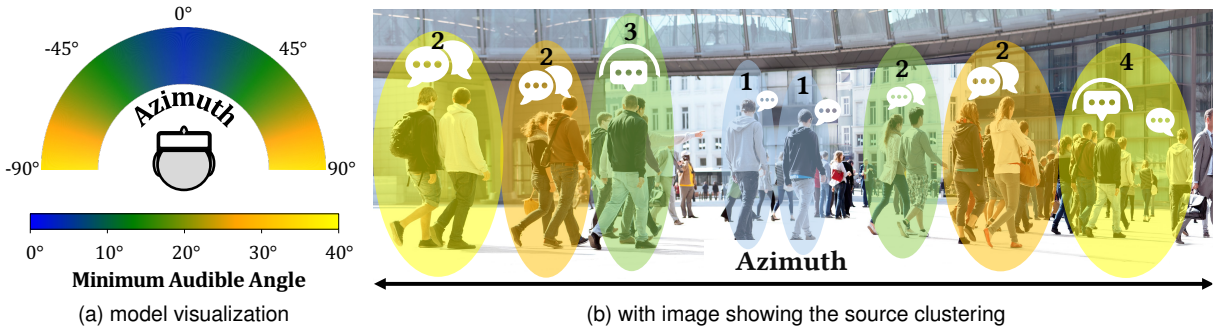
Perceptually-Guided Acoustic "Foveation"

Xi Peng[1,2] * Kenneth Chen[2] Iran Roman[2,3] Juan Pablo Bello[2,3] Qi Sun[2†]
Praneeth Chakravarthula[1†]

[1] UNC-Chapel Hill [2] Tandon School of Engineering, New York University
[3] Music and Audio Research Lab, New York University



(a) model visualization (b) with image showing the source clustering

Figure 1: *Azimuth-based audio perceptual acuity guided sound source clustering. (a) visualizes our model-predicted human auditory perception of spatial discrimination threshold along azimuth eccentricity in degrees. (b) illustrates the model-derived audio source clustering method. Based on listeners' heading direction (assuming forward here), we cluster audio sources that are spatially indistinguishable. Clusters were highlighted by colors corresponding to the human's minimum audible angle. The number of sound sources for each cluster is marked in the figure. Our measurement shows a* 53% *computational saving at the presented scene.*

Abstract

Realistic spatial audio rendering improves immersion in virtual environments. However, the computational complexity of acoustic propagation increases linearly with the number of sources. Consequently, real-time accurate acoustic rendering becomes challenging in highly dynamic scenarios such as virtual and augmented reality (VR/AR). Exploiting the fact that human spatial sensitivity of acoustic sources is not equal at azimuth eccentricities in the horizontal plane, we introduce a perceptually-aware acoustic "foveation" guidance model to the audio rendering pipeline, which can integrate audio sources that are not spatially resolvable by human listeners. To this end, we first conduct a series of psychophysical studies to measure the minimum resolvable audible angular distance under various spatial and background conditions. We leverage this data to derive an azimuth-characterized real-time acoustic foveation algorithm. Numerical analysis and subjective user studies in VR environments demonstrate our method's effectiveness in significantly reducing acoustic rendering workload, without compromising users' spatial perception of audio sources. We believe that the presented research will motivate future investigation into the new frontier of modeling and leveraging human multimodal perceptual limitations — beyond the extensively studied visual acuity — for designing efficient VR/AR systems.

Index Terms: Perception, Virtual reality, Mixed/Augmented reality.

*e-mail: xipeng@cs.unc.edu
†co-corresponding authors.

1 Introduction

Wearable graphics systems, such as untethered VR/AR headsets, often struggle to balance high-fidelity rendering demands against low computational resources. Adaptively allocating compute to render content in alignment with human perceptual sensitivity, an area in perceptual graphics, has emerged as a promising approach to tackle this challenge. For example, in the past decade, visual foveated rendering techniques have been extensively proposed, exploiting the visual acuity degradation in the periphery [40, 54, 26, 61]. Such perceptually-guided methods successfully reduce compute demands without compromising visual experience. However, perceptually-aware audio propagation and rendering is still under-investigated, despite its importance in real-time, interactive and immersive applications such as VR spatial audio rendering.

With computational demand of acoustic rendering increasing with the number of audio sources, efficiently rendering high-fidelity spatial audio in complex auditory environments continues to be a significant challenge. Particularly, generation of late reverberation (LR) [51], which has to be computed for each audio source propagating through the 3D scene, poses a major bottleneck. While there are methods that improve acoustic rendering by filtering or clustering audio sources [59, 51], they do not fully take into account the human perceptual ability for spatial audio localization. Consequently, prior techniques fall short in achieving perceptually optimal computation and compromise perceived audio quality.

In this work, we introduce "foveation" to acoustic rendering. Specifically, we exploit the gradual decline in human spatial auditory acuity as sound sources move away from the central azimuth axis, and establish a probabilistic computational model that estimates human spatial audio localizing sensitivity. Our work goes beyond existing psychoacoustic findings, and provides a novel framework for clustering audio sources based on perceptual thresholds identified from a large-scale human audio perception data collected via subjective studies. Informed by the model, we developed a

perceptually-guided audio source clustering approach that reduces the computational load — up to 94.9% with as large as 300 simultaneous audio sources — without compromising listeners' perception of spatial sound.

To this end, we first design and conduct a series of psychophysical studies on the **minimum audible angle** (**MAA**) [37] to assess the human's spatial sensitivity to auditory sources across the azimuth plane. We observe that the results consistently exhibit a positive correlation between the discrimination threshold and azimuthal eccentricity. From numerical analysis, we derive a real-time perceptually-aware approach that adaptively manages available audio sources in the scene. In particular, guided by our model, we cluster audio sources in a way that is imperceptible by humans with the tracked information of users' head position and heading direction. Subjective evaluation studies demonstrate our method's effectiveness in significantly reducing audio rendering workload without compromising spatial audio quality. We hope that our research will inspire new frontiers in perceptual graphics to study and leverage multimodal perceptual abilities, beyond vision.

Scope of Work. The primary goal of this work is to provide a probabilistic mathematical framework for clustering audio sources that cannot be spatially distinguished due to human perceptual sensitivity of spatial audio. The proposed clustering framework offers a more efficient spatial audio rendering approach. Although we do not derive new audio propagation algorithms in this paper, sound synthesis and audio propagation catered to spatial audio rendering with acoustic "foveation" in complex 3D environments is an exciting future work.

2 Related Work

2.1 Audio Rendering and Propogation

Realistic sound propagation simulation significantly enhances the immersion of a virtual scene by computing the environmental response to a specific source signal, thereby enhancing a user's sense of presence in VR. The simulation of sound propagation includes multiple methods. Wave-based methods solve the acoustic wave equation to simulate acoustic effects accurately [32, 33]. On the other hand, the geometric methods assume sound travels along a straight line with specific energy attenuation [38, 51, 16]. These methods are generally faster to compute than the wave-based techniques, but are incapable of simulating low-frequency acoustic phenomena such as diffraction accurately. Currently, their application is limited to static scenes with few objects, and they are impractical for scenes containing numerous sources. Spatial impulse response rendering (SIRR) is another way to render localized sound events in reverberant environments [35, 44]. These audio propagation algorithms have achieved precise acoustic effect simulation considering complicated virtual scenes and have been developed in multiple software packages and implemented in VR applications [49, 48, 50, 10, 55, 46]. However, these techniques still face a challenging dilemma – the computational requirements of sound propagation algorithms increase linearly with the number of audio sources [51].

Many techniques have been proposed to reduce these computational requirements, including through interpolation [63], adaptive techniques [6], or by considering analogs of optimizations to traditional visual rendering methods [23, 58]. Earlier techniques took advantage of the fact that sound sources may mask each other [4, 39]. Similar to our technique, many approaches aim to cluster multiple audio sources to reduce the number of sources that need to be rendered. For example, [59] proposed a technique to cluster sound sources based on spatial deviation, following the [20] heuristic, which leverages a recursive method to improve the clustering approach. [51] observe that distant or occluded sound sources can be difficult to distinguish individually and cluster the sound sources based on this phenomenon. However, these clustering techniques follow heuristics and are not necessarily optimized with human audio perception.

2.2 Foveation

The performance of the human visual system is non-uniform across the visual field with visual acuity decreasing in the periphery. Taking advantage of this visual performance falloff, foveated rendering vastly reduces compute demands in VR by allocating a higher rendering budget near the fovea and less in the periphery [40, 53]. Further leveraging the visual acuity difference by eye-dominance [34], attention [26], retinotopy [64], local contrast [61], visual metamers [57, 56], and scene-awareness [15], foveated rendering can achieve better perceptual quality with reduced compute cost. This technique has also been extensively used in multiple VR/AR applications, such as AR holography [7], light field [54], display power [14, 11], and view synthesis in VR [13].

While these concepts have been extensively studied for vision, the concept of "foveation", which reduces computational costs in low-acuity areas while maintaining perceptual quality, has been largely overlooked in audio research. To our knowledge, we still haven't found a comprehensive computational model that can reduce the number of audio sources without compromising perceived spatial audio quality to provide acoustic "foveation" guidance for VR/ AR applications.

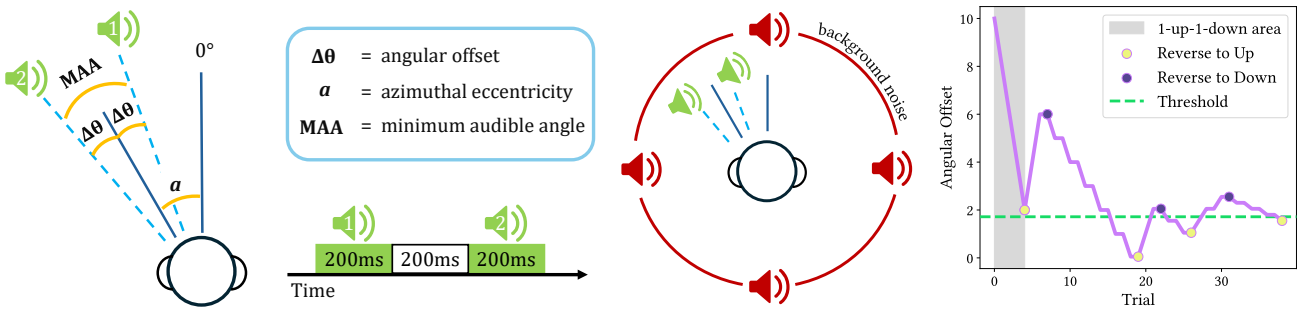
2.3 Human Auditory Perception

Spatial audio is crucial in Virtual/ Augmented Reality [12, 19]. For example, prior art have studied human sound localization ability in web VR [45]. Studies have found that audio significantly affects people's experience in VR [5, 9, 8] and improves users' depth judgment in AR [3]. Integrating with the visual cue, audio could lower the reaction latency [24, 41] and increase the target localization accuracy [3] in VR interactive applications. However, mistakenly rendered audio, like spatially incongruent sound, can degrade visual performance [31]. Therefore, understanding audio perception is essential to provide algorithmic guidance for next-generation VR/AR applications

We want to discover in what range the audio source clustering is not noticeable by humans. The **minimum audible angle** (**MAA**), defined as the smallest detectable angular difference between two identical sources of sound [37], has been proposed to measure human audio localization capability [42, 52, 29, 18]. Several variables have been found to increase MAA, including increased azimuthal eccentricity [37] and increased age [28]. A study by [43] found that the MAA is significantly higher when sound sources are distributed on the horizontal rather than on the vertical plane. Although the above works thoroughly explain the minimum audible angle, to our best knowledge, there is still no algorithmic model that can predict the MAA as a function of the audio's spatial direction of arrival considering natural background noise. Our work aims to establish a MAA prediction model, which can be used to provide guidance for the sound sources clustering method in audio rendering.

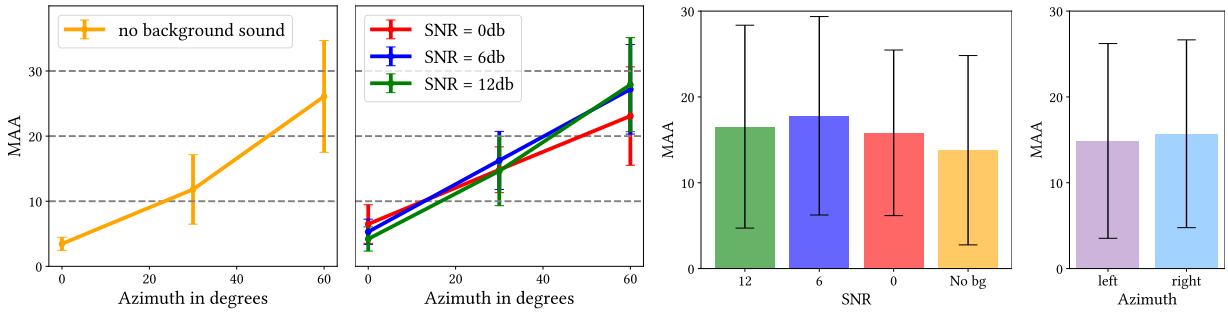
3 Pilot Study: Measuring Human Spatial Acuity of Auditory Sources

In order to acquire the guidance of the source clustering method, we studied human auditory perception from various perspectives. First, we conducted a **main study** for identifying the spatial audio localization acuity along azimuth (Section 3.1). Then, considering the complexities in natural environments where multiple sound sources may concurrently exist, we conducted another **generalizability study** to measure the spatial auditory sensitivity with background noises (Section 3.2). We recruited a large group of participants for a large-scale data collection to derive a perceptual-aware sound source clustering algorithm (Section 4). Additionally, we



(a) main study setup (b) generalizability study setup (c) example staircase procedure

Figure 2: *Study protocol and procedure.* (a) shows the protocol and stimuli of our main study (detailed in Section 3.1). Participants were asked to determine the relative directions of two spatial audio events played sequentially. The minimum audible angle, MAA, and azimuthal eccentricity, a, are defined as illustrations. Angular offset, $\Delta\theta$, was controlled by the staircase procedure based on the participant's responses during the study. (b) presents the stimuli of our generalizability study (detailed in Section 3.2), which consisted of 4 audio sources placed at a participant's north/south/east/west to simulate background noise. (c) shows our study procedure. The 1-up-2-down staircase procedure decreases the angular offset $\Delta\theta$ presented in (a) after a user answers correctly twice (purple scatter points), and increases $\Delta\theta$ if the user answers incorrectly once (yellow scatter point). The procedure was continued until either a total of seven reversals were reached or 50 trials were completed, whichever occurs first. We implemented a 1-up-1-down (grey shaded region) before the first reversal for faster convergence. The discrimination threshold (dashed green line) was obtained by averaging the last 3 reversals in the study.



(a) main study (b) generalizability study (c) by SNR (d) by sides

Figure 3: *Pilot Study results.* (a) visualizes the measured MAA (y-axis) with regard to azimuth (x-axis) in the main study. (b) shows the measured MAA with different SNR in the generalizability study. Error bars represent the standard deviation. (c) illustrates the aggregated MAA without/ with background sounds in different SNR. (d) shows the aggregated MAA on different sides with regard to the zero azimuthal eccentricity. Please refer to Figure 4 and Figure 10 for per-subject results.

also included a small participant group for larger individual sample sizes to validate the masking effect of background noise. Since the audio localization capability is significantly higher when audio sources differ in azimuth than in elevation [43], our study concentrated on azimuth to measure the minimum audible angle.

3.1 Main Study: Auditory Discrimination Across Azimuths

Setup and participants Our psychophysical study was conducted with audio played on the Sony WH-1000XM5 headphones. We recruited 20 participants (ages 21-31, 10 female, 10 male) with normal hearing. During the study, participants remained seated in a completely quiet room and perceived spatial audio stimuli through the over-ear headphones, minimizing the environmental noise. We disabled the active noise-canceling function to eliminate any potential interference from its filtering effects.. Participants interacted with the study software using a keyboard. The study was approved by an Institutional Review Board (IRB).

Tasks and Stimuli As shown in Figure 2a, we measured the participant's minimum audible angle (**MAA**) using a two-interval forced choice, (2IFC) task. During the study, the participants received

1. an audio event followed by another audio event, and
2. pressed a button on a keyboard to report whether the second audio was to the left or right of the first audio.

The audio events were two 20 Hz–20K Hz white noises, both with an intensity of 56dB, rendered in the Unity engine. Same as prior VR audio perception study, [17, 24] We adopted the Meta Oculus SDK for audio spatialization. The 2 audio events were played sequentially per trial, each lasting 200ms with a 200ms break between each. The experiment studied three azimuthal eccentricities at $a = 0°, \pm 30°, \pm 60°$. The audio events were presented around this direction on the azimuth plane, with one played at $a + \Delta\theta$ and the other at $a - \Delta\theta$ (shown in Figure 2a).

Procedure The experiment followed a one-up-two-down 2IFC staircase procedure. Each trial included a 2IFC task where participants needed to discriminate the relative directions of sound sources. We adopted the MAA (as described in Section 2.3 and illustrated in Figure 2a) to determine human audio-spatial perception

([30, 37]). At the start of each trial, we shuffled the relative directions of two sequentially presented sounds, with the first played to the left or right of the azimuthal reference and the second played to the opposite side with the same angular offset $\Delta\theta$.

The one-up-two-down staircase procedure determines the angular offset (illustrated as $\Delta\theta$ in Figure 2a) between audio event and the reference azimuthal eccentricity and makes the task harder or easier depending on the user's response. As shown in Figure 2c, starting from an angular offset of 10 degrees ($\Delta\theta = 10°$), the 1-up-2-down staircase procedure increases $\Delta\theta$ after one incorrect response and decreases it after two consecutive correct responses. The step for these adjustments of $\Delta\theta$ was preset and changed after each reversal following a descending order at $\{2, 2, 1, 1, 0.5, 0.5, 0.25\}$ degrees during the staircase procedure. The staircase terminates after 7 reversals of this procedure or a maximum of 50 trials. We implemented a 1-up-1-down before the first reversal for faster convergence. The discrimination threshold was determined by averaging the last three reversals in the study. The measured threshold of MAA was calculated as $2 \cdot \Delta\theta$.

Conditions We assume that MAA is symmetrical with respect to the zero azimuthal eccentricity, and treat conditions at $a = \pm 30°$ and $a = \pm 60°$ as two repeats at $a = 30°$ and $a = 60°$. For fair cross-condition data collection, we repeated the staircase for zero azimuthal eccentricity ($a = 0°$). Participants completed 6 (3 azimuthal eccentricities $\times$ 2 repeats) staircases with breaks in between. The study takes approximately 15 minutes for each participant to complete.

Results As shown in Figure 3a, the minimum audible angle (MAA) increases with azimuthal eccentricity from 3.46° at $a = 0°$ to 26.09° at $a = 60°$. A one-way ANOVA shows the effect of azimuthal eccentricity on MAA is significant ($F_{2,117} = 148.69, p < .01$), and a Pearson correlation test confirmed a statistically significant positive correlation ($r_{118} = .84, p < .01$). Please refer to Figure 4 for measured MAA per subject.

3.2 Generalizability Study: Validating Masking Effect of Background Audio

The study above provides us with data that suggests MAA significantly degrades along azimuthal eccentricities. Another consideration is that in both natural world and computer graphics applications (such as animation and gaming), massive numbers of sound events may occur concurrently [32]. The *cocktail party effect* [1] is a phenomenon in which humans exclude auditory targets from conscious awareness which are not the main focus. This effect inspired us to validate the feasibility of masking non-target background audio.

Stimuli and conditions We simulated background noise by placing 4 additional audio sources at 90° steps around the user ($a = 0°, 180°, -90°, 90°$), shown in Figure 2b. The background noise was the same full-spectrum white noise as in Section 3.1, but with lower intensity volume, $v_{bg} = 50$dB. Background noise was played constantly per trial. For the foreground audio event, we included two extra novel intensities not seen in Section 3.1, resulting in 3 intensities for foreground sound events ($v_{fg} = 50/56/62$ dB) and 3 signal-to-noise ratios (SNR $= 0/6/12$ dB). The procedure remained the same as the 1-up-2-down 2IFC staircase. We included 18 staircases (3 SNR $\times$ 3 Azimuthal eccentricities $\times$ 2 repeats) conducted per participant. The study takes approximately 45 minutes for each participant to complete. We recruited 6 users (ages 18-30, 2 female, 4 male) for this generalizability study.

Results As shown in Figure 3b and Figure 3c, the average MAA aggregated across azimuthal eccentricities was $14.8°/16.2°/15.5°$ at SNR$= 0/6/12$, respectively. Within the range of $0 \leq$SNR≤ 12, the ANOVA test didn't evidence that the SNR has a significant effect on the MAA($F_{2,105} = 0.18, p = .83$). Compared with the measured data in the no-background-sound condition (aggregated MAA $= 13.79°$ in average), the ANOVA test shows no significant difference in SNR $= 0$ ($F_{1,154} = 0.25, p = .61$), SNR $= 6$ ($F_{1,154} = 1.41, p = .24$), and SNR $= 12$ ($F_{1,154} = 0.72, p = .40$) conditions, shown in Figure 3c. We compared all collected data between left and right azimuthal eccentricity conditions, visualized in Figure 3d. The ANOVA test ($F_{2,243} = 0.32, p = .56$) indicates that the sign of azimuthal eccentricity ($sign(a)$) does not have a significant effect on MAA. See Figure 10 for per-subject results.

3.3 Discussion

The statistical analysis and visualization leads us to several findings. First, the analysis doesn't show the sign of azimuthal eccentricity affects MAA significantly. This verifies our pre-study hypothesis (MAA is symmetrical with respect to the zero azimuthal eccentricity) and provides evidence for repeating the study in ($a = 0$) conditions. Second, the SNR doesn't significantly affect spatial audio localization sensitivity. This is likely due to the cocktail party effect [1], in which background noises are filtered out when humans pay attention to certain sound events. This discovery motivates us to separate the heard audio into *foreground sound* and *background sound*, in which the spatial localization sensitivities of the former are not influenced by the latter. Third and most importantly, the azimuthal eccentricity significantly changes the spatial audio perception given the direction. This azimuth effect aligns with prior work [30] and indicates the increasing difficulty for humans to distinguish audio spatial location at larger azimuthal eccentricities.

4 Implementing Perceptually-Aware Audio Clustering

The experimental results and statistical analysis from Section 3 indicate the significant degradation of auditory acuity along farther azimuthal eccentricities, independent of background sounds. The findings aligned with prior discoveries, as discussed in Section 2.3, verifying the correctness of our study's data and proving its effectiveness. This motivates us to further transform the quantified discoveries toward a generalizable model and run-time algorithm that adaptively clusters audio sources for enhanced computational efficiency.

We found that the measured value of MAA varies across different individuals. In order to establish a robust model for all, we included a percentile index P as an input to our model, with a lower percentile corresponding to a more conservative discrimination threshold prediction. Therefore, given input P, we regressed a specialized polynomial f_{P} on the P$^{\text{th}}$ percentile value of our measured MAA data distribution.

The observed data shows a non-linear increase in the discrimination threshold of MAA in response to azimuthal eccentricities (with 8.35° increase from $a = 0°$ to $a = 30°$, but 14.38° increase from $a = 30°$ to $a = 60°$). Therefore, we fit a 2-degree polynomial,

$$f_{\mathrm{P}}(a) = e_{0,\mathrm{P}} a^2 + e_{1,\mathrm{P}} a + e_{2,\mathrm{P}}, \tag{1}$$

to our MAA threshold data. The coefficients $\{e_{0,\mathrm{P}}, e_{1,\mathrm{P}}, e_{2,\mathrm{P}}\}$ were fitted by minimizing the squared error on the P$^{\text{th}}$ percentile value of our data collected in Section 3.1 as

$$\min_{\{e_{0,\mathrm{P}}, e_{1,\mathrm{P}}, e_{2,\mathrm{P}}\}} \| f_{\mathrm{P}}(a) - MAA_{(a,\mathrm{P})} \|^2, \tag{2}$$

where $MAA_{(a,p)}$ is the P$^{\text{th}}$ percentile value of MAA measured at a azimuthal eccentricity. Overall, our model can predict the minimum audible angle threshold with multiple fitted 2-degree polynomials.

$$F(a, \mathrm{P}) \coloneqq f_{\mathrm{P}}(a), \tag{3}$$

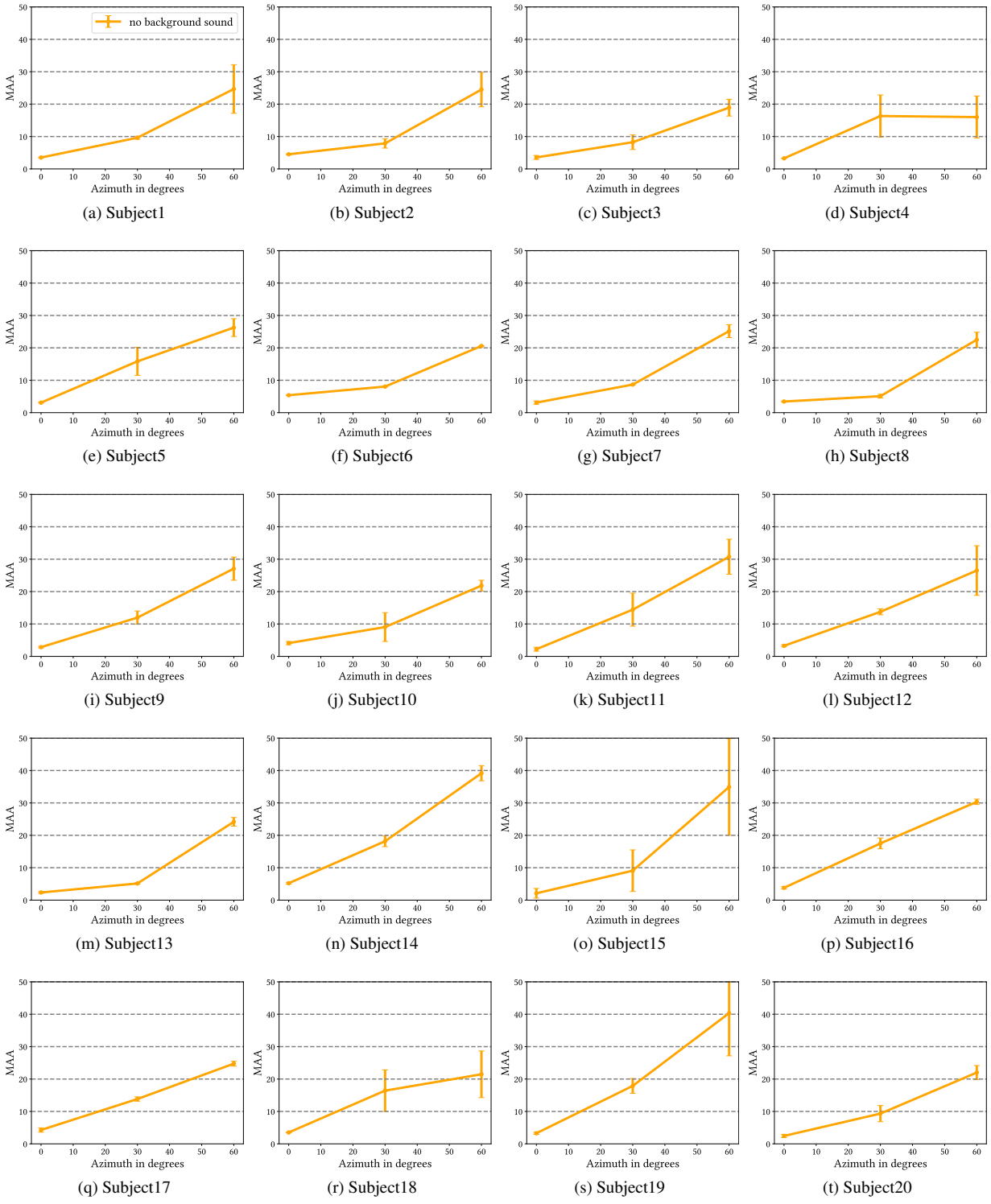


Figure 4: *Main study results per subject.* The main study (as described in Section 3.1) shows the measured MAA (Y-axis) with regard to azimuthal eccentricity (X-axis). The MAA increases as the azimuthal eccentricity is increased.

where F is the model, and a and P are the input azimuthal eccentricity and percentile index, respectively. Our model F is visualized at Figure 5.

Perceptual Sound Source Clustering Guided by the model, we further design a perception-based clustering algorithm to improve performance by reducing the number of audio sources over a large azimuthal eccentricity range. We assumed the user's focus is

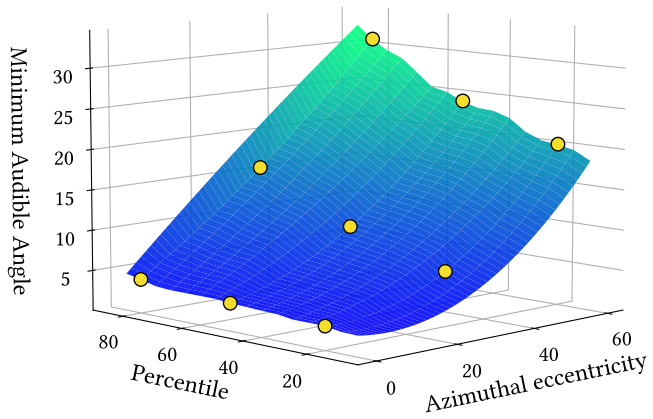


Figure 5: *Computational model.* A visualization of our model prediction of minimum audible angle interpolated across azimuthal eccentricities and percentiles. The measured data from our perceptual study at 20th, 50th, and 80th percentile are presented as yellow scatter points. (We plotted the points with higher z-order than the surface for easier visualization.)

Perceptual Sound Source Clustering

function
  P = GETPERCENTILE
  D = GETFOCUSDIRECTION
  for S_i in S_{list} **do**
    ▷ S_{list} includes all sound sources
    $S_i.angle = \cos^{-1}\left(\frac{(\vec{S_i}-\vec{L})\cdot \mathrm{D}}{\|\vec{S_i}-\vec{L}\|\|\mathrm{D}\|}\right)$
  end for
  $S_{list}.sort(key = \ lambda\ S_i : S_i.angle)$
    ▷ Sort S_{list} based on focus
  Initialize **Clusters**
  while S_{list} is not empty **do**
    Initialize C
    Take first S in S_{list}
    $T = F(a_S, \mathrm{P})$
      ▷ a_S is the azimuthal eccentricity of S
    for each other S' in S_{list} **do**
      if $a(S)$ is within T **then**
        Add S' to C
      end if
    end for
    Add C to **Clusters**
    Remove all sources in C from S_{list}
  end while
  return **Clusters**
end function

Figure 6: *Pseudocode of our audio clustering implementation.*

in the D direction. We cluster sound sources S_i according to how closely they align with D. Specifically, considering a sound source S_f which has minimum angular distance from D along the azimuth, we cluster all sound sources S_i with S_f if they are within the range predicted by our model,

$$a_{S_f} - \Delta\theta < a_{S_i} < a_{S_f} + \Delta\theta, \forall S_i : \Delta\theta = \frac{1}{2}F(a_{S_f}, \mathrm{P}) \quad (4)$$

where a_{S_i} is the audio source's azimuthal eccentricity. The detailed clustering implementation is illustrated as pseudocode in Figure 6. After determining the clusters, we characterized their spatial directions based on the sources within them. Similar to [59], we use the intensity v_i of an audio source as a criterion to place the source direction of cluster C,

$$a_C = \sum_i a_{v_i S_i},\ \phi_C = \sum_i \phi_{v_i S_i} \quad (5)$$

where a, ϕ are the audio source's azimuthal eccentricity and elevation to the listener, respectively. We apply this clustering algorithm in a large scale simulation, described in Section 6.

5 Application Case Study: Accelerating Audio Processing in VR Environment

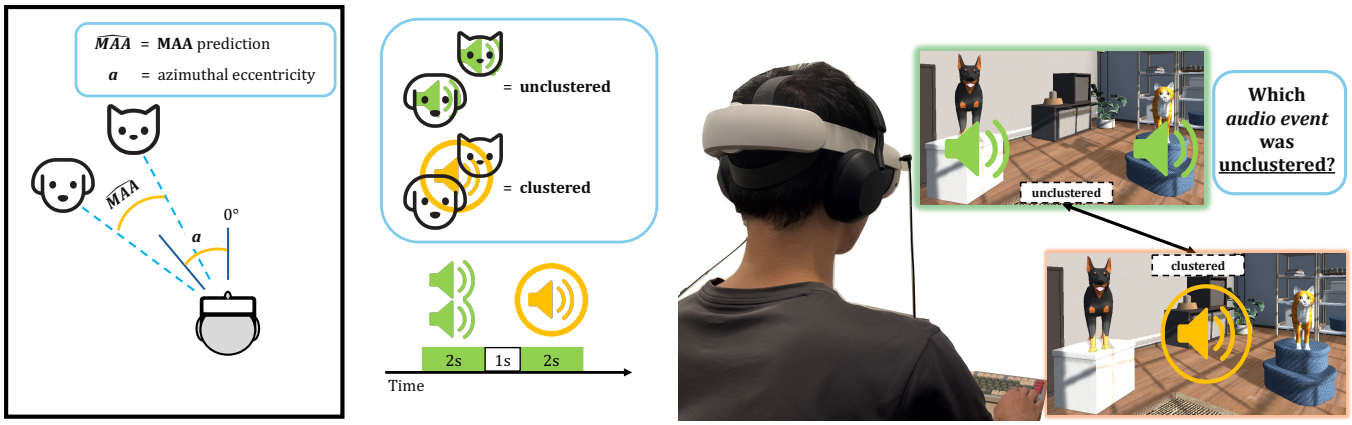
We evaluate the effectiveness and generalizability of our perceptual sound source clustering method (Figure 6) in a more realistic scenario by conducting a psychophysical study in a different environmental setup, protocol, and audio stimuli compared to that mentioned in Section 3. In particular, we measure users' ability to distinguish whether or not the audio sources are clustered.

Stimuli We simulated spatial audio with *pyroomacoustics* [49], which uses an image-source model used in the prior art for sound propagation prediction [21] and spatialized sound rendering in virtual environments [36]. We placed two microphones 16.2 centimeters apart (measured as the average head breadth of men in [62]) to simulate the distance between the two ears of participants. Visualized in Figure 7b, participants were placed in a virtual room that included a cat and a dog. The audio sources in the events were the cat meowing and the dog barking. The audio events were simulated with two conditions: **unclustered** and **clustered** (shown in Figure 7a). For **unclustered** audio event, the two audio sources were spatially separated relative to the participant with a given degree ($\widehat{MAA}$, based on our model prediction) and centered at a given eccentricity (a). See both in "*Conditions*". For **clustered** audio event, the two audio sources were clustered at a. Since the audio clustering algorithm runs per each interactive frame [51], our method applies equally irrespective of whether the sound sources and listeners are static or dynamic. Consequently, conducting our experiment in a static scene does not compromise its generalization to dynamic scenarios.

Participants and hardware Participants remained seated and perceived audio event using the same headphones as in Section 3.1. Users wore the Meta quest 3 to view the immersive virtual scene (as shown in Figure 7b) while completing the task. We recruited 12 participants (ages 22-27, 7 female, 5 male) to perform the study. None of the participants were aware of the research hypothesis.

Task As illustrated in Figure 7a and Figure 7b, the task was a two-interval forced-choice (2IFC). In each trial, participants sequentially experienced two 2-second sound events (one using **clustered** and another using **unclustered**) in a random order, with a 1-second pause in between. Participants were instructed to identify which event involved spatially separated sound sources (i.e., **unclustered**) by responding via keyboard.

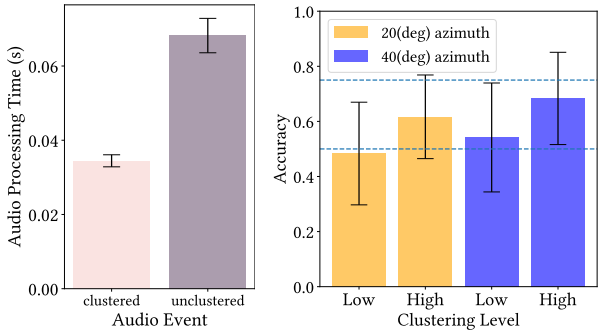
Conditions We included two azimuthal eccentricities ($a = 20°, a = 40°$) that were not presented in the pilot study to additionally test our model's ability to generalize to unseen conditions. As shown in Figure 7a. we characterized the angular distance between two audio sources by our model prediction $\widehat{MAA} = F(a, \mathrm{P})$ (i.e. cat-dog pair were placed at the limit of $\widehat{MAA}$). By leveraging the model, we used two percentile indices P to generate two levels of $\widehat{MAA}$ prediction, resulting in two clustering levels. At the low clustering level (**L-CLUSTER**), we used the P = 10 percentile



(a) application study setup

(b) application study visualization

Figure 7: *Application Study.* (a) shows the audio events in **clustered** and **unclustered** conditions. They are played sequentially with randomized order to the participant for each trial. Participants were instructed to distinguish which one of the audio events is **unclustered**, and gave the answer via keyboard. (b) visualizes the hardware and virtual scene. The study was performed using a Meta Quest 3 VR headset with Sony XM5 noise-canceling headphones. Notably, the visual stimuli remained unchanged across varying audio events and were spatially aligned with the **unclustered** condition for each trial.



(a) sound processing time

(b) results by clustering level ($\widehat{MAA}$)

Figure 8: *Application Study results.* (a) shows the mean audio processing time in **clustered** and **unclustered** audio events. The time consumption was reduced by 50% after clustering the audio sources. (b) shows the aggregated accuracy of participants correctly distinguishing the **clustered** and **unclustered** audio events. It compares between **L-CLUSTER** (Low) and **H-CLUSTER** (High). The 50% and 75% accuracy were presented as dotted lines in the figure. Refer to Figure 11 for individualized results.

index to predict a conservative $\widehat{MAA}$ threshold, in which most participants could not distinguish whether the audio was clustered in this range. At the high clustering level (**H-CLUSTER**), we used the $\text{P} = 90$ percentile index to provide a higher $\widehat{MAA}$ in which users are more likely to perceive which audio event was clustered or not.

The study had 40 (2 azimuthal eccentricities $\times$ 2 $\widehat{MAA}$ $\times$ 10 repeats) trials in total and took approximately 10 minutes for each participant to complete.

Results Figure 8b shows all 12 of the participants' aggregated 2IFC results. The mean accuracy of **L-CLUSTER** ($48.3 \pm 18.0\%$ at $a = 20°$, and $54.1 \pm 19.7\%$ at $a = 40°$) is near to random guess (50%) for the 2IFC task. A binomial test with 50% hypothesized probability shows the random guess null hypothesis cannot be rejected, with $p \gg .05$ ($p = .78$ at $a = 20°$, and $p = .41$ at $a = 40°$). In contrast, in **H-CLUSTER**, users were more likely to distinguish the clustered condition and obtain higher accuracy ($61.7 \pm 15.2\%$ at $a = 20°$, and $68.3 \pm 16.7\%$ at $a = 40°$). The binomial test evidenced that the results in **H-CLUSTER** are not based on random guess ($p = .01$ at $a = 20°$, and $p < .01$ at $a = 40°$). We also observed that the measured mean accuracy increases with azimuthal eccentricity. However, a one-way within-subjects ANOVA didn't show this effect to be significant ($F_{1,22} = 0.50, p = .48$ in **L-CLUSTER** and $F_{1,22} = 0.95, p = .33$ in **H-CLUSTER**). Visualized results for each subject are shown in Figure 11 As shown in Figure 8a, the processing time for *pyroomacoustics* to generate the audio of the **clustered** event (0.034 ± 0.002 seconds) is double that of the **unclustered** event (0.068 ± 0.004 seconds), and the aggregated mean accuracy is $58.12 \pm 19.2\%$.

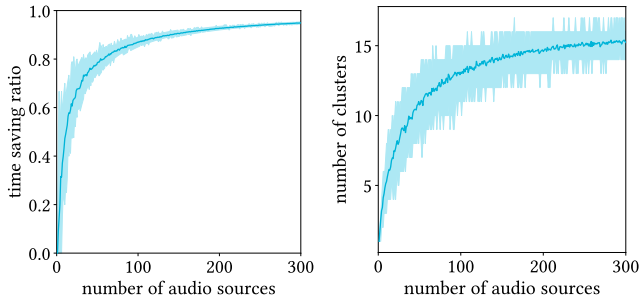
Discussion The statistical analysis indicates that participants were unable to distinguish between the original unclustered auditory space vs. the conservative clustering **L-CLUSTER** predicted by our model (10^{th} percentile index). While the more aggressive clustering **H-CLUSTER** became more distinguishable than random guesses, participants were still making uncertain choices (65%).

These observations reveal that our perceptually-guided auditory clustering can reduce computation time by 50%, *without significantly compromising the perceived auditory space*. Meanwhile, introducing the population percentile P to our model offers probabilistic and controllable guidance on balancing cross-population significance with computational savings. This controllable knob allows for the model's applicability in auditory rendering applications that are constrained by given computational resources, such as untethered VR/AR scenarios. Meanwhile, the differences in sound types (cat meowing and dog barking) validate the model's robustness to the various types of audio stimuli.

6 Simulated Performance Assessment

As verified by the user study in Section 5, our audio clustering method is able to efficiently reduce the number of audio sources without altering the user's audio perception. Here, we measure the method's performance with a large-scale audio source clustering simulation as a pressure test.

Audio source clustering simulation We randomly positioned audio sources around the listener and clustered the audio sources



(a) time saving ratio (b) number of clusters

Figure 9: *Performance Evaluation.* We simulated the usage of our perceptual sound source clustering method, detailed in Figure 6. The audio processing simulation was run 50 times, each time with randomly placed audio sources. (a) presents the range and mean of the time-saving ratio in our simulation, and (b) shows the range and mean of the number of clusters being grouped from our method, both as a function of the number of audio sources.

using our method, outlined in Figure 6. We assumed the listener's focus is on the zero azimuthal eccentricity direction in the simulation. The ratio r of time saved during audio propagation can be formalized as: $r = 1 - N_C/N_S$, where N_C and N_S are the number of clusters and sound sources, respectively. For each number of audio sources, we randomized directions of arrival of the audio sources 50 times to simulate all possible spatial audio arrangements. The results of this experiment are visualized in Figure 9.

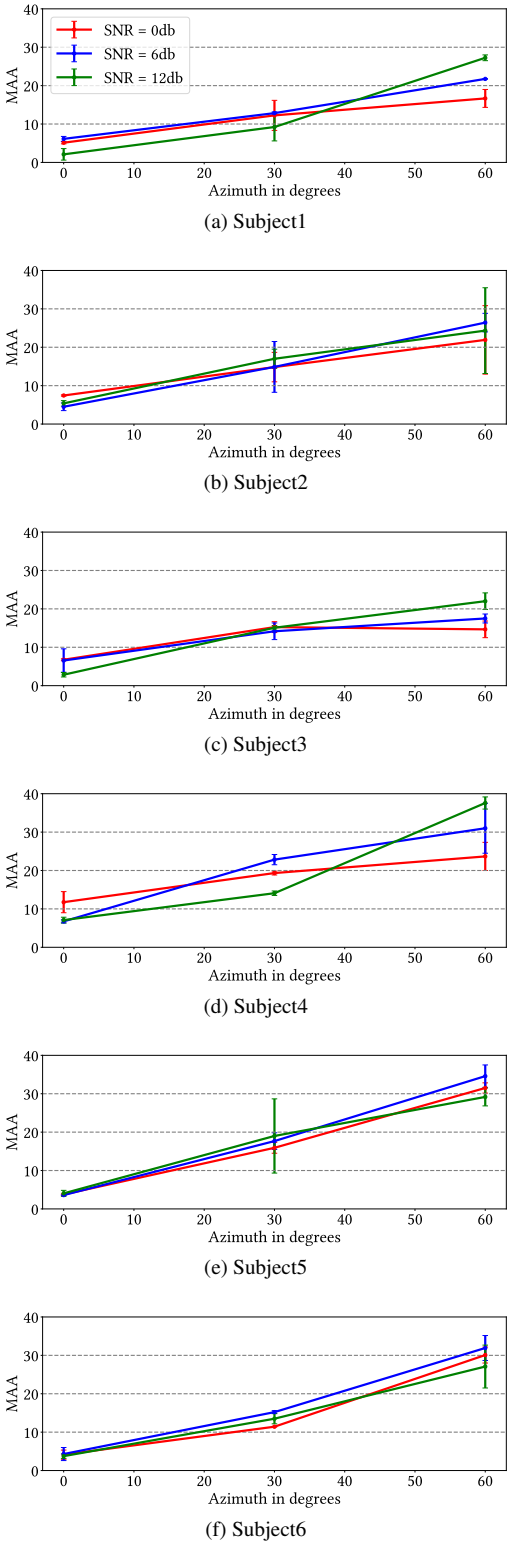
Our source clustering method's performance improves with the number of audio sources N_S, as shown in Figure 9a. The time-saving ratio r increases from $57.8 \pm 16.6\%$ at $N_S = 15$ to $94.92 \pm 0.3\%$ at $N_S = 300$. Figure 9b presents the number of clusters N_C being generated by our method across the number of sound sources. The results show that our method can efficiently reduce 300 sound sources to 15 clusters on average. We calculated the moving variance with a 10-step window to assess the convergence of N_C. The moving variance of the number of clusters N_C decreases under 0.1 after $N_S = 53$, indicating that N_C tends to converge to a stationary upper bound. It shows our method is able to merge a large number of sound sources into a controllable number of clusters.

7 Limitations and Future Work

Attention in audio targets. In our run-time audio source clustering algorithm (summarized in Figure 6), we assume the knowledge of listeners' focus. For example, a video game player may focus on the sound of an enemy's steps to determine their location. The heuristics provide us with the initial pivot points to perform spatial clustering. However, in task-free applications, it is uncertain how to determine where a user is attentive. In the future, we plan to investigate solutions for estimating auditory saliency [25, 60], which may shed light on adaptive weights of individual sources to guide the clustering algorithm.

Multi-order audio. We estimate the audio's direction based on azimuthal eccentricities without spatial reverb. Although this estimation measures the most significant order, the accuracy also depends on how reflective the environment geometry and materials are. Therefore, future work could consider the higher-order post-reverb audio eccentricities as a function of spatial layout [27] and barrier materials [47].

Additional auditory characteristics. While all audio sources were full spectrum white noise in Section 3, factors such as audio frequency and motion may also influence the perceptual sensitivity



(a) Subject1 (b) Subject2 (c) Subject3 (d) Subject4 (e) Subject5 (f) Subject6

Figure 10: *Generalizability study results per subject.* Generalizability study (as described in Section 3.2) shows the measured MAA in various SNR conditions. Raw data illustrates the measured MAA (Y-axis) with regard to azimuthal eccentricity (X-axis) per subject. The measured MAA is not significantly influenced by SNR.

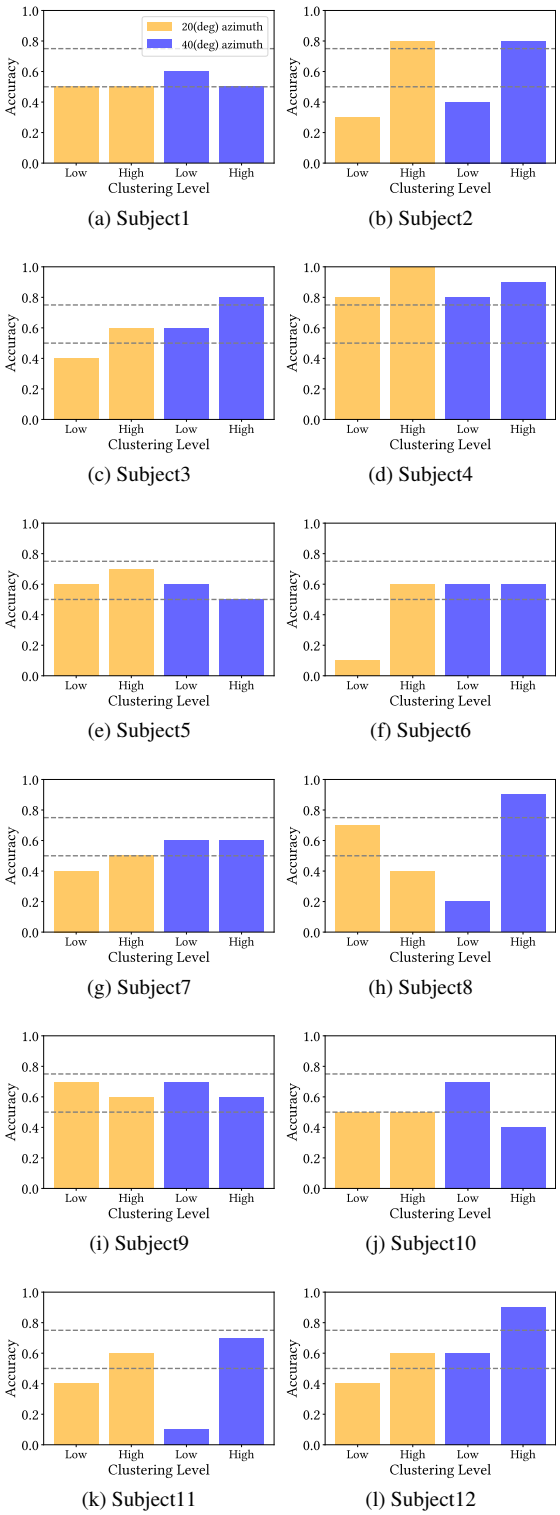


Figure 11: *Application study results per subject.* Application study (as described in Section 5) shows the accuracy (Y-axis) of participant's responses with regard to clustering level (X-axis) between **L-CLUSTER** and **H-CLUSTER**. In each plot, 2 bars on the left were measured at 20 degrees, with the others at 40.

to source direction. We envision future investigations of these factors may further enhance the accuracy of the modeling as well as clustering efficiency, but is beyond the scope of this work.

Age limitations of the user group Across the studies, the participants' ages were in the young adult range (18-31). [2] found that other age groups were less accurate at locating visual and auditory stimuli than young adults, indicating that our model could predict conservative MAA considering all ages. We will consider studying the MAA with a larger age range.

Selections of azimuthal eccentricity We studied the azimuthal eccentricity with a sample range ($a = 0°, \pm 30°, \pm 60°$) that aligns with visual field of view [22]. Extending the sample ranges of audio events out of the field of view or even behind the user is an exciting future direction, but beyond the scope of this research as the first attempt to transform the idea of "foveation" from visual to auditory.

8 Conclusion

In this paper, we present the first perceptually guided acoustic "foveation" approach that leverages the non-uniform spatial sensitivity across the azimuthal plane. We derived a computational model and a sound source clustering algorithm using the data collected from our psychophysical studies. Our clustering technique can save 50% on audio computation while not compromising listeners' perceived sound space evaluated in our user study. We hope the research will create future collaboration in the community to develop comprehensive and multimodal immersive systems that optimize for human perception, beyond the previously studied vision.

Acknowledgments

This work has been supported by National Science Foundation (NSF) grant #2107454.

References

[1] B. Arons. A review of the cocktail party effect. *Journal of the American Voice I/O society*, 12(7):35–50, 1992. 4

[2] M. M. Barrett and F. N. Newell. Task-specific, age related effects in the cross-modal identification and localisation of objects. *Multisensory research*, 28(1-2):111–151, 2015. 9

[3] E. Bernal-Berdun, M. Vallejo, Q. Sun, A. Serrano, and D. Gutierrez. Modeling the impact of head-body rotations on audio-visual spatial perception for virtual reality applications. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 2

[4] K. Brandenburg. Mp3 and aac explained. In *Audio Engineering Society Conference: 17th International Conference: High-Quality Audio Coding*. Audio Engineering Society, 1999. 2

[5] W.-P. Brinkman, A. R. Hoekstra, and R. van EGMOND. The effect of 3d audio and other audio techniques on virtual reality experience. *Annual Review of Cybertherapy and Telemedicine 2015*, pp. 44–48, 2015. 2

[6] C. R. A. Chaitanya, J. M. Snyder, K. Godin, D. Nowrouzezahrai, and N. Raghuvanshi. Adaptive sampling for sound propagation. *IEEE transactions on visualization and computer graphics*, 25(5):1846–1854, 2019. 2

[7] P. Chakravarthula, Z. Zhang, O. Tursun, P. Didyk, Q. Sun, and H. Fuchs. Gaze-contingent retinal speckle suppression for perceptually-matched foveated holographic displays. *IEEE Transactions on Visualization and Computer Graphics*, 27(11):4194–4203, 2021. 2

[8] E. Chang, X. Che, C. Qu, R. Xiao, D. Hu, and Z. Li. Where is the sound: User sound source perception in virtual reality environment. In *2023 IEEE Smart World Congress (SWC)*, pp. 1–8. IEEE, 2023. 2

[9] F.-Y. Chao, C. Ozcinar, C. Wang, E. Zerman, L. Zhang, W. Hamidouche, O. Deforges, and A. Smolic. Audio-visual perception of omnidirectional video for virtual reality applications. In *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 1–6. IEEE, 2020. 2

[10] C. Chen, C. Schissler, S. Garg, P. Kobernik, A. Clegg, P. Calamia, D. Batra, P. W. Robinson, and K. Grauman. Soundspaces 2.0: A simulation platform for visual-acoustic learning. In *NeurIPS 2022 Datasets and Benchmarks Track*, 2022. 2

[11] K. Chen, T. Wan, N. Matsuda, A. Ninan, A. Chapiro, and Q. Sun. Peapods: Perceptual evaluation of algorithms for power optimization in xr displays. *ACM Trans. Graph.*, 43(4), jul 2024. doi: 10.1145/3658126 2

[12] M. Cohen, S. Aoki, and N. Koizumi. Augmented audio reality: Telepresence/vr hybrid acoustic environments. In *Proceedings of 1993 2nd IEEE International Workshop on Robot and Human Communication*, pp. 361–364. IEEE, 1993. 2

[13] N. Deng, Z. He, J. Ye, B. Duinkharjav, P. Chakravarthula, X. Yang, and Q. Sun. Fov-nerf: Foveated neural radiance fields for virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3854–3864, 2022. 2

[14] B. Duinkharjav, K. Chen, A. Tyagi, J. He, Y. Zhu, and Q. Sun. Color-perception-guided display power reduction for virtual reality. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 41(6):144:1–144:16, 2022. 2

[15] R. Fan, X. Shi, K. Wang, Q. Ma, and L. Wang. Scene-aware foveated rendering. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 2

[16] T. Funkhouser, I. Carlbom, G. Elko, G. Pingali, M. Sondhi, and J. West. A beam tracing approach to acoustic modeling for interactive virtual environments. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pp. 21–32, 1998. 2

[17] M. Gerritse, M. Rietzler, C. Van Nimwegen, and J. Frommel. The effect of spatial audio on curvature gains in vr redirected walking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–10, 2024. 3

[18] T. M. Grieco-Calub and R. Y. Litovsky. Sound localization skills in children who use bilateral cochlear implants and in children with normal acoustic hearing. *Ear and hearing*, 31(5):645–656, 2010. 2

[19] R. Gupta, J. He, R. Ranjan, W.-S. Gan, F. Klein, C. Schneiderwind, A. Neidhardt, K. Brandenburg, and V. Välimäki. Augmented/mixed reality audio for hearables: Sensing, control, and rendering. *IEEE Signal Processing Magazine*, 39(3):63–89, 2022. doi: 10.1109/MSP.2021.3110108 2

[20] D. S. Hochbaum and D. B. Shmoys. A best possible heuristic for the k-center problem. *Mathematics of operations research*, 10(2):180–184, 1985. 2

[21] M. Hodgson. On the accuracy of models for predicting sound propagation in fitted rooms. *The Journal of the Acoustical Society of America*, 88(2):871–878, 1990. 6

[22] I. P. Howard and B. J. Rogers. *Binocular vision and stereopsis*. Oxford University Press, USA, 1995. 9

[23] D. L. James, J. Barbič, and D. K. Pai. Precomputed acoustic transfer: output-sensitive, accurate sound generation for geometrically complex vibration sources. *ACM Transactions on Graphics (TOG)*, 25(3):987–995, 2006. 2

[24] D. Jiménez Navarro, X. Peng, Y. Zhang, K. Myszkowski, H.-P. Seidel, Q. Sun, and A. Serrano. Accelerating saccadic response through spatial and temporal cross-modal misalignments. In *ACM SIGGRAPH 2024 Conference Papers*, pp. 1–12, 2024. 2, 3

[25] C. Kayser, C. I. Petkov, M. Lippert, and N. K. Logothetis. Mechanisms for allocating auditory attention: an auditory saliency map. *Current biology*, 15(21):1943–1947, 2005. 8

[26] B. Krajancich, P. Kellnhofer, and G. Wetzstein. Towards attention–aware foveated rendering. *ACM Transactions on Graphics (TOG)*, 42(4):1–10, 2023. 1, 2

[27] B. S. Liang, A. S. Liang, I. Roman, T. Weiss, B. Duinkharjav, J. P. Bello, and Q. Sun. Reconstructing room scales with a single sound for augmented reality displays. *Journal of Information Display*, 24(1):1–12, 2023. 8

[28] R. Y. Litovsky. Developmental changes in the precedence effect: Estimates of minimum audible angle. *The Journal of the Acoustical Society of America*, 102(3):1739–1745, 1997. 2

[29] R. Y. Litovsky, P. M. Johnstone, S. Godar, S. Agrawal, A. Parkinson, R. Peters, and J. Lake. Bilateral cochlear implants in children: localization acuity measured with minimum audible angle. *Ear and hearing*, 27(1):43–59, 2006. 2

[30] R. K. Maddox, D. A. Pospisil, G. C. Stecker, and A. K. Lee. Directing eye gaze enhances auditory spatial cue discrimination. *Current Biology*, 24(7):748–752, 2014. 4

[31] S. Malpica, A. Serrano, J. Guerrero-Viu, D. Martin, E. Bernal, D. Gutierrez, and B. Masia. Auditory stimuli degrade visual performance in virtual reality. In *ACM SIGGRAPH 2022 Posters*, pp. 1–2. 2022. 2

[32] R. Mehra, N. Raghuvanshi, L. Antani, A. Chandak, S. Curtis, and D. Manocha. Wave-based sound propagation in large open scenes using an equivalent source formulation. *ACM Transactions on Graphics (TOG)*, 32(2):1–13, 2013. 2, 4

[33] R. Mehra, A. Rungta, A. Golas, M. Lin, and D. Manocha. Wave: Interactive wave-based sound propagation for virtual environments. *IEEE transactions on visualization and computer graphics*, 21(4):434–442, 2015. 2

[34] X. Meng, R. Du, and A. Varshney. Eye-dominance-guided foveated rendering. *IEEE transactions on visualization and computer graphics*, 26(5):1972–1980, 2020. 2

[35] J. Merimaa and V. Pulkki. Spatial impulse response rendering i: Analysis and synthesis. *Journal of the Audio Engineering Society*, 53(12):1115–1127, 2005. 2

[36] J. D. Miller and E. M. Wenzel. Recent developments in slab: A software-based system for interactive spatial sound synthesis. In *Proceedings of the 2002 international conference on auditory display*, pp. 403–408, 2002. 6

[37] A. W. Mills. On the minimum audible angle. *The Journal of the Acoustical Society of America*, 30(4):237–246, 1958. 2, 4

[38] Q. Mo, H. Yeh, and D. Manocha. Tracing analytic ray curves for light and sound propagation in non-linear media. *IEEE transactions on visualization and computer graphics*, 22(11):2493–2506, 2015. 2

[39] T. Painter and A. Spanias. A review of algorithms for perceptual coding of digital audio signals. In *Proceedings of 13th International Conference on Digital Signal Processing*, vol. 1, pp. 179–208. IEEE, 1997. 2

[40] A. Patney, M. Salvi, J. Kim, A. Kaplanyan, C. Wyman, N. Benty, D. Luebke, and A. Lefohn. Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics (TOG)*, 35(6):1–12, 2016. 1, 2

[41] X. Peng, Y. Zhang, D. Jiménez-Navarro, A. Serrano, K. Myszkowski, and Q. Sun. Measuring and predicting multisensory reaction latency: A probabilistic model for visual-auditory integration. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 2

[42] D. R. Perrott. Concurrent minimum audible angle: A re-examination of the concept of auditory spatial acuity. *The Journal of the Acoustical Society of America*, 75(4):1201–1206, 1984. 2

[43] D. R. Perrott and K. Saberi. Minimum audible angle thresholds for sources varying in both elevation and azimuth. *The Journal of the Acoustical Society of America*, 87(4):1728–1731, 1990. 2, 3

[44] V. Pulkki. Spatial sound reproduction with directional audio coding. *Journal of the Audio Engineering Society*, 55(6):503–516, 2007. 2

[45] C. Rajguru, G. Brianza, and G. Memoli. Sound localization in web-based 3d environments. *Scientific Reports*, 12(1):12107, 2022. 2

[46] A. Ratnarajah and D. Manocha. Listen2scene: Interactive material-aware binaural sound propagation for reconstructed 3d scenes. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 254–264. IEEE, 2024. 2

[47] Z. Ren, H. Yeh, and M. C. Lin. Example-guided physically based modal sound synthesis. *ACM Transactions on Graphics (TOG)*, 32(1):1–16, 2013. 8

[48] I. R. Roman, C. Ick, S. Ding, A. S. Roman, B. McFee, and J. P. Bello. Spatial scaper: a library to simulate and augment soundscapes for sound event localization and detection in realistic rooms. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024. 2

[49] R. Scheibler, E. Bezzam, and I. Dokmanić. Pyroomacoustics: A python package for audio room simulation and array processing algorithms. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 351–355. IEEE, 2018. 2, 6

[50] C. Schissler and D. Manocha. Gsound: Interactive sound propaga-

tion for games. In *Audio Engineering Society Conference: 41st International Conference: Audio for Games*. Audio Engineering Society, 2011. 2

[51] C. Schissler and D. Manocha. Interactive sound propagation and rendering for large multi-source scenes. *ACM Transactions on Graphics (TOG)*, 36(4):1, 2016. 1, 2, 6

[52] P. Senn, M. Kompis, M. Vischer, and R. Haeusler. Minimum audible angle, just noticeable interaural differences and speech intelligibility with bilateral cochlear implants using clinical speech processors. *Audiology and Neurotology*, 10(6):342–352, 2005. 2

[53] R. Singh, M. Huzaifa, J. Liu, A. Patney, H. Sharif, Y. Zhao, and S. Adve. Power, performance, and image quality tradeoffs in foveated rendering. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 205–214. IEEE, 2023. 2

[54] Q. Sun, F.-C. Huang, J. Kim, L.-Y. Wei, D. Luebke, and A. Kaufman. Perceptually-guided foveation for light field displays. *ACM Transactions on Graphics (TOG)*, 36(6):1–13, 2017. 1, 2

[55] Z. Tang, R. Aralikatti, A. Ratnarajah, , and D. Manocha. Gwa: A large geometric-wave acoustic dataset for audio processing. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings (SIGGRAPH '22 Conference Proceedings)*, 2022. 2

[56] T. Tariq and P. Didyk. Towards motion metamers for foveated rendering. *ACM Transactions on Graphics (TOG)*, 43(4):1–10, 2024. 2

[57] T. Tariq, C. Tursun, and P. Didyk. Noise-based enhancement for foveated rendering. *ACM Transactions on Graphics (TOG)*, 41(4):1–14, 2022. 2

[58] N. Tsingos. Precomputing geometry-based reverberation effects for games. In *Audio Engineering Society Conference: 35th International Conference: Audio for Games*. Audio Engineering Society, 2009. 2

[59] N. Tsingos, E. Gallo, and G. Drettakis. Perceptual audio rendering of complex virtual environments. *ACM Transactions on Graphics (TOG)*, 23(3):249–258, 2004. 1, 2, 6

[60] T. Tsuchida and G. Cottrell. Auditory saliency using natural statistics. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 34, 2012. 8

[61] O. T. Tursun, E. Arabadzhiyska-Koleva, M. Wernikowski, R. Mantiuk, H.-P. Seidel, K. Myszkowski, and P. Didyk. Luminance-contrast-aware foveated rendering. *ACM Transactions on Graphics (TOG)*, 38(4):1–14, 2019. 1, 2

[62] K. P. Wankhede, V. P. Anjankar, M. P. Parchand, N. Kamdi, and S. T. Patil. Estimation of stature from head length and head breadth in central indian population: an anthropometric study. *Int J Anat Res*, 3(1):954–7, 2015. 6

[63] F. Wefers and M. Vorländer. Efficient time-varying fir filtering using crossfading implemented in the dft domain. In *Proceedings of the 2014 7th Medical and Physics Conference Forum Acusticum, Cracow, Poland*, pp. 7–12, 2014. 2

[64] Y. Zhang, K. You, X. Hu, H. Zhou, K. Kiyokawa, and X. Yang. Retinotopic foveated rendering. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 903–912. IEEE, 2024. 2