Relevance and Prediction of Externalization Based On Binaural Cues

Arttu Pahta
Aalto Universtiy
Master's Programme CCIS / AAT

arttu.pahta@aalto.fi

Abstract

This seminar paper explores sound externalization and ways to predict it. Externalization is a perceptual phenomenon in which sound is perceived as coming from outside the head. It is crucial for creating credible virtual audio environments. The paper studies the interplay of interaural level differences (ILDs), spectral details, and reverberation effects, on externalization. Analysis of binaural cues highlights the significance of these components in creating realistic virtual acoustic environments. The paper showcases a predictive model that uses various psychoacoustic cues to forecast the level of sound externalization perceived. These predicted levels of externalization are compared to subjective results from test subjects to check the validity of the model. The study underscores the importance of accurate ways to predict externalization level. Which allows for real-like soundscapes that support not only the visual content in entertainment but also can be sufficient on its own to create credible surround feel to the user. This work not only contributes to our understanding of auditory perception but also underscores its importance in the development of advanced audio technologies and hearing aids.

1 Introduction

Externalization of sound is a term that refers to the perception that a sound source is located outside the listener's head. Contrary to internalization, an effect where the sound is perceived to be coming from the inside of a person's head. Externalization plays a crucial role in locating the sound source to a certain distance or direction. Externalization is needed for creating credible virtual sound systems through headphones. [1]

Binaural technology enhances the realism of virtual sound systems by simulating the way sound is perceived in natural environments. Unlike monaural systems, which use a single sound source typically heard through one earpiece, binaural sound systems employ two audio channels to create a more expansive aural scene. Use of binaural sound systems is necessary for externalization of the sound but doesn't guarantee externalized sound. To produce credible binaural sound systems understanding of binaural modelling is necessary. Binaural modelling involves computational algorithms that simulate the auditory cues humans use to locate sounds in three-dimensional space. These binaural cues are classified into two main types: interaural and monaural cues. Interaural cues, which include *interaural time differences* (ITD), *interaural level differences* (ILD), and *interaural phase differences* (IPD), require analysis from both ears to assess the direction and distance of sound sources. Monaural cues, on the other hand, are processed by a single ear and contribute to the perception of sound elevation and timbre. Effective binaural modeling not only replicates these auditory cues but also incorporates them into localization models. These models estimate the positions of sound sources by mimicking the auditory processing of the human brain, thus enhancing the listener's sense of spatial awareness in a virtual environment. By refining these models, binaural sound systems can more accurately reproduce the auditory landscape, significantly improving the user's experience. [2]

Credible binaural sound sources are typically created using *head-related impulse responses* (HRIR) with anechoic audio signals. These two signals are convolved to create a credible spatial virtual sound. The frequency domain representation of HRIRs, *head-related transfer functions* (HRTF) consists of psychoacoustic cues. These cues play a crucial role in creating well externalized virtual sound sources. HRTFs need to be measured individually for each person as they are unique for each person. For these individually measured HRTFs the virtual sounds are perceived as externalized. [3] Externalization relies on ITDs and IPDs at lower frequencies and on ILDs at all frequencies. None of these cues are capable of producing a well-externalized virtual sound source at their own. The spectral information is also needed to achieve an externalized sound. [4] Reverberation of the virtual sound also plays a part in creating externalized sound [5]. Monaural reverberation cues are sufficient enough to produce externalized lateral sounds but for frontal externalization the reverberation cues require binaural information signals. [6]

Evaluating binaural cues is essential for advancing the externalization of sound in virtual sound systems. The complex interplay of interaural and monaural cues, along with the spectral and reverberation information, forms the foundation for creating a convincingly externalized auditory environment. As such, our understanding and precise modeling of these cues are critical. They allow us to tailor sound experiences to individual listeners by accounting for the unique acoustic signatures captured by HRTFs. This individualization is vital as it significantly enhances the credibility and spatial accuracy of virtual auditory environments, thereby improving user immersion and satisfaction. By focusing on the detailed study of binaural cues and their integration into sound system design, we can push the boundaries of auditory virtual reality, making it as lifelike and immersive as possible.

The ability to detect spatial cues decays can decline with aging. Age-related changes in auditory function, including declines in monaural temporal processing and neural synchrony, as well as reduced central inhibition, have been associated with poorer performance on binaural tasks that require precise temporal processing. This affects tasks involving lateralization, localization, and detecting signals in noise, indicating that aging impacts the neural encoding and processing of binaural and spatial cues. [7] Also, hearing aids can disturb the externalization [8]. Therefore, there is a need for understanding of the levels of externalization not only in the entertainment industry but in the hearing aid industry as well.

2 Key Factors in Sound Externalization

Interaural level differences are critical auditory cues. ILDs determine the location of sounds. These cues are based on sound intensity levels received by two separate ears. From ILDs the position of sound source can be analyzed by the psychoacoustic system. Even though ILD is the oldest theory of directional hearing, it isn't vastly utilized in computer-based systems. The estimation ILD is done by calculating the relative energy difference between two sound signals mimicking the separate ears. This can be used to determine the sound source's location by comparing the intensity of these signals. [9]

The perception of externalization of sound can be modelled by a conceptual localization model. The model utilizes the differences between long-term and short-term memory and the difference which information related to auditory models are stored in which. [10] The HRIRs of person are stored in the long-term memory and the acoustic cues caused by the reverberation are stored in the short-term memory. To analyze the incoming sound the information in long-term memory and the constantly changing analysis of short-term memory are both needed. A common technique for creating binaural audio involves taking a clean (anechoic) sound recording and then applying a set of (HRIRs) or binaural room impulse responses (BRIRs). This method encounters problems if the spatial properties of virtual sound aren't matching with the listeners. The result in externalization is a distorted sound image or the sound images are perceived inside of the head of test subject. To perceive well-externalized sound images the information received from virtual sound sources needs to contain similar information that is stored in both memories. [3]

Li, et al. [3] conducted an experiment to study the influence of ILD externalization. The study was conducted to five (5) test persons who had previous experience in same kind of studies. These test subjects were aged between 24-30 years and had no hearing impairments. In the study the study they focused on three acoustics cues ILD, spectral information and reverberation. The externalization rating scale used in these studies are shown in table.

Table 1: Scale for externalization used by Li et al.

Degree	Meaning of the degree
3	The sound is externalized and at the position of the loudspeaker.
2	The sound is externalized but not as far as the loudspeaker.
1	The sound is not well-externalized. It is at my ear.
0	The sound is in my head.

2.1 Influence of ILD

Keeping the changes in sound signal only in the ILD, without altering the spectral information, the sound level was adjusted for each ear independently. This was achieved by adjusting the sound level at the right (contralateral) ear by 0, 5, 10, 15, and 20 dB across different frequency ranges (broadband [0.2–16 kHz], low frequency [0.2–3 kHz], and high frequency [3-16 kHz]) while maintaining the sound pressure level (SPL) of the left (ipsilateral) ear signal constant. This method allowed for the isolation of the ILD effect without altering the spectral content in the head-related transfer functions (HRTFs). [3]

Externalization levels decreased with increasing ILD difference across all frequency ranges. Notably, the impact of ILD expansion was particularly significant in the broadband and low-frequency ranges, where an increase of 5 dB in ILD notably reduced the perception of externalization. The sound image shifted closer to the ear, losing its externalized quality with further ILD increase. In contrast, the high-frequency range exhibited a lesser impact on perceived externalization, where even with a 20 dB ILD increase, the sound source maintained a degree of externalization. These outcomes underscore the nuanced role ILDs play across different frequency spectrums in shaping the externalization of sound images. The impact of ILD alternations to externalization are shown in Figure 1. ”BB” referring to Broadband, ”LO” for Low frequency range, and ”HI” for high frequency range. [3]

2.2 Influence of Spectral Details

Spectral details, particularly those associated with HRTFs, are fundamental for externalization. The degradation or smoothing of the spectral cues typically provided by the natural HRTFs can significantly impact the externalization of sound sources. When the spectral details in the HRTFs are altered, especially in the ipsilateral ear (the ear closer to the sound source), the perceived externalization is notably reduced. This outcome can be detected if the spectral information is systematically manipulated, keeping interaural level differences (ILDs) constant. [3]

Li et al manipulated the spectral magnitude by applying different smoothing levels. They observed that as the spectral cues became less distinct due to increased

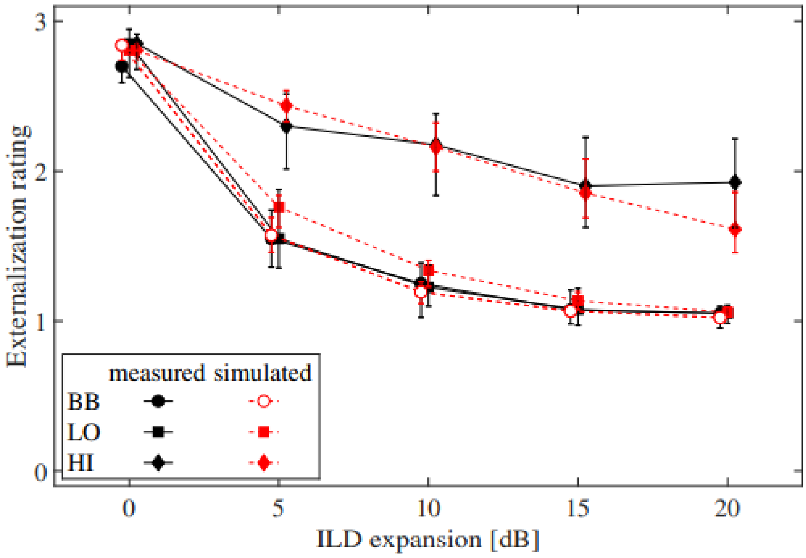

Figure 1: Median values of externalization ratings (solid lines) and Li, et al. prediction model for level of externalization (dashed lines) for ILD expansions in three different frequency ranges ("BB", "LO" and "HI").[3]

smoothing, the listeners reported a reduced sense of externalization. The results highlighted that fine spectral details play a crucial role in the auditory system's ability to localize sounds in the external environment. [3]

Furthermore, the influence of spectral details is not uniform across all frequencies. The study indicated that spectral cues at lower frequencies contribute differently to the perception of externalization compared to those at higher frequencies. The effect of spectral smoothing in HRTF was scaled on an equivalent rectangular bandwidth (ERB) scale for ERB $\in$ 0,1,4,16,64. Where for ERB $= 0$ no spectral smoothing was applied. [3] The effects of spectral smoothing in different levels of spectral smoothing are shown in Figure 2. ERB is used to mimic the frequency resolution of human auditory system. It represents the bandwidth of a rectangular filter that would pass the same amount of power as a given auditory filter modeled as a gammatone filter. The ERB scale is closely related to how humans perceive differences in frequencies. [11]

2.3 Effects of Reverberation

Reverberation plays a pivotal role in the externalization of sound, profoundly influencing the listener's perception of auditory space. In virtual environments, the manipulation of reverberation is crucial for mimicking the acoustic characteristics of real-world spaces, thereby enhancing the realism and spatial accuracy of sound. Li et al. [3] conducted extensive experiments to quantify the impact of reverberation

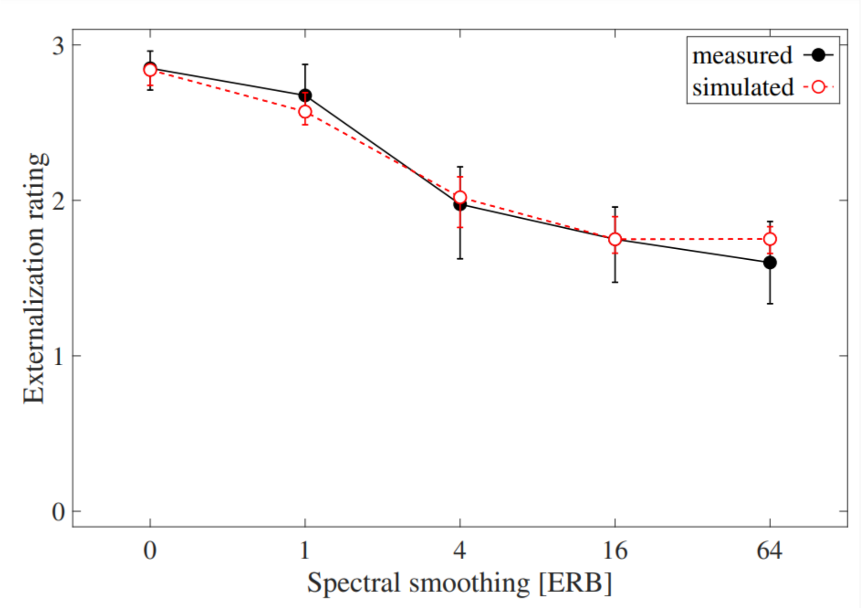

Figure 2: *Externalization rating related to Spectral smoothing of the HRTFs on ipsilateral ear. Measured and simulated results.* [3]

on perceived externalization, particularly focusing on its interaction with monaural spectral cues and interaural level differences (ILDs). Through controlled adjustments of reverberation levels in anechoic and reverberant conditions, the study demonstrated that reverberation significantly modifies the effectiveness of these binaural cues. While reverberation generally diminishes the clarity of monaural and ILD cues, it does not fully mask their influence on externalization perceptions. The experimental results showed that increasing reverberation time led to a reduction in the degree of perceived externalization. This suggests that while reverberation adds a layer of spatial context, excessive reverberation can cloud the auditory cues that contribute to the localization and externalization of sounds. [3] Reverberation alone isn't sufficient to produce externalization. This can be seen in the Figure 3 as the spectral smoothed signal with 0 % reverberation reduction isn't perceived as externalized. Reverberation on the other hand is necessary for externalization to happen. Even the signals containing all of the spectral information aren't perceived as externalized if the reverberation is reduced greatly.

3 Predicting the Level of Externalization

Predicting the level of perceived externalization in auditory systems is essential for developing more realistic and immersive virtual acoustic environments. Li et al. proposed a comprehensive model that integrates various psychoacoustic cues to predict externalization effectively. The results received with this model where successful. The accuracy of predicted externalization ratings was reported to be higher than 90%.

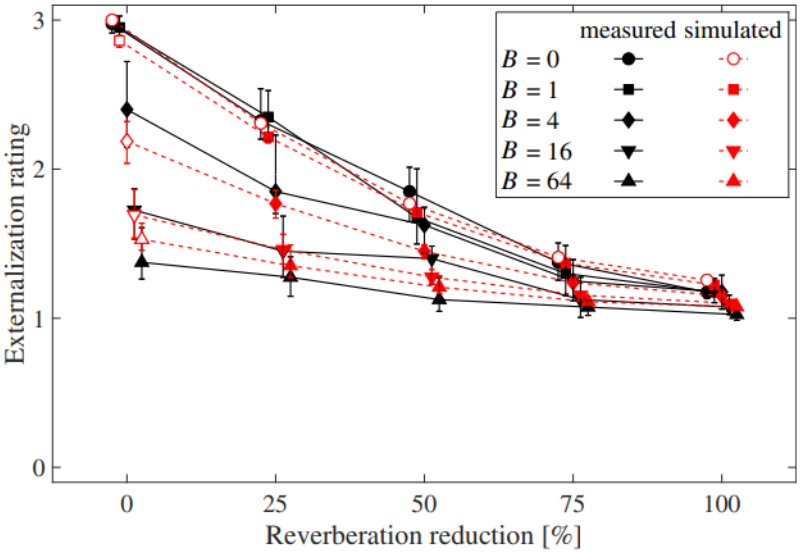

Figure 3: *Externalization rating for combined Spectral smoothing and reverberation reduction. Measured and simulated results* [3]

3.1 Predictive Model

The predictive model is based on the interplay of ILDs, monaural spectral cues, and the temporal fluctuations of these cues. As shown in the subjective tests these all play a crucial role. The model suggests ways to evaluate the levels of these cues and their effect on the level of externalization perceived. [3]

The model involves comparing the modified acoustic cues from the target sound with those of an unprocessed template, which represents the listener's natural acoustic environment. By evaluating the discrepancies in ILD and spectral information, the model predicts the level of externalization.

ILD is primarily significant at all frequencies, and monaural spectral cues, which provide information on sound elevation and timbre, are also considered in the model. The model also uses reverberation characteristics of the environment, which rely on the clarity of other cues, still significantly influence the perceived externalization of the sound.

3.2 Computational Implementation

For the computational prediction of externalization level all of the three mentioned attributes have to be calculated.

3.2.1 Spectral Gradient

The spectral information can be calculated for each ear k for each pair of frequency channels with the equation 1 [12].

$$\xi_k(i) = M_k(f_{c,i}) - M_k(f_{c,i-1}) \tag{1}$$

where $M_k(f_{c,i}$ is the excitation in a single frequency band and $i \in 2, 3, ...N$ These received spectral gradients are compared to template spectral gradients to received normalized spectral gradients for each ear according to equation 2 [3].

$$\Delta\xi = \frac{\sum_{i=2}^{N} |\xi_{\text{received},i} - \xi_{\text{template},i}|}{\sum_{i=2}^{N} |\xi_{\text{template},i}|} \tag{2}$$

To receive the weighting factor of spectral gradient based effect on the externalization, the normalized values for each ear are compared according to equation 3 [3].

$$\overline{\Delta\xi}(w) = w\overline{\Delta\xi}_{\text{left}} + (1-w)\overline{\Delta\xi}_{\text{right}} \tag{3}$$

where w is the binaural weighting factor limited between zero to one.

3.2.2 ILD

For the evaluation of ILDs effect on the reverberation the normalized ILD deviations have to be calculated. These are calculated and averaged throughout the frequency bands with equation [13].

$$\Delta\overline{\text{ILD}} = \frac{1}{N}\sum_{i=1}^{N} \frac{|\text{ILD}_{\text{target}}(f_{c,i}) - \text{ILD}_{\text{template}}(f_{c,i})|}{|\text{ILD}_{\text{template}}(f_{c,i})|} \tag{4}$$

where $ILD_{target}(f_{c,i})$ is the target signal $ILD_{template}(f_{c,i})$ is the template signal. Also, to predict the attribute of ILD the temporal fluctuations in it need to be calculated. The receive these the target binaural signals need to processed using an echo-suppression. Echo-suppression is done by setting a time window with following values: 1 from 0 to 2.5 ms (direct sound duration), 0 from 2.5 to 10 ms (echo suppression), transitioning from 0 to 1 from 10 ms to 15 ms using a raised-cosine window. The echo-suppressed signals are filtered through a gammatone filter bank with bandwidths equal to one Equivalent Rectangular Bandwidth (ERB) to mimic cochlear filtering. ILDs are computed for each frequency band using a 20 ms Hann window with 50% overlap, processed over the signal's duration to produce 99 frames for a 1-second signal. The ILD temporal fluctuations are standard deviation across the short-term ILDs. The standard deviations are for each frequency band center are defined in equation 5. [3]

$$ILD_{TSD}(f_c) = \sqrt{\frac{1}{\text{N}_{\text{frame}} - 1}\sum_{n=1}^{\text{N}_{\text{frame}}} (\text{ILD}(f_c, n) - \overline{\text{ILD}}(f_c))^2} \tag{5}$$

where, N_{frame} is the number of frames, $\text{ILD}(f_c, n)$ is the ILD at the n-th frame for frequency f_c, and $\overline{\text{ILD}}(f_c)$ is the average ILD at frequency f_c. Thse need to be normalized according to the equation 6. [3]

$$\overline{\Delta ILD}_{TSD} = \frac{\sum_{i=1}^{N} |ILD_{TSD,\text{target}}(f_{c,i}) - ILD_{TSD,\text{template}}(f_{c,i})|}{\overline{ILD}_{TSD,\text{reference}}} \tag{6}$$

3.2.3 Reverberation

Spectral gradients and ILDs have effect on externalization in anechoic environment but their influence decreases when reverberation is also introduced. The weighting factor for reverberation can be computed following. [3]

$$\gamma = 1 - b_\gamma \frac{\overline{\text{ILD}}_{\text{TSD,template}}}{\overline{\text{ILD}}_{\text{TSD,reference}}} \tag{7}$$

where b_y is a weighting factor, $\overline{\text{ILD}}_{\text{TSD,template}}$ is the current acoustic environment and $\overline{\text{ILD}}_{\text{TSD,reference}}$ is reference acoustic environment.

3.2.4 Final mapping

With information of weightings of spectral gradients, ILDs, and their temporal fluctuations the final externalization level can be predicted. This is done by summing up the weighting factors according to the equation [3]:

$$\Delta m = \gamma(b_{ILD}\overline{\Delta ILD} + b_\xi \overline{\Delta\xi}(w)) + b_{ILDTSD}\overline{\Delta ILD}_{TSD} \tag{8}$$

where b_{ILD}, b_n and $b_{\text{ILD TSD}}$ are weighting factors for deviations of acoustic cues. The mapping between the objective measures and the externalization ratings is represented by an exponential function presented in the equation 9.

$$E = ae^{-\Delta m} + c \tag{9}$$

where a and c are mapping parameters set as 2 and 1 in the simplest model. [3]

3.3 Evaluation and Discussion

The prediction model presented by Li et al. shows good accuracy. The results of the prediction model are compared to ones received in subjective tests are shown in figures 1-3. [3]

Even though the model received great accuracy through the test Li et al computed, the test group of their test was limited. The number of participants was 5 people and all of these test subjects came from background with some previous auditory tests. Therefore, it would be good to evaluate the prediction model even more before applying it to user products.

4 Importance of Predicting the Level of Externalization

In the evolving field of auditory technologies, understanding and predicting the level of externalization is critical. Having well-externalized audio signals not only enriches the realism of virtual environments and enhances user interactions with audio systems but also helps the development of safety and assistive technologies.

4.1 Enhanced Spatial Awareness in Virtual Environments

Even though Virtual Reality (VR) has typically had its focus on the visual display of the systems it significantly benefits from immersive audio systems as well. The spatial awareness can be enhanced by simulating realistic soundscapes. These soundscapes rely on credible level of externalization as well externalized sounds help users to position themselves in the space around them. This improves navigation, interaction and overall engagement with the virtual world. [8] The importance of these systems can be also seen in investments. Google and Facebook have invested in open-source ambisonics audio, crucial for the development of VR audio formats. [14]

Creator of these immersive audio systems have to have some way to evaluate the systems they have created. For this, ability the predict the level of externalization can be helpful. Based on the prediction models Li et al have suggested the level of achieved externalization in these systems can be evaluated without subjective testing of the systems.

4.2 Externalization and Hearing Aids

The impact of hearing aids on sound externalization is a critical but not so deeply researched area in auditory science. The introduction of hearing aids disrupts this natural auditory processing, often leading to sounds being perceived as internalized, or occurring within the head. [8] It is also common for people who need hearing aids to have hearing impairments. The ability to detect spatial cues which help us to perceive sounds as externalized decay with age and hearing impairments. [7]

The reason why hearing aids can cause distortion in externalization is not well understood. There are several theories that have proposed the reason for the effect. Occlusion Effect: Many hearing aids use earmolds that block the ear canal either partially or fully. This can potentially lead to the occlusion effect where external sounds are perceived as coming from inside the head. The effect changes the listener's connection to their environment, which alters the natural perception of sound locations. Binaural Cue Distortion: Hearing aids can also distort the natural binaural cues, which are crucial for sound localization and externalization. These include interaural time differences (ITDs) and interaural level differences (ILDs).

Distortion in these cues due to hearing aid processing can lower the level of externalization. The processing of dynamic-range compression in hearing aids has been shown to reduce ILDs. As previously discussed ILDs have a great impact on the level of perceived externalization. Microphone Placement: In behind-the-ear hearing aids, the microphone is placed above the pinna which can prevent the capture of natural pinna-related spectral cues that have been stored to the person's long time memory. These spectral cues are crucial for sound localization and externalization. [8]

The possibility to understand and evaluate the relevance of ILDs, spectral details and reverberations on perceived externalization level can benefit the design of hearing aids. Finding the crucial levels of auditory cues that need to exist for externalization can help with planning of microphone placements, designs of the cavity and better the processing mechanisms of hearing aids.

5 Conclusions

The sound externalization is a complex interplay between interaural level differences (ILDs), spectral details, and reverberation. On the other hand, sound externalization is crucial for creating credible sound images and has various other applications. The analysis underscores the significance of each component in enhancing the realism and immersive quality of auditory experiences in virtual environments.

ILDs play crucial part in determining the spatial characteristics of sound, giving a foundation for auditory localization. Still, the integration of precise spectral cues, especially those derived from HRTFs, are essential for a truly externalized sound perception. Also, the reverberation in the sound signals have their affection the externalization level of the sound, but the sound can be perceived as externalized even without reverberation.

Detailed understanding and precise modeling of these auditory cues is important, as they allow for carefully tailored soundscapes that enhance user immersion and satisfaction in virtual reality and other audio-intensive applications. The future of auditory technology lies in further refining these models to achieve more lifelike and immersive auditory environments, making the prediction and application of these cues more accurate and effective. Through ongoing research and technological advancements, the potential to fully harness these auditory cues opens up new possibilities for improving not only entertainment and virtual interaction but also practical applications in accessibility and public safety.

References

[1] N.I. Durlach, A. Rigopulos, X.D. Pang, W.S. Woods, A. Kulkarni, H.S. Colburn, and E.M. Wenzel. On the externalization of auditory images. *Presence: Teleoperators Virtual Environments*, 10(2):251–257, 1992.

[2] J. Blauert. *The Technology of Binaural Listening*. Springer, Heidelberg, Berlin, 2013.

[3] S. Li, R. Baumgartner, and J. Peissig. Modeling perceived externalization of a static, lateral sound image. *Acta Acustica*, 4(Article 21):1–15, 2020. https://doi.org/10.1051/aacus/2020020.

[4] W.M. Hartmann and A. Wittenberg. On the externalization of sound images. *J. of the Acoustical Society of America*, 99:3678–3688, 1996.

[5] A.J. Kolarik, B.C.J. Moore, P. Zahorik, S. Cirstea, and S. Pardhan. Auditory distance perception in humans: A review of cues, development, neuronal bases, and effects of sensory loss. *Attention, Perception, Psychophysics*, 78(2):373–395, 2016.

[6] J. Catic, S. Santurette, and T. Dau. The role of reverberation-related binaural cues in the externalization of speech. *J. of the Acoustical Society of America*, 138:1154–1167, 2015.

[7] A.C. Eddins, E.J. Ozmeral, and D.A. Eddins. How aging impacts the encoding of binaural cues and the perception of auditory space. *Hearing Research*, 369:79–89, 2018. https://doi.org/10.1016/j.heares.2018.05.001.

[8] Virginia Best, Robert Baumgartner, Mathieu Lavandier, Piotr Majdak, and Norbert Kopčo. Sound externalization: A review of recent research. *Trends in Hearing*, 24:2331216520948390, 2020. PMID: 32914708.

[9] Stan T. Birchfield and Rajitha Gangishetty. Acoustic localization by interaural level difference. In *2005 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1109–1112, [Place of publication not identified], 2005. IEEE. [Online].

[10] G. Plenge. Über das problem der im-kopf-lokalisation (the problem of in-head localization). *Acustica*, 26:241–252, 1972.

[11] B.C.J. Moore and B.R. Glasberg. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, 74:750–753, 1983.

[12] R. Baumgartner, P. Majdak, and B. Laback. Modeling the effects of sensorineural hearing loss on sound localization in the median plane. *Trends in Hearing*, 20:1–11, 2016.

[13] H. G. Hassager, F. Gran, and T. Dau. The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment. *Journal of the Acoustical Society of America*, 139:2992–3000, 2016.

[14] Xuejing Sun. Immersive audio, capture, transport, and rendering: a review. *APSIPA Transactions on Signal and Information Processing*, 10:e13, 2021.