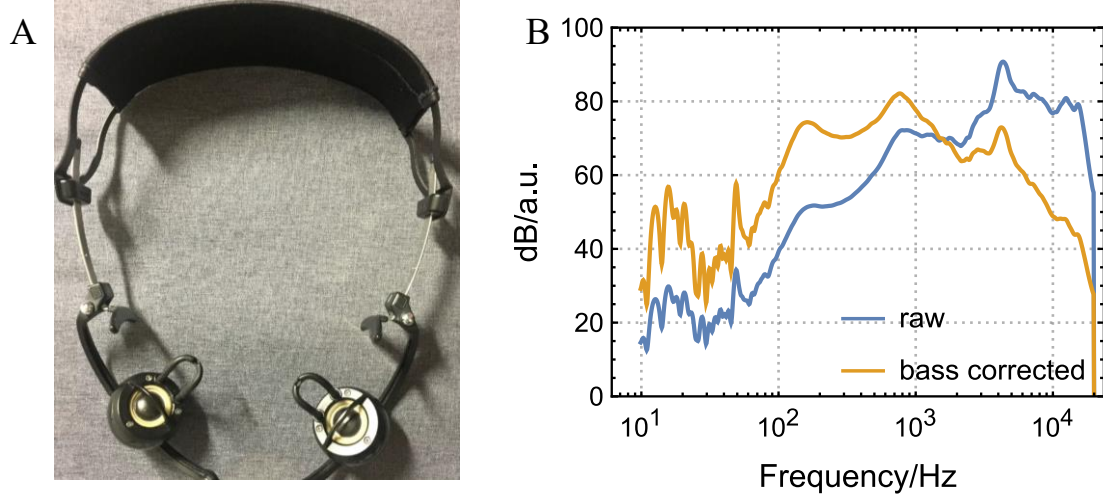


Figure 4. A: Sony PFR-V1; B: raw frequency response (blue) and bass corrected frequency response (orange).

Apart from KK, Sony also produced some impact on open-front designing. PFR-V1 is the model that shares a similar idea to KK with a fixed, unusual incident angle of approximately  $67.5^\circ$ . With the “PFR” stands for “Personal Filed Speaker”, PFR-V1 uses a one-inch dome as a full-range dynamic acoustic unit with a bass waveguide to extend the low response [0]. However, the raw FR shows a serious bass leaking issue compared to the KK shown as the blue curve in Figure 4B. Equalizing corrections are applied and shown in Figure 4B as the orange curve, with ignorance of the harmonic distortion problem, respectively. It can be noticed that even with the bass correction, the cut-off is still high at around 150 Hz which is due to the high  $F_0$  of the dome-type dynamic driving unit. The 100—1000 Hz band shows a similar trend to KK which targets an FF (free field) coupling but with bass lifting. This free-field response similarity enables V1 to share the same room simulation parameters, however, due to the leaking, the audience experience is not as incredible as AKG KK with a much smaller perceptual sound field and the unusual displacement angle only shifts the perceptual sound field to the front a bit. Nevertheless, the audience can always claim an advance in the surrounding sound and apply Dolby or DTS to gain an immersive experience which will be reviewed in the later chapter.

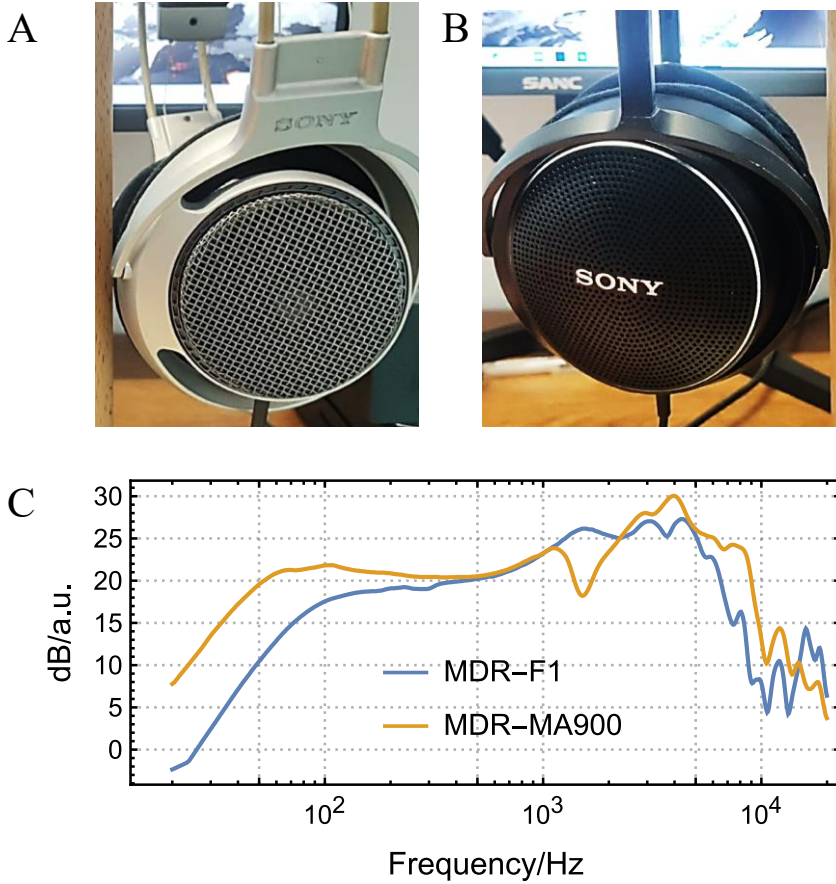


Figure 5. A: Sony MDR-F1; B: Sony MDR-MA900; C: frequency response of F1 and MA900.

To solve this leaking issue, semi-open-front design was introduced as MDR-F1 and the successor MDR-MA900. Like PFR-V1, MDR-F1 also uses a fixed angle but with approximately 20° only, which is close to the angle study of KK [49]. A large bio-dynamic driving unit with a wider band is used to avoid the high cut-off, as shown in Figure 5C. One can easily correct it to a flat band such that the reproduced amplitude distribution remains the same. Here, the semi-open-front is named because the front resonance chamber only has part of it exposed to the air. Better performance of FR can be found in MA900 because of the less-opening design. Reducing the degree of openness made MA900 a more traditional headphone but with a much larger bass-leaking threshold. Hence, although with a boundless sensing of the perceptual sound field, the MA900 still sounds like a traditional headphone.

### THE LARGE-CLOSE-FRONT DESIGN

Modern headphone design prefers the large-close-front. Although open-front design can provide a large perceptual sound field, disadvantages arise from the design like interference from the environment [50]. Early designs of this type can be

traced back to the 1980s with a famous example of Stax SR-Sigma (Figure 6A). The schematic of Sigma is presented in Figure 6B as a subplot. Stax placed the electric driving unit in front of the ear canal which is like the maximum opened KK but with the front closed. Physical reflection and reverberance are introduced by the design and controlled by the absorbing foam inside the chamber. Therefore, this type of design naturally simulates a room by pairing headphone impulse-response (HpIR) to a room impulse-response (RIR). Sigma's FR is presented in Figure 6B which shows that the overall trend follows the room static curve with the characteristic resonance at about 60 Hz. This characteristic implies that the target room is at least the size of 5.72-by-5.72 square meters. However, the listening test did not suggest that this large room target was achieved since the sensing of Sigma's perceptual sound field is smaller than expected. Unlike PFR-V1 still has sound localization from the back, all of the localization by Sigma comes from the front but hard to distinguish the front-and-rear distance between the drum and the vocal, like squeezed into the same plane, making the sensed sound distributed like an ellipsoid. Other than the worse performance on the longitudinal perceptual sound field, the lateral is not large enough either. It could be because of the reflection and reverberance which also lower the resolution. Also, the distance between the driving unit and the ear canal entrance can affect the room size sensing that the closer the driver, the smaller the size.

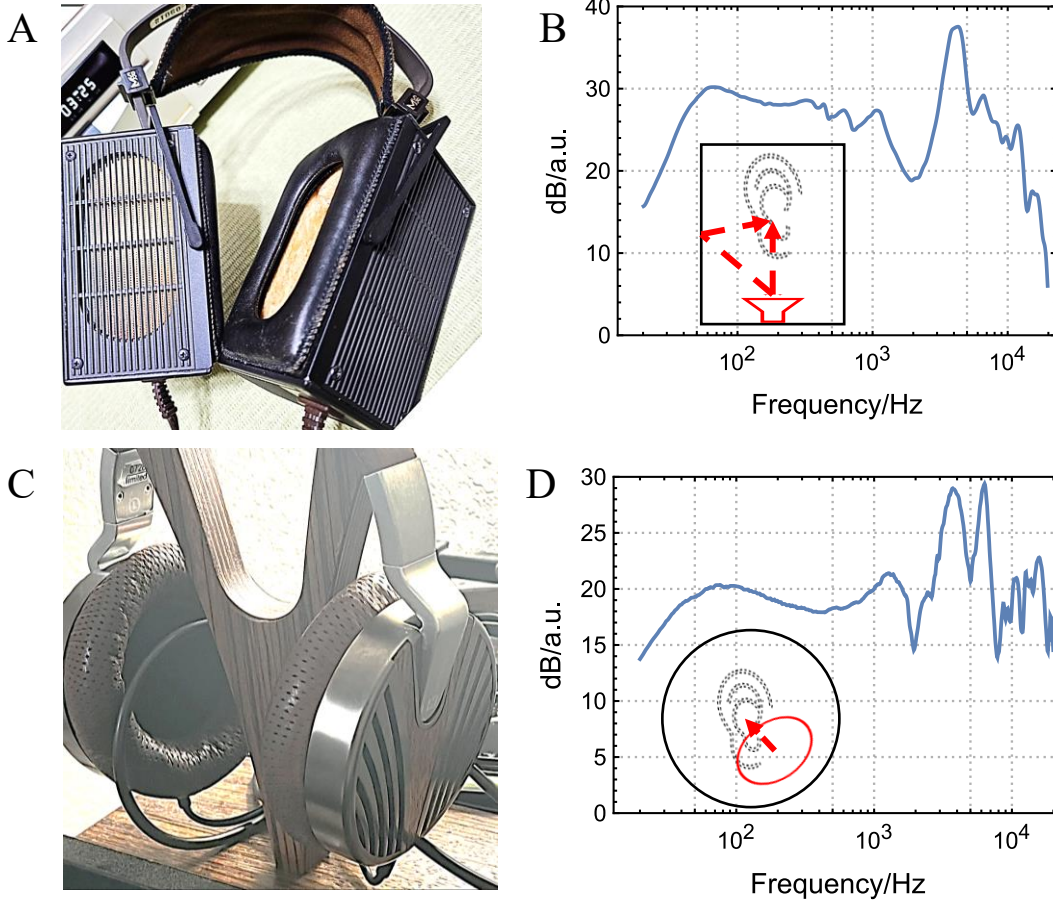


Figure 6. A: Stax SR-Sigma; B: frequency response of Sigma with subplot shows the schematic; C: Ultrasonic Edition 10; D: frequency response of Edition 10 with subplot shows the schematic.

Similar structures could be found in Ultrasonic S-logic techniques. As shown in Figure 6D, the S-logic technique placed the dynamic driving unit asymmetrically but pointed to the ear canal entrance. This design followed the idea of Stax Sigma but modern in terms of unit displacement. Unlike Sigma's extraordinary case, S-logic tuned the displacement angle and the leather pads to balance the reflection, reverberance, and direct sound such that a closer imitation could be present. This work tested the headphone model Edition 10 (Figure 6C) to review the S-logic technique. Figure 6D presents the FR of Edition 10. Like the STAX Sigma, Edition 10 follows the room static response curve with characteristic resonance at around 50- 60 Hz, and consequently, the perceptual sound field is not large either. Improvement has been introduced via the structural advance: tilted displacement would not squeeze the sound field as the front displacement and the clarity is undoubtedly enhanced due to the shorter reverberate relaxation time. In all, neither extraordinary attempt met the target of reproducing a room feeling using headphones. Still, such attempts overcome the in-head localization effect and are followed by modern headphone designs.

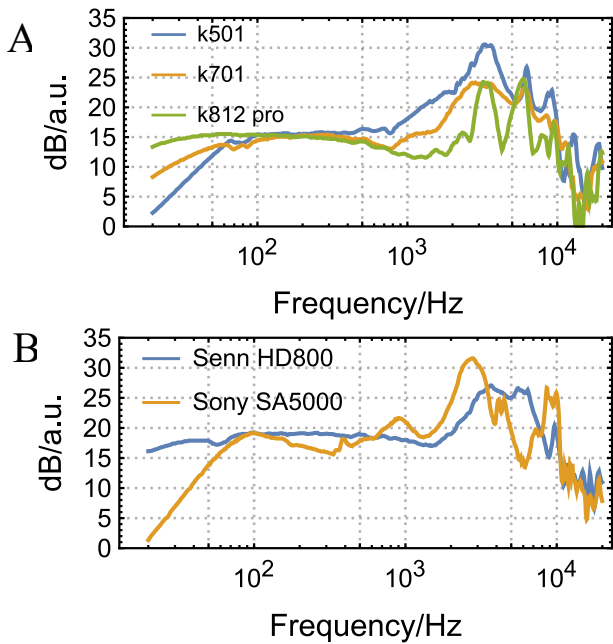


Figure 7. A: Frequency response of AKG K501, K701, and K812 pro;  
B: Frequency response of Sony MRD-SA5000 and Sennheiser HD 800.

Figure 7A shows the FR of three generations of AKG headphones which reflect the modern headphone evolution. All three generations use a large closed-front design but with a certain percentage of leakage to reduce the unwanted reverberance for clarity. Early products like K501 share the target FR with KK (free field) and hence are considered another KK or small KK. The K501 used a fabricated front chamber and earpads such that the restriction and reflection at the boundary were minimized. Evolution from K501 to K701 has been done by replacing the earpads with valve material and using a denser front chamber design to lift the bass response. In consideration of higher reflection, it's reasonable to change the target curve to diffusion field response. Modern products like K812 pro do not prefer a single target, they use a combination of FF and DF, or DF and room static curve which can also be seen in Figure 7B. Both Sony MDR-SA50000 and Sennheiser HD 800 place their dynamic drivers at approximately 10°. The distinctive FR of these two should be analyzed on the designer's philosophy. The SA5000 was introduced with SACD and stressed the treble reproduction quality, therefore the dynamic driver was equipped with a light coil and thin membrane which introduced characteristic peaks at around 4000 Hz and 10, 000 Hz. The designer preferred the speaker performance in an anechoic chamber hence the FR trend at a combination of FF and room static curve. On the other hand, Sennheiser's designer preferred the chamber room for classical music and chose the combination of FF and DF to present a realistic feeling when designing HD 800. In contrast to SA5000, HD 800 uses an extremely large ring radiator (56 mm in size) which focuses on the bass reproduction quality and therefore less peak at treble, only a broadened 3000 Hz [0]. All five headphones here can reproduce a wide range of perceptual sound fields comparable to a 3-inch near-field monitor and become a headphone design guideline. A very recent commercialized design of Drop OAE1 in Figure 8 shows a different case. This model also proposed a large, tilted angle displacement (Figure 8B) with the FR shown in Figure 8C to simulate a front-coming diffusion field. The

author claimed that this design can imitate a neutral feeling in a real acoustic environment. Although with a large degree of tilted angle (about  $60^\circ$ ) as in Figure 8B, OAE1 does not show any superiority in the perception of the sound field whose reason is complex. One possibility is the reverberation time.

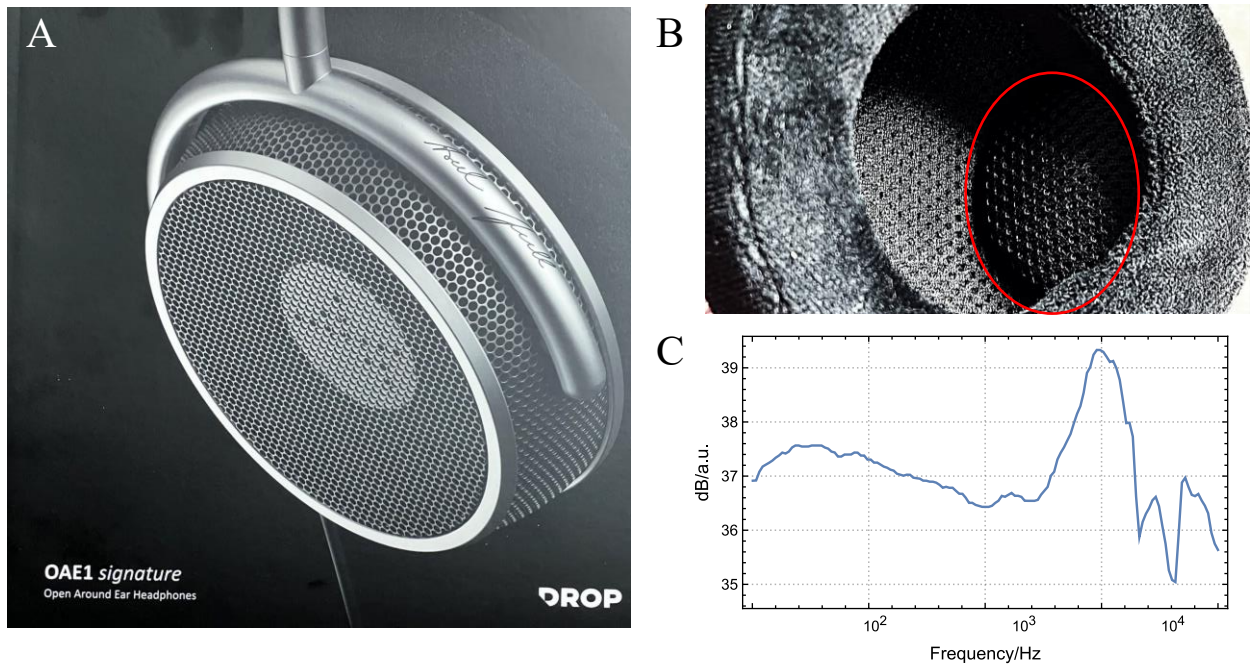


Figure 8. A: Drop OAE1; B: the image of tilted angle (red circle); C: frequency response of OAE1.

As followed by the design guideline of Sennheiser HD 650, OAE1 performs a longer but uniform reverberation than other headphones. Longer reverberation time leads to a narrow sound field perception in HD 650 and similarly in OAE1, nevertheless, according to the lessons from AKG KK and STAX Sigma, larger tilted angle displacement can introduce a wider horizontal but narrower lateral sound field which gratefully fits the drawback of HD 650 and successfully realize a circular perceptual sound field, and consequently, the in-head localization becomes a problem of OAE1 in contrast to other modern headphones. Another possible reason arises from the distance between the driving unit and the ear canal, like the STAX Sigma [51].

#### Virtual 3D surrounding technology

After digital audio technologies were introduced to audio mixing and reproducing, virtual surrounding technology became popular with people for enjoying music and movies in 3D sensing [13]. Early virtual surrounding technology is based on vector-based amplitude panning (VBAP), which is the same as Dolby and DTS's 5.1/7.1 surrounding technology. The idea is simple: by imitating the acoustic environment at a surrounding theater, the virtual surrounding experience can approach the real one using headphones. Early surrounding headphones like AKG K290 (Figure 9A) and the ASUS ROG theta 7.1

keep the physical channel and  
A B



Figure 9. Virtual surrounding headphones A: AKG K290; B: ASUS ROG theta 7.1; C: Pioneer DRS-3000C; D: Sony DS-8000.

surrounding work. This design guarantees a directional sound coming from the side and back, but cons arise as well: interference and reverberation destroy the experience of surroundings [18].

Pioneer and Sony proposed improvement in 2000, with Dolby and DTS cooperating in virtual surround technology. As shown in Figures 9C and 9D, the proposed system uses a single driver unit on each side of the headphone and creates a 5.1 virtual surround sound with the mechanism in Figure 10A. One may note that it's not always necessary to have HRTF modification, in fact, the Sony DS-8000 system does not apply the HRTF [52]. Owing to the same open-front design

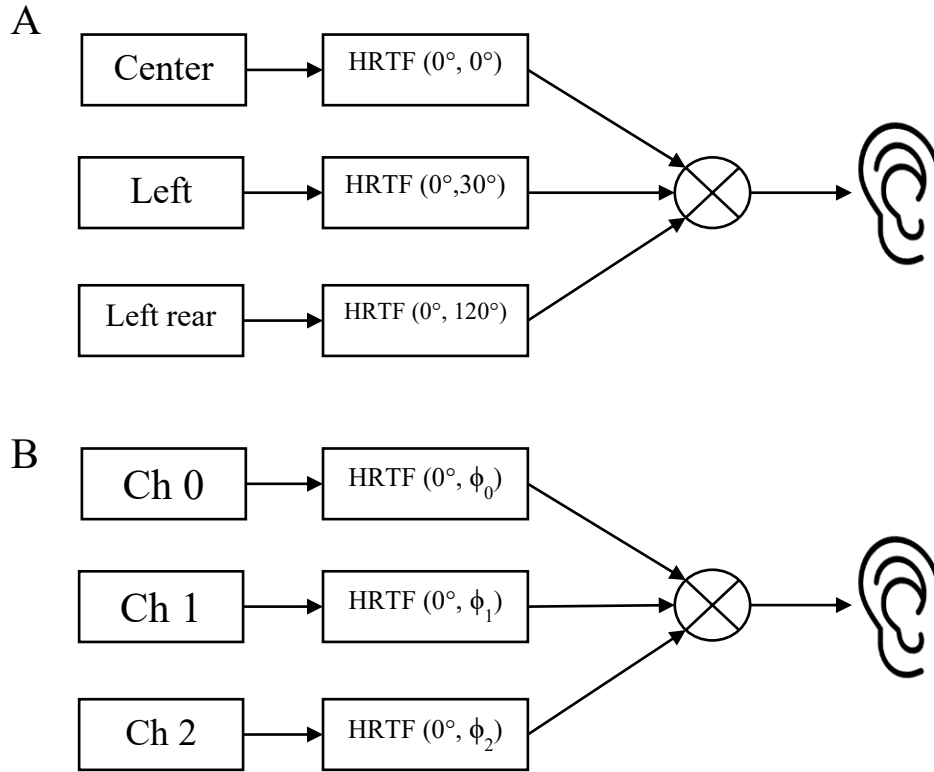


Figure 10. Virtual surrounding headphones diagram using HRTF filter. A: 5.1 Surrounding system;  
B: First order ambisonics system.

as MDR-F1, the DS-8000 system originally had a speaker-comparable perceptual sound field and simply applying the VBAP with matched ITD-ILD pair is enough to reproduce a surrounding sound field via headphones. Moreover, the DS-8000 system is equipped with a dual-channel IR system, allowing head-tracking functionality, and providing a more realistic feeling. More common designs are like Pioneer DRS-3000C: compact front chamber design to suppress reflection and reverberation; gaming aid usually features the close-back while home theater prefers the open-back for better clarity in space-sensing [53]. In some cases like Google VR video, first-order ambisonics (FOA) is applied and the decoding mechanism is shown in Figure 10B with the incident angle panning in a spherical way. In general, FOA uses less microphone array to provide similar information as the Dolby/DTS system, but the user's experience is close.



Figure 11. Advanced virtual surrounding headphones using Dolby Atmos or Higher order ambisonics. A: Apple AirPods Max; B: Audeze Mobius and C: HRTF customization setup for Mobius.

The listening test reports that the DS-8000 system could approach the realistic feeling of a home theater best among the four followed by the AKG K290 due to the DF target curve [54]. The DRS-3000C reproduced sound does show a directional feeling, but the imaging is softened and vague with narrowed sound field sensing. The physical multi-channel headphones ranked the worst for driver interference. So far, it is clear that realistic sound field perception requires the headphones to be wide-filed, i.e., extra channels can improve sound localization but not sound field externalization.

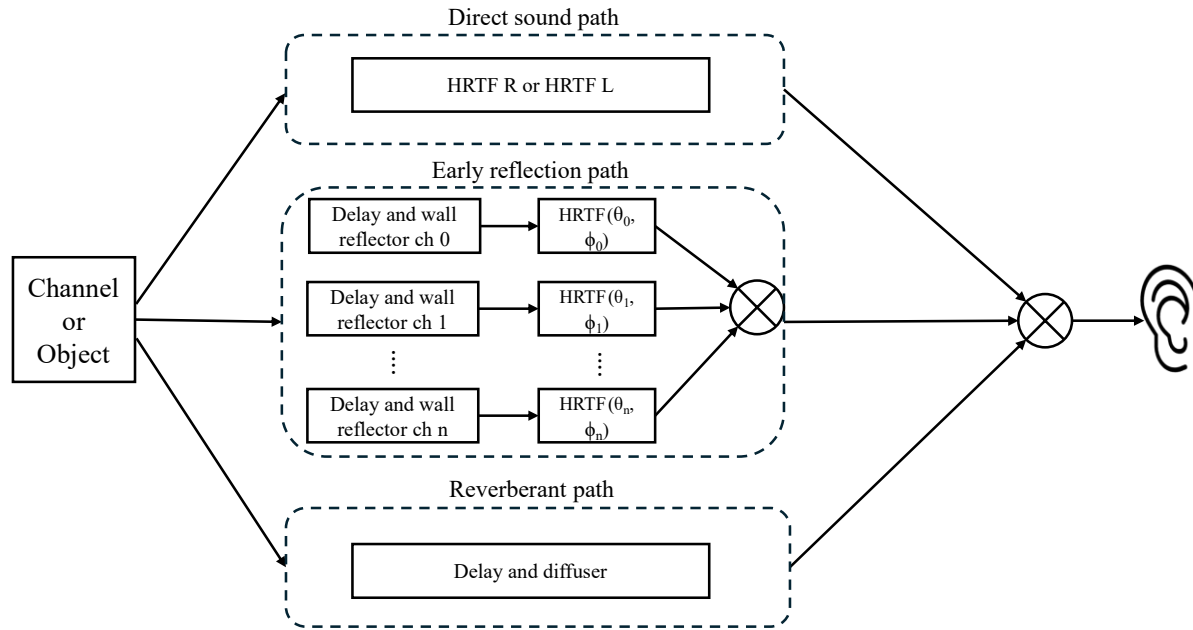


Figure 12. Virtual surrounding headphones diagram using Dolby Atmos or Higher-order ambisonics.

Based on these observations, advanced headphone models for spatial sound reproduction are proposed in Figure 11. Figure 11A presents the Apple AirPods Max (APM), which uses Dolby Atmos technology with a built-in and easy-to-set HRTF measurement. Figure 11B presents the Audeze Mobius using the Higher-order ambisonics (HOA) technology from Wave NX. Figure 11C presents the customization panel for each user's individual HRTF using several parameters like head circumference [55]. These techniques allow more channels to localize the sound object (Dolby Atmos of 60, HOA of 64) than the surround sound system in Figure 10. The improved sound field reproducing process is shown in Figure 12. The HRTF will first modulate each channel or sound object to form the direct sound for sharp imaging, and then a delayed multi-channel scattering process based on HRTF simulates the early reflection from the wall, and finally a reverberator reproduces the sound diffusion and reverberant on the room size impression [12]. Such a process therefore rebuilds the acoustic process of a real home theater and, as long as the headphone's design suppresses all kinds of reflection and reverberation, the user's perception will be hard to distinguish from the real one.

Practical listening agrees with that. Although the sound field depends on the source, both APM and Mobius preserve the sense of the space with no in-head localization effect. They can hear the reproduced "speak person" step close or distant clearly from the back without any dangling, even though both are designed in a close-back way [55]. Additionally, APM and Mobius track the head movement using a gyroscope sensor, increasing the "trustability" like the DS-8000 system. Comparisons to the KK system (with room simulation) and the DS-8000 system are done here. No doubt that the Dolby Atmos and HOA systems show a great ability to sense the back and side, but the front situations can be different [56,57,58]. Both the KK and the DS-8000 systems win the game since they have an open-front chamber with a tilted driver displacement [59,60]. This design allows the outer ear to do its job, or in other words, much finer individual HRTF than the simple measurement or statistical model with a hint from the aerodynamic ---- airflow from the front may imply a

front-coming sound [61,62]. An AR/VR preferred design could be drawn from the comparison: open or closed, this design must suppress everything other than the directly delivered sound with a tilted driver to maximize the use of the outer ear, then the same system measured HRTF should be applied in reproducing the early reflection and reverberant[63].

#### Neural networks approach

Although it's a new field of acoustic studies, neural networks have already taken a place in noise pattern classification and cancellation in industrial products and the myth is simple [4]. Neural networks are considered as an approach to any function depending on training data. Feeding networks with a sampled speech in noise environments and desired pure speech signals then can train a network to filter out the background noise with the vocal voice keeping [56].

A recent thesis by Teodors Kerimovs tapped into this field and proposed a Recurrent Neural network (RNN) model to simulate an actual loudspeaker in acoustic environments. The method is presented as the following [0]:

Prepare several stereo waveform files that cover different music genres and human listening bandwidth.

Record the loudspeaker system using a dummy head using the inner microphone and prepare the input-response dataset.

Train the RNN model using 90% of the dataset and use the rest 10% as a test.

One may recognize this process that the RNN emulates a Binaural Room Impulse Response (BRIR) of such a speaker system. Studies on BRIR and headphones have pointed out that convoluting BRIR before the signal drives the headphones can reproduce the acoustic field characteristics and consequently reproduce the feeling of a real sound field in a virtual world [15,57].

A design of computational method to enlarge and externalize the perception of sound field then can be proposed:

Prepare several stereo waveform files that cover different music genres and human listening bandwidth.

Record the headphones system (direct sound only) using a dummy head and prepare the input dataset.

Record the Dolby Atmos or HOA loudspeaker system using a dummy head and prepare the response dataset.

Train the RNN model using the input-response dataset and emulate a mapping function that transfers the field of a headphone to a field of theater.

## SUMMARY

In this work, several headphones are studied to expand the externalized virtual sound field for the future use of AR/VR. Based on the theory of reverberation field reproduction and HRTF, three types of headphones are proposed, reviewed, and listed the open-front design, large-closed front design, and virtual surrounding technology.

Overall the open-front design wins the game as this type presents the largest reproduction of the sound field over the other two. Especially, studies from AKG K1000 have shown that the best matching incident angle between the ear canal entrance and the headphone driving unit is about  $22^\circ$ . Several large-closed-front-designed headphones violate this angle like STAX Sigma and Ultrasonic Edition 10, and the listening test did not show positive for these designs, no matter with larger angle or asymmetric displacement. However, these displacements developed modern headphones and inspired Sony and Sennheiser to better products, despite the Drop OAE1. One can notice that these two designs are distinguished: one for the free field and one for the reverberant field, and modern headphones thus realize a reproduction of acoustic room response by adjusting the physical process in the front chamber.

Virtual surround sound is based on the model of how the human ear senses the room. Therefore, the acoustic tuning of this type of sound mostly only contains direct sound, so the reflection and reverberation can be customized to any room. An interesting finding arising from testing virtual surrounding headphones is that although algorithms can cheat human ears, the largest sound field reconstruction is agreed to be the Sony DS-8000 system that combines room simulation and open-

front design. In fact, an appropriate incident angle and simply applying cross-feed can simulate a near-field monitor in reviewing the AKG K1000. Thus, this work proposed that this type of combination serves the AR/VR use better, with the neural network constructing a mapping from headphone response to the desired room's response.

#### **DATA AVAILABILITY STATEMENT**

No new data were created or analysed in this study.

## References

1. Chen, Je Ru, S. J. Pawar, and Jin H. Huang. "Structural modifications of headphone front chamber for better frequency response: Experimental and simulation studies." *Acoustics Australia* 49.1 (2021): 69-82.
2. Hiipakka, Marko. "Estimating pressure at the eardrum for binaural reproduction." (2012).
3. Downs, Jacob Kingsbury. *Headphone listening: space, embodiment, materiality*. Diss. University of Sheffield, 2021.
4. Kerimovs, Teodors. *Loudspeaker Modelling with Recurrent Neural Networks*. MS thesis. 2023.
5. Li, Song, and Jürgen Peissig. "Measurement of head-related transfer functions: A review." *Applied Sciences* 10.14 (2020): 5014.
6. Olive, Sean E. "Selecting a reference headphone for 3D audio reproduction." *The Journal of the Acoustical Society of America* 152.4\_Supplement (2022): A150-A150.
7. Kinnunen, Teemu. "Headphone development research." (2022).
8. Pfanzagl-Cardone, Edwin. "The DOLBY® "Atmos™" System." *The Art and Science of 3D Audio Recording*. Cham: Springer International Publishing, 2023. 143-188.
9. Birnie, Lachlan, et al. "Mixed source sound field translation for virtual binaural application with perceptual validation." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29 (2021): 1188-1203.
10. Eiderbo, Ian. "How does binaural audio mixed for headphones translate to loudspeaker setups in terms of listener preferences?." (2021).
11. McCormack, Leo, and Symeon Delikaris-Manias. "Parametric first-order ambisonic decoding for headphones utilising the cross-pattern coherence algorithm." *EAA Spatial Audio Signal Processing Symposium*. 2019.
12. Bauer, Valentin, et al. "Binaural headphone monitoring to enhance musicians' immersion in performance." *Advances in Fundamental and Applied Research on Spatial Audio* (2022).
13. Pei, Yukun, You Li, and Pablo Ripollés. "Automotive Audio System Evaluation over Headphones Based on the BVIRs of Different Listening Positions: A Case Study of a Specific Audio System." *Audio Engineering Society Conference: AES 2022 International Automotive Audio Conference*, Audio Engineering Society. 2022.
14. Yang, Jun, Ming Wu, and Lu Han. "A review of sound field control." *Applied Sciences* 12.14 (2022): 7319.
15. Everest, F. Alton. *Master handbook of acoustics*. 2022.
16. Møller, Aage R. "Improved technique for detailed measurements of the middle ear impedance." *The Journal of the Acoustical Society of America* 32.2 (1960): 250-257.
17. Menzies, Dylan, and Marwan Al-Akaidi. "Nearfield binaural synthesis and ambisonics." *The Journal of the Acoustical Society of America* 121.3 (2007): 1559-1563.
18. Otani, Makoto, et al. "Binaural Ambisonics: Its optimization and applications for auralization." *Acoustical Science and Technology* 41.1 (2020): 142-150.
19. Cliffe, Laurence, et al. "Materialising contexts: virtual soundscapes for real-world exploration." *Personal and Ubiquitous Computing* 25 (2021): 623-636.
20. Roquet, Paul. "Ambiance In and Around the Virtual Reality Headset." *Proceedings of the 4th International Congress on Ambiances, Alloaesthesia: Senses, Inventions, Worlds. Vol. 2. Réseau International Ambiances*, 2020.
21. Li, Song. "On Externalization of Virtual Sound Images Presented via Headphones." (2021).
22. Aguirre, Sergio Luiz. "Iceberg: a loudspeaker-based room auralization method for auditory research." (2023).
23. da Silveira Zanin, Camila Elizabeth, et al. "Experimental evidences of nonlinear programmable electroacoustic loudspeaker." *Building Acoustics* 30.3 (2023): 249-263.
24. McCormack, Leo, and Symeon Delikaris-Manias. "Parametric first-order ambisonic decoding for headphones utilising the cross-pattern coherence algorithm." *EAA Spatial Audio Signal Processing Symposium*. 2019.
25. Menzies, Dylan, and Marwan Al-Akaidi. "Nearfield binaural synthesis and ambisonics." *The Journal of the Acoustical Society of America* 121.3 (2007): 1559-1563.
26. Wightman, Frederic L., and Doris J. Kistler. *Hearing in three dimensions: Sound localization*. No. NAS 1.26: 187706. 1990.
27. Li, Song. "On Externalization of Virtual Sound Images Presented via Headphones." (2021).
28. Nowak, Patrick. *Spatial audio through headphones based on HRTFs approximated by parametric IIR filters*. Diss. Universitätsbibliothek der HSU/UniBwH, 2022.
29. Roquet, Paul. "Ambiance In and Around the Virtual Reality Headset." *Proceedings of the 4th International Congress on Ambiances, Alloaesthesia: Senses, Inventions, Worlds. Vol. 2. Réseau International Ambiances*, 2020.
30. Brahimaj, Detjon, et al. "Enhancing Object Localization in VR: Tactile-Based HRTF and Vibration Headphones for Spatial Haptic Feedback." *IEEE Transactions on Haptics (ToH)* (2024).
31. Kinnunen, Teemu. "Headphone development research." (2022).
32. Brilhante, Davi da Silva, et al. "A literature survey on AI-aided beamforming and beam management for 5G and 6G systems." *Sensors* 23.9 (2023): 4359.
33. Nowak, Patrick. *Spatial audio through headphones based on HRTFs approximated by parametric IIR filters*. Diss. Universitätsbibliothek der

HSU/UniBwH, 2022.

34. Sumner, Sydney. "The Digital Ears: A Binaural Spatialization Plugin." (2022).
35. dos Santos, A. N., D. Fernandes, and B. S. Masiero. "3-D Audio Synthesis: A DIY Approach for HRIR Database Acquisition."
36. Soeta, Yoshiharu, and Sho Ooi. "Spatial audio quality factors for headphone evaluation." 2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA). IEEE, 2021.
37. Grimaldi, Vincent, et al. "Externalization of virtual sounds using low computational cost algorithms for hearables." (2020).
38. Pei, Yukun, You Li, and Pablo Ripollés. "Automotive Audio System Evaluation over Headphones Based on the BVIRs of Different Listening Positions: A Case Study of a Specific Audio System." Audio Engineering Society Conference: AES 2022 International Automotive Audio Conference, Audio Engineering Society. 2022.
39. Johnston, James D., and Yin Hay Vicky Lam. "Perceptual soundfield reconstruction." Audio Engineering Society Convention 109. Audio Engineering Society, 2000.
40. Dylan Menzies, Marwan Al-Akaidi; Nearfield binaural synthesis and ambisonics. J. Acoust. Soc. Am. 1 March 2007; 121 (3): 1559–1563.
41. Sean E. Olive; Selecting a reference headphone for 3D audio reproduction. J. Acoust. Soc. Am. 1 October 2022; 152 (4, Supplement): A150.
42. Lachlan Birnie, Thushara Abhayapala, Vladimir Tourbabin, and Prasanga Samarasinghe. 2021. Mixed Source Sound Field Translation for Virtual Binaural Application With Perceptual Validation. IEEE/ACM Trans. Audio, Speech and Lang. Proc. 29 (2021), 1188–1203.
43. Jin, Craig T. "A tutorial on immersive three-dimensional sound technologies." Acoustical Science and Technology (2020): n. pag.
44. S. D. You and F. -H. -C. Fang, "Using Semi-personalized Loudness Difference for Improved Surround Sound Rendered by Headphones," 2021 3rd International Conference on Computer Communication and the Internet (ICCCI), Nagoya, Japan, 2021, pp. 168-170
45. 山高正烈, 坂本修一, and 鈴木陽一. "頭部伝達関数におけるノッチ幅と中心周波数の偏移が音色知覚と音像定位に及ぼす影響." 愛知工科大学紀要= Bulletin of Aichi University of Technology 19 (2022): 19-24.
46. 西山織絵, and 飯田一博. "正面方向の個人化パラメトリック頭部伝達関数の生成—ノッチとピークのレベル・尖鋭度設定における固定値の適用性検証—." 聴覚研究会資料= Proceedings of the auditory research meeting. Vol. 51. No. 2. 日本音響学会, 2021.
47. Begault, Durand R., and Leonard J. Trejo. 3-D sound for virtual reality and multimedia. No. NASA/TM-2000-209606. 2000.
48. Jianjun, H. E., Ee Leng Tan, and Woon-Seng Gan. "Natural sound rendering for headphones: integration of signal processing techniques." IEEE Signal Processing Magazine 32.2 (2015): 100-113.
49. Nishimura, Ryouichi, et al. "An attempt to calibrate headphones for reproduction of sound pressure at the eardrum." IEEE transactions on audio, speech, and language processing 19.7 (2011): 2137-2145.
50. Liitola, Toni. "Headphone sound externalization." Science in Technology, Tampere, XP055267926, pp. 1-74 (2006).
51. Johnston, James D., and Yin Hay Vicky Lam. "Perceptual soundfield reconstruction." Audio Engineering Society Convention 109. Audio Engineering Society, 2000.
52. Zaunschirm, Markus, Matthias Frank, and Franz Zotter. "Binaural rendering with measured room responses: First-order ambisonic microphone vs. dummy head." Applied Sciences 10.5 (2020): 1631.
53. Narbutt, Mirosław, et al. "Ambiquat: Towards a quality metric for headphone rendered compressed ambisonic spatial audio." Applied Sciences 10.9 (2020): 3188.
54. Vorländer, Michael. "Are virtual sounds real." Acoustics Today 16.1 (2020): 46-54.
55. Wang, Jiarui, et al. "Binaural Reproduction Using Multi-Driver Headphones." 2022 International Workshop on Acoustic Signal Enhancement (IWAENC). IEEE, 2022.
56. Bernardini, Alberto, Lucio Bianchi, and Augusto Sarti. "Loudspeaker virtualization—Part I: Digital modeling and implementation of the nonlinear transducer equivalent circuit." Signal Processing 202 (2023): 108720.
57. Miccini, Riccardo, and Simone Spagnol. "HRTF individualization using deep learning." 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE, 2020.
58. York University. "SADIE Project." SADIE Project, University of York.
59. Tsuchiya, Keisuke, et al. "Nearfield Binaural Synthesis and Ambisonics." The Journal of the Acoustical Society of America, vol. 121, no. 3, 2007, pp. 1559-1571.
60. Ito, Yasushi, et al. "AST: Journal of Advanced Science and Technology." Advanced Science and Technology, vol. 41, no. 1, 2021, Article E19229.
61. Zhu, Bin, et al. "A Deep Learning Approach for Human Activity Recognition Based on Smartphones." IEEE Transactions on Mobile Computing, vol. 19, no. 1, 2020, pp. 141-158.
62. Li, Song. "Dissertation: Advanced Structural Dynamics." Leibniz University Hannover, 2021.
63. Niskanen, Markku, et al. "Dissertation: Technological Innovations in Sustainable Architecture." Aalto University, 2021.