



Audio Engineering Society Conference Paper

Presented at the Conference on
Audio for Virtual and Augmented Reality
2018 August 20 – 22, Redmond, WA, USA

This paper was peer-reviewed as a complete manuscript for presentation at this conference. This paper is available in the AES E-Library (<http://www.aes.org/e-lib>) all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Acoustic perturbations in HRTFs measured on Mixed Reality Headsets

Andrea Genovese, Gabriel Zalles, Gregory Reardon, and Agnieszka Roginska

New York University, Music and Audio Research Lab

Correspondence should be addressed to Andrea Genovese (genovese@nyu.edu)

ABSTRACT

Materials that obstruct the path of acoustic waveforms in free-field to the human ear, may introduce distortions that can modify the natural Head-Related Transfer Functions. In this paper, the effect of wearing commercially available Head-Mounted Displays for Mixed and Augmented Reality has been measured via a dummy head mannequin. Such spectral distortions may be relevant for mixed reality environments where real and virtual sounds mix together in the same auditory scene. The analysis revealed that the measured HMDs affected the fine structure of the HRTF (> 3-6 kHz) and also introduced non-negligible distortions in the interaural level difference range mostly at the contralateral ear. Distortion patterns in HRTFs and cue modifications are reported and discussed across incidence angles and frequency bands.

1 INTRODUCTION

Mixed reality (MR) is defined as the merging of physical (real) and virtual sensory environments. MR environments are created by imposing virtual objects onto a physical environment (Augmented Reality (AR)) or physical objects onto a virtual environment using a Head-Mounted Display (HMD) [1]. The blending of real and virtual auditory environments is a complex problem that requires both room-acoustic matching and accurate audio spatialization.

Spatialization of audio objects in virtual auditory environments, such that the objects appear to be naturally occurring in the physical or virtual environment, is a large area of research. The placement of audio objects in 3D space is typically accomplished through applying location-dependent directional filters that transfer

human localization cues onto the audio content. These digital FIR filter-pairs are known as Head-Related Transfer Functions (HRTFs) and describe how an acoustic waveform at a point in space transfers to both ears of a listener. In a nutshell, the main localization cues contained within an HRTF pair consist of *Interaural Time Delays* (ITDs) for low frequency localization, *Interaural Level Differences* (ILDs) for high frequency localization and spectral cues for elevation and front-back discrimination [2].

Most of the research on HRTFs, and their implementation in virtual environments, has focused on reproduction in non-obstructed scenarios. However, in mixed reality reproduction settings, an HMD may obstruct the direct path of the physical waveform to the ears, possibly altering the spectrum of the sound received. This might affect the spatialization process in mixed

reality settings in two ways. Firstly, localization of real sources might be degraded because acoustical distortions are imposed upon the listener's natural localization cues. Secondly, spectral differences between the HRTF cues used to spatialize the virtual audio - and the now-altered natural localization cues of the listener - might diverge enough to prevent a plausible reproduction of 3D-audio content, even under optimal settings [3] [4]. Whenever the goal is to achieve an immersive mixed auditory environment, a coherent inclusion of the distortions caused by HMDs may be desired to allow a perfect blending of a virtual auditory environment with the current physical one [5, 3]. Such mixed reality scenarios require a transparent sound reproduction method as standard headphones would otherwise themselves occlude the path to the ear.

This work presents an investigation into the effect of MR HMDs on HRTF measurements. The HRTFs of a dummy head in free-field (no HMD) were measured and compared against those of a dummy head affixed with a Microsoft HoloLens and a Metavision Meta-2¹. The HRTFs and their differences were analyzed for 200 azimuth and 6 elevation positions around the head. This analytical approach to the problem can inform on whether further perceptual studies are worth to be carried out in order to determine the impact of the measured perturbations.

1.1 Psychoacoustics Background

Part of this study is to examine whether the spatial auditory cues modified by the presence of HMDs are objectively affected beyond the point of discrimination from the free-path case. Previous psychoacoustics literature provides reference *just noticeable difference* values (JNDs) that indicate the perceptual threshold at which physical differences in signals become audible. Human spatial hearing depends on the anatomical features of the head and body [2]. These features obstruct the path of a waveform in free space and result in two different waveforms received at each ear. The ITD and ILD are the main cues for localization in the horizontal-plane and are optimally tuned for low-frequency localization (< 1.5 kHz) and high-frequency localization (> 2 kHz), respectively [2].

The ITD cue refers to differences in the time arrival of signals at ear. The rigid nature of the head (and body)

diffracts and scatters the incident waveform and results in frequency-dependent microsecond differences between waves received at each ear [6]. The JND is as small as 10 μ s for pure tone frequencies below 1.5 kHz. For pure tones above 1.5 kHz, differences in ITD are usually not perceptible due to spatial aliasing [7, 8]. However, listeners can, in particular cases, still be sensitive to the ITD of high-frequency non-pure tones. For instance, in [9], the ITD JND of narrowband noise centered at high-frequencies was found to be substantially affected by the simultaneous presence of narrowband noise centered at low-frequencies.

Generally, as a sound source moves away from the front and back of the head towards the side, the ILD increases because the head casts an acoustic shadow on the contralateral ear (the ear farther away from the sound source). This shadow is frequency-dependent, with the head posing a small obstacle to large wavelengths and a much more significant obstacle to smaller wavelengths [10]. The JND for changes in ILD is around 0.5 dB - 1 dB and is frequency-independent between approximately 200 Hz and 10 kHz [11, 12].

The outer ear acts as a direction-dependent filter and helps the auditory system to discriminate elevation, distance, and sources located along auditory *cones of confusion*, where the ITD and ILD cues are in theory nearly constant [2]. Specifically, sound reflects and scatters off of and resonates with the outer ear and its various folds and cavities (including the pinnae and its resonant cavity the concha) resulting in intricate spectral modifications on the signal received at the ear. Given the structure of the outer ear, spectral cues are most useful in the range of 4 kHz to 17 kHz [10]. These cues can typically be evaluated using only one ear (*monaural cues*) [13] and are key elements of vertical localization perception [2].

An HRTF filter embeds the composite of all the complex interactions of the waveform with the head on its path to each ear. HRTF measurements are typically gathered from real listeners (individual HRTFs) or using dummy-head microphones (generic HRTFs). A dummy-head microphone is shaped to match the main anatomical features of a human head and has microphones placed inside the artificial ears. While dummy-heads only average the population of natural HRTFs, they are still very useful for generating HRTF sets for use in mass-distributed virtual auditory environments and to study the acoustic field environment of the head and near-head.

¹Latest developer models as of December 2017

2 PREVIOUS WORK

2.1 Distortion metrics

Two metrics have been used to understand the measured HRTF distortions on a dummy head affixed with MR HMDs. These two metrics are the *HRTF Difference* (HRTFD) and *Spectral Distortion* (SD). The HRTFD has been defined as the quotient of magnitude spectra of two HRTFs from the same direction and has been found to appropriately characterize differences between HRTFs [13]. The HRTFD is useful for understanding deviations in HRTFs due to obstructions within the acoustic environment as the contribution of the specific dummy-head are divided out. A slightly modified version used for this study is, computed as follows:

$$\text{HRTFD}(\phi, \delta) = 20 \log_{10} \left(\frac{|\text{HRTF}_{\text{mod}}(\phi, \delta)|}{|\text{HRTF}_{\text{ref}}(\phi, \delta)|} \right) \quad (1)$$

Where $|\text{HRTF}(\phi, \delta)|$ is the frequency magnitude response of the HRTF located at azimuth ϕ and elevation δ . HRTF_{ref} is the free-field dummy-head HRTF measurement and HRTF_{mod} is the HRTF of the dummy-head that has been altered (with an HMD in the context of this work). The *Spectral Distortion* (SD) is a metric that has been used to study differences between the structures of different HRTFs sets [14] and is a global measure of the spectral differences in dB of a HRTFD within a given frequency range. The SD is defined as follows:

$$\text{SD}(\phi, \delta) = \sqrt{\frac{1}{K} \sum_{k=1}^K (\text{HRTFD}(\phi, \delta, k))^2} \quad \text{dB}. \quad (2)$$

Summarizing spectral distortions over multiple frequency bands is useful when exploring and characterizing the differences between HRTF datasets [15]. Both metrics can be used to understand the perturbations in acoustical measurements as due to wearable HMDs.

2.2 Previous Studies

Past authors have looked at the effect of the near-head acoustic environment on HRTFs. Wersényi and Illényi [16] measured the HRTFs of a dummy-head microphone in free-field and when fitted with glasses, baseball caps, and toupees. No differences between the HRTFs were found below 1.5 kHz. Significant spectral

distortions, as high as 20dB, were found behind the head at low elevations ($\delta = 0 - 15^\circ$) for the baseball cap. Each of the objects amplified and dampened the height of existing peaks and valleys of the HRTF. The HRTFD for ipsilateral positions was mainly affected in the high-frequency range (> 3 kHz). The authors also reported a dB amplification of up to 10dB in mid-frequency content (1.5 kHz - 3 kHz) of the HRTFD at the acoustic “bright spot”. The bright spot is the contralateral response to a source located around 100° on the horizontal plane, where the diffracted waveforms combine in phase, resulting in a local maximum [6]. Thus, the explored wearables displayed an effect on the scattering behavior of the dummy head in key frequency zones necessary for median-plane vertical localization.

Treeby et al. [17] studied the effect that human hair had on the ITD and ILD of measured HRTFs in the median plane. They found that the hair produced asymmetries in the ITD and ILD of the measured HRTFs of up to 4dB for the anterior and posterior hemispheres. Further, the hair introduced asymmetries in the HRTF at the positions adjacent to the bright spot, with the anterior side of the position having more pronounced spectral notches than the posterior. Differences of up to 10dB at the contralateral ear for frequencies in the ILD range (> 1.6 kHz) due to hair have been reported, but this effect was much less pronounced at elevations above 30° [16]. Similar findings are reported in [18] where hair and hats were found to have an effect above 7 kHz and 5 kHz respectively. None of these studies performed a direct listening experiment to measure the perceptual impact of their findings.

3 HRTF Measurements

In order to accomplish the measurements present herein, a stepper motor with a custom microphone boom attachment was utilized. The stepper motor was controlled via an Arduino UNO and MATLAB using the Psychtoolbox [19] and the Arduino MATLAB support package. The *ScanIR* tool [20] was modified to allow for multi-speaker measurements in tune with automatized motor rotations. Namely, six Genelec 8030A speakers were set up in a spiral configuration of intervals of 22.5° in the azimuthal plane and 18° in the median plane, for 1 meter radial distance to a Neumann KU-100 dummy head. A MOTU UltraLite MK3 Hybrid interface was used to send a 2-seconds

sinesweep to each of the six speakers and to record the binaural signals at 96 kHz from 20 Hz to 20 kHz. All speakers were calibrated at 85 dBC SPL.

The motor was set to the finest available resolution of 1.8° per step, for a total of 200 steps. This resulted in a total of 1200 measurements per dataset (200 azimuths \times 6 elevations). Firstly, a set of KU-100 measurements without any HMD were performed, this allowed for later calculation of the HRTFD with and without HMDs. Several AR/VR HMDs were placed to the best ergonomic tight fit possible and the HRTFs were again measured - this paper is concerned only with the measurements performed on the Hololens and Meta-2 mixed-reality headsets². All measurements were conducted at the NYU MARL research lab facility. The same room was measured in [21] where a T_{60} of 0.14 seconds was found. The dimensions of this rectangular room are 4.5x3.7x2.5 m. Measurement was repeated twice and averaged in order to improve the SNR. Each HMD measurement took approximately 6 hours.

The final sets of measurements were saved as minimum-phase filters, processed for correct angle-alignment and normalized to the loudest measurement of all data sets combined. For the purpose of this paper, the collected HRIRs were first truncated to 256 samples and scaled by the second half of a Tukey window with a 25% cosine taper amount [17] to remove potential reflections. HRTF data was computed over the windowed right-ear HRIRs (NFFT = 8192) and smoothed over logarithmic fractional 12th-octave-bands using the algorithm indicated in [22].

4 Results

4.1 Preliminary Analysis

A rough first analysis on the resulting spectra permitted to immediately identify whether any change could be detected at all within frequency ranges of interest at sampled positions, guiding further investigations. The sampled position in fig. 1 shows how the general spectral behavior of acoustic distortions are not expected to diverge in the lower frequency-range until about 1.5 kHz, thus beyond the main ITD-range. Similar patterns were found in all measurements regardless of the azimuth and elevation source position.

²The HTC Vive, GearVR and Oculus Rift VR HMDs were also measured. Data to be released and discussed in a future publication.

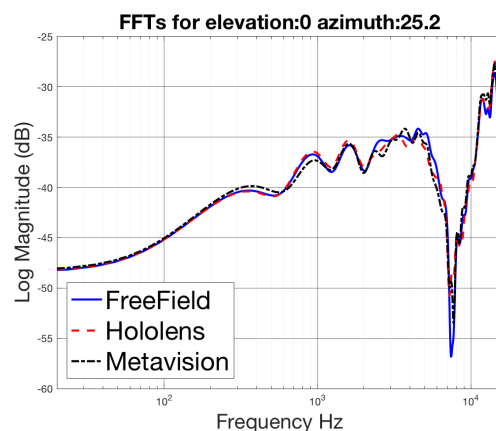


Fig. 1: Spectrum comparison for a sample position

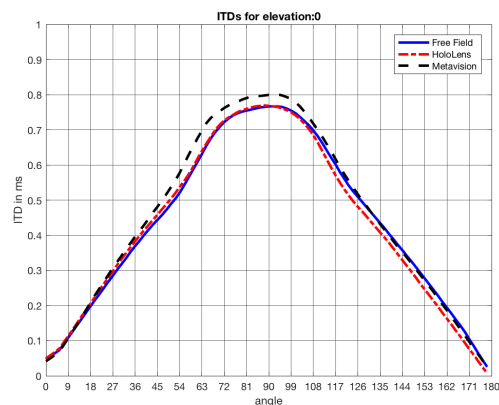


Fig. 2: ITD levels for the horizontal plane

The effects on ITDs were further explored by computing time delay curves on the horizontal plane. ITDs were computed as the maximum inter-aural cross-correlation (*maxIACC*) on the windowed broadband HRIRs low-pass filtered at 3 kHz [23]. The IACC returns the index of maximum coherence which is in turn converted to milliseconds. Results (fig 2) showed that the ITD cues were not affected for the Hololens HMD (barring negligible deviations of about $2\mu\text{s}$), while the Meta-2 presented a slight divergent behaviour starting from 45° to 110° with a peak difference of about $30\mu\text{s}$ occurring at $\phi = 100^\circ$. The ITDs of the Meta-2 HMD at all other elevations presented similar divergent behavior. However, no divergence at all was found for ITDs calculated on HRIRs low-passed at a 1.5 kHz cutoff frequency, isolating the range where the ITDs

form the dominant localization cue.

4.2 HRTFD analysis

The right-ear HRTFDs (expressed in dB) were obtained via equation (1), setting the free-field measurements as the reference HRTFs and the HMD measurements as the altered versions. An initial visualization on the absolute HRTFD permitted to confirm the findings of the preliminary analysis and also to identify the approximate frequency regions of maximum perturbation. From the horizontal plane analysis of the absolute HRTFD, fig. 3, it is clear that the majority of the differences occur beyond the 2 kHz mark. Similar trends in both HMDs show higher perturbations in the anterior contralateral region (between 270° and 360° , where most of the physical interactions supposedly occur) at frequencies > 4 kHz, up to approximately 16dB. From table 1, the HRTFD has been highlighted to be consistently more severe at the contralateral anterior region ($270^\circ - 360^\circ$) for the 4-8 kHz octave band. High ipsilateral posterior ($90^\circ - 180^\circ$) perturbations are also present in the range of 3 to 6 kHz, peaking around the 100° azimuth mark. Interestingly, both HMDs present a high azimuth-independent HRTFD between 7-8 kHz immediately followed by a relatively unaffected area spanning from 9 to 14 kHz (with the exclusion of the anterior contralateral angles). As found in [17], a lower degree of change can be seen at the acoustic “bright spot” in the contralateral region, here found as a local minimum about $\phi = 270^\circ$. Another minimum is found at the cross-median posterior region (centre-back of the head around $\phi = 180^\circ$) where the direct path of the source to the ears is minimally occluded by the wearable devices.

Similar HRTFD curves as those shown in [16] were produced to illustrate the spectral behavior within a close azimuth span and inspect on whether the measured deviations were caused by a spectral notch or a spectral peak at the locations of interest. Sub-figures 4 a) and b), show the raw HRTFD across a section of the ipsilateral posterior range (source located in the region spanning from $\phi = 90^\circ$ to 180° degrees, the plots show up to 150°). Both headsets show a frequency boost in the range of 3.5 to 5.5 kHz and a fluctuation behavior between 7-8 kHz. Sub-figures c) and d) instead show the behavior for a section of the contralateral anterior region (region spanning $\phi = 270^\circ$ to 360° degrees, the plots show up to 330°). In this case the effect in both

HMDs is generally that of an attenuation pattern between 4 to 6 kHz followed by a sharp boost between 7-8 kHz.

Other elevation planes show more pronounced absolute peaks that reach up to 23dB of absolute change (see Table 1 and Appendix A and B for a more comprehensive view). In respect to the contralateral anterior region, lower elevations also present a ~ 7 -8 kHz cross-azimuth peak in both headsets while increasing the boosting behaviour, shifting it towards lower frequencies. Higher elevations would instead reduce the 7kHz peak and cause a higher degree of dampening, as well as reducing the general boost behaviour. Differently from the Hololens, the Meta-2 seemed to be affected more at high frequencies > 9 kHz and shows higher dampening in the posterior range from 110° to 250° at higher incidence angles.

4.3 Spectral Distortions

Spectral distortions (SD) were first calculated over select octave-bands of interest, namely 1-2 kHz, 2-4 kHz, 4-8 kHz and 8-16 kHz, using the formula given in equation (2). This metric is useful to have a more robust measure of the magnitude of distortion within particular frequency bands, by averaging spurious peaks in the HRTFD. Results were inspected through several visualizations capable of highlighting potential patterns or perturbations. Figure 5 illustrates the spectral distortions plotted across azimuth-elevation maps for selected frequency octave-bands, in this case the 2-4 kHz band and 4-8kHz band³. The 8-16 kHz octave band was also affected albeit to a lesser degree than the 4-8 kHz octave while the SD levels between 1-2 kHz appeared negligible at every location for both HMDs.

In general, a high azimuth-dependency pattern is easily discernible at all elevation levels. All figures again pointed to the fact that fewer disturbances are introduced at the ipsilateral positions than the contralateral azimuth range, regardless of octave band and elevation level. Another discernible observation is that the anterior contralateral azimuth range ($270^\circ - 360^\circ$) is the most affected region in all situations. The two sets differ slightly in terms of distortions in the higher bands and behavior at the extreme elevation levels, with the Hololens showing more distortions in the 4-8 kHz band

³Appendix C. illustrates the SD on a refined narrow-band linear frequency zoom (1 kHz-wide bands from 2 to 6 kHz) to further explore the most affected frequencies through polar visualizations.

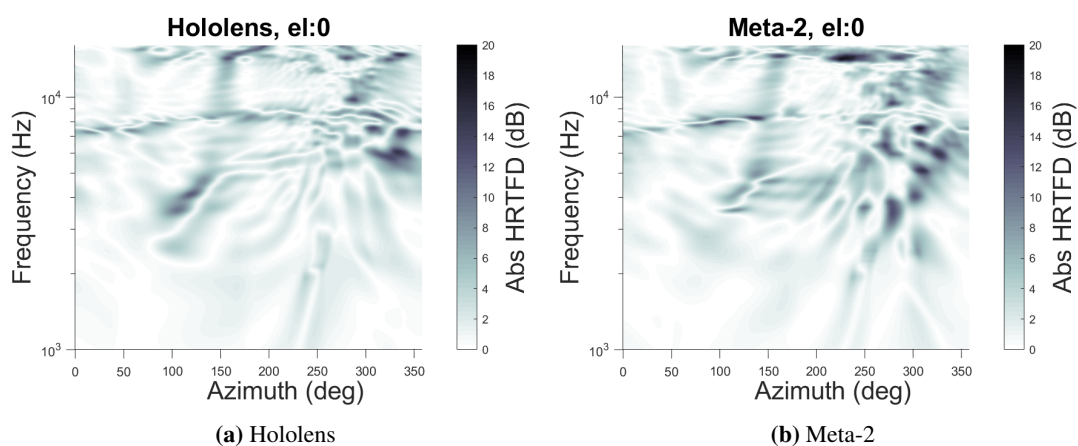


Fig. 3: Absolute HRTFDs across the horizontal plane

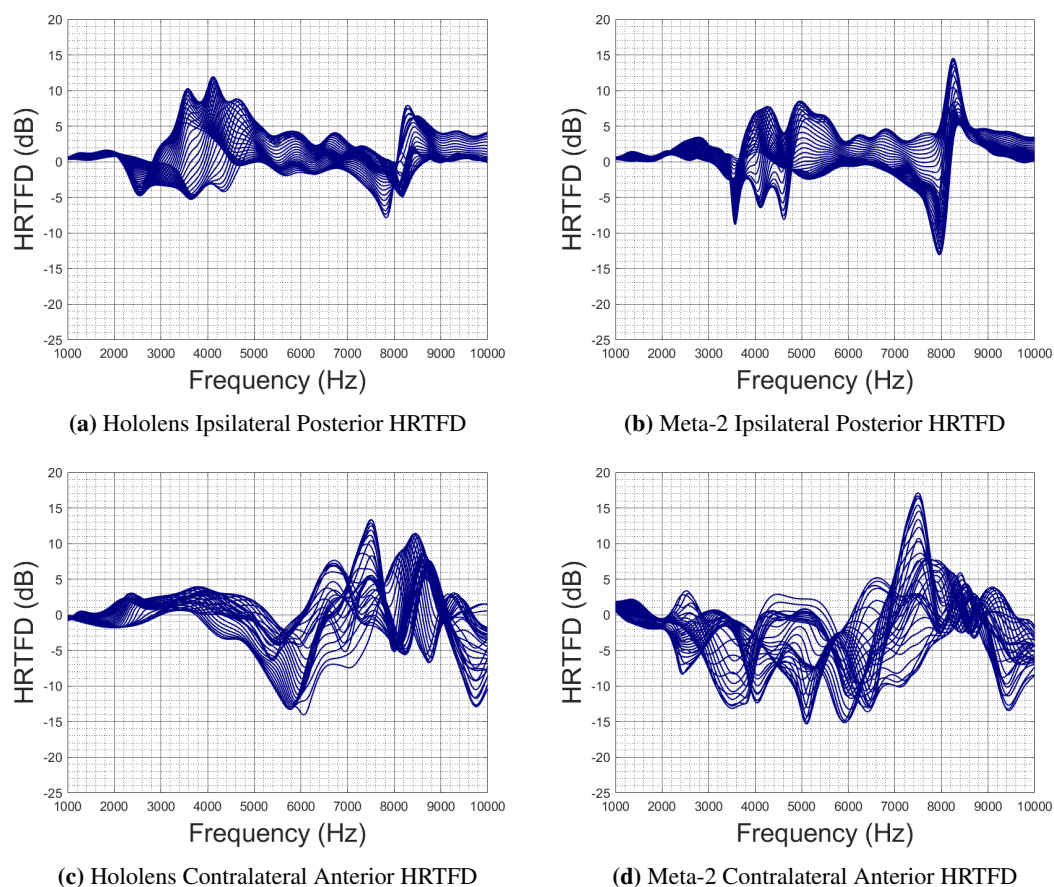


Fig. 4: HRTFDs for two sensitive azimuth angle ranges on the horizontal plane. Top row: subset within the Ipsilateral Posterior region $\phi = 90^\circ - 150^\circ$. Bottom row: subset within the Contralateral Anterior region $\phi = 270^\circ - 330^\circ$

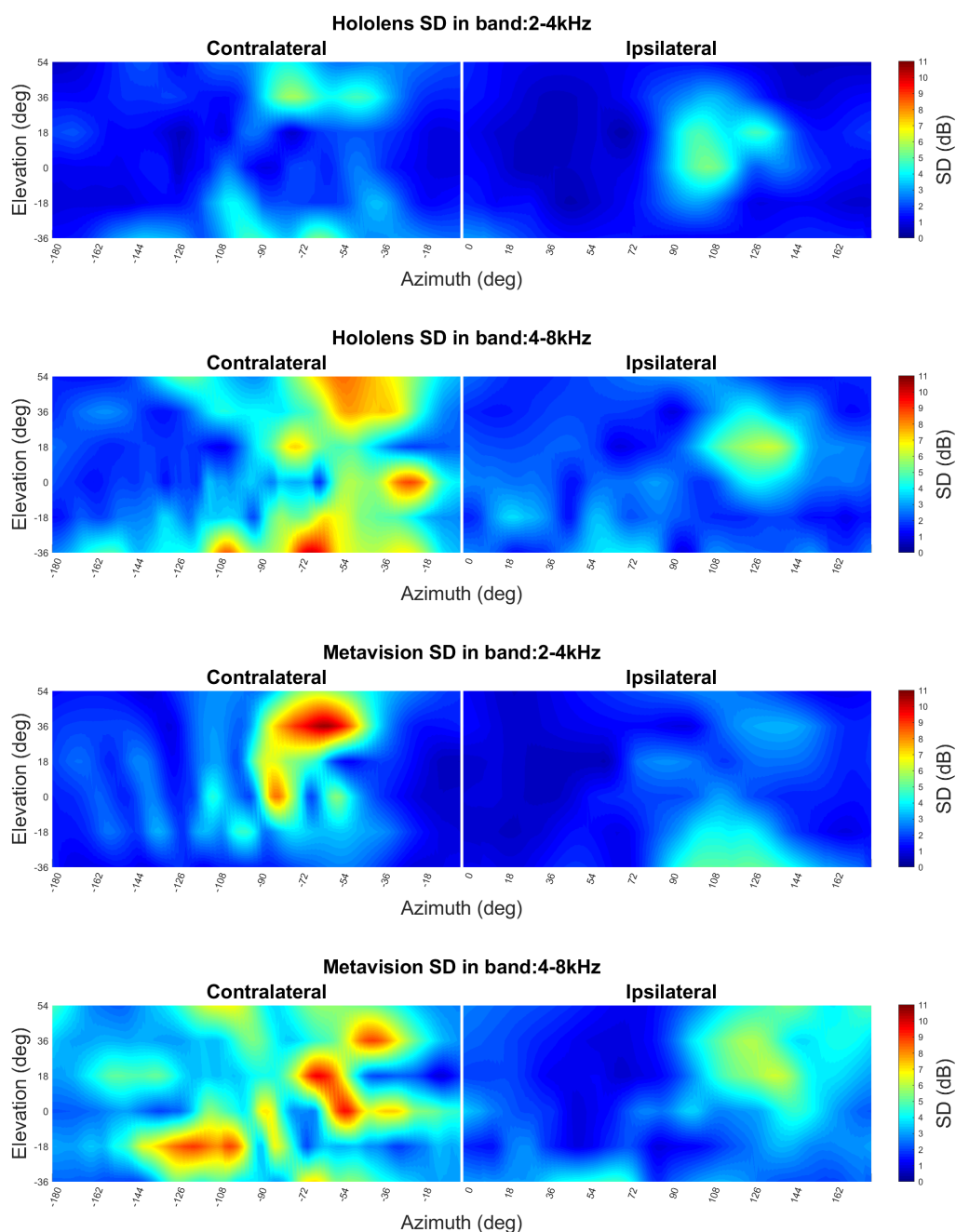


Fig. 5: Spectral Distortion maps for two select octave-band frequency ranges spanning azimuth and elevations. Top two graphs: Hololens. Bottom two graphs: Meta-2. Each graph represents the SD from the right ear perspective for source-locations folding around the head from the center white line marking the median plane.

Table 1: Distortion summary table for the right-ear recordings with the two HMDs in selected octave bands and azimuth ranges. The source-location azimuth ranges are divided into four quadrants: *Ipsilateral Anterior* (IA, $0^\circ - 90^\circ$), *Ipsilateral Posterior* (IP, $90^\circ - 180^\circ$), *Contralateral Posterior* (CP, $180^\circ - 270^\circ$) and *Contralateral Anterior* (CA, $270^\circ - 360^\circ$). Reported values comprise peak and spherical location of the HRTFD and Spectral Distortion (SD) (in terms of azimuth ϕ and elevation δ), and the mean and standard deviation of the SD across the specified azimuth quadrant. Most of the bigger distortions (marked in boldface) are seen to occur in the contralateral anterior region. Generally, the most affected octave band is that within 4-8 kHz for both HMDs. The effect of elevation seems to be weak for one of the headsets indicating that the impact of this factor might correlate with specific physical dimensions.

HMD	Az.Range	Freq.Range	HRTFD (dB)			Spectral Dist. (SD) (dB)				
			peak (dB)	ϕ_{Peak}	δ_{Peak}	max (dB)	ϕ_{max}	δ_{max}	mean	std.
HL	IA	1-2 kHz	1.69	0°	18°	1.00	0°	-18°	0.39	0.21
		2-4 kHz	-6.02	0°	-36°	3.05	88.2°	0°	1.21	0.54
		4-8 kHz	-13.77	21.6°	-18°	3.9	66.6°	-36°	2.26	0.65
		8-16 kHz	13.12	57.6°	-36°	4.71	55.8°	-36°	1.73	0.84
	IP	1-2 kHz	2.38	178.2°	18°	1.5	167.4°	0°	0.81	0.34
		2-4 kHz	-12.16	135°	18°	5.38	106.2°	0°	1.97	1.23
		4-8 kHz	14.19	113.4°	18°	6.25	133.2°	18°	2.57	1.13
		8-16 kHz	12.66	158.4°	0°	5.89	158.4°	0°	2.25	1.1
	CP	1-2 kHz	-4.71	261°	18°	3.21	259.2°	18°	1.21	0.57
		2-4 kHz	-8.33	266.4°	-36°	4.65	266.4°	-36°	1.79	0.85
		4-8 kHz	15.59	261°	-36°	8.57	255.6°	-36°	2.85	1.37
		8-16 kHz	12.64	196.2°	0°	6.12	255.6°	18°	2.9	1.08
	CA	1-2 kHz	-3.85	280.8°	54°	1.93	270°	18°	1.03	0.4
		2-4 kHz	-15.14	295.2°	-36°	5.81	284.4°	36°	2.56	1.19
		4-8 kHz	-21.81	304.2°	54°	10.04	293.4°	-36°	5.28	1.88
		8-16 kHz	18.64	318.6°	36°	8.21	318.6°	36°	4.2	1.73
MV	IA	1-2 kHz	1.56	88.2°	-18°	1.12	88.2°	-36°	0.47	0.22
		2-4 kHz	-10.3	88.2°	-36°	3.45	88.2°	-36°	1.23	0.55
		4-8 kHz	12.1	68.4°	-36°	4.65	66.6°	-36°	1.99	0.91
		8-16 kHz	13.07	55.8°	-36°	5.16	54°	-36°	1.89	0.98
	IP	1-2 kHz	2.59	178.2°	18°	1.45	122.4°	-36°	0.85	0.33
		2-4 kHz	-17.01	131.4°	-36°	5.13	124.2°	-36°	2.52	1.09
		4-8 kHz	-20.87	140.4°	54°	6.26	136.8°	18°	3.47	1.19
		8-16 kHz	17.27	158.4°	-18°	5.84	135°	54°	2.99	1.3
	CP	1-2 kHz	5.33	237.6°	0°	3.53	268.2°	54°	1.17	0.51
		2-4 kHz	-10.31	250.2°	0°	5.08	268.2°	18°	2.07	0.77
		4-8 kHz	15.99	243°	-18°	9	241.2°	-18°	4.06	1.68
		8-16 kHz	18.5	226.8°	0°	6.9	253.8°	18°	3.78	1.4
	CA	1-2 kHz	-8.97	288°	54°	4.55	284.4°	54°	1.39	1.07
		2-4 kHz	-19.22	284.4°	36°	10.65	298.8°	36°	3.28	2.35
		4-8 kHz	-23.33	293.4°	18°	9.68	297°	18°	4.58	1.94
		8-16 kHz	-16.64	293.4°	18°	7.26	333°	36°	4.02	1.45

at $\delta = -36$ and $\delta = 54$ than $\delta = 0$. Higher elevations seemed to have a more pronounced effect than lower elevations. When putting the SD plots into context with the HRTFDs (appendix B.) the deviations seen in the upper locations seem to stem from attenuation. Appendix D. shows SD boxplots across the contralateral azimuth range. It suggest that for almost all elevation rings the majority of severe distortions are contained within the 4-8 kHz band and 8-16 kHz octave band.

Table 1 provides a summary of the SD levels across selected octave bands, their location within quadrant macro-regions, and the mean across the quadrant. The highest perturbations (> 7 dB) are highlighted. Both the table and a deeper narrow-band decomposition of 1-kHz wide bands (Appendix C.), which calculated the SD over smaller ranges, confirmed the regions of highest perturbation to be the ipsilateral posterior (IP) and contralateral anterior (CA). In Appendix C. the magnitude of the SD is shown to often reach a high level 6-10dB for the IP quadrant, and 11-13dB for the CA quadrant (with a sharp peak of 17dB in a particular case of the Meta-2, fig C.(j)). The most affected bands consistently were those of 4-5 kHz and 5-6 kHz.

5 Discussion

Although this paper looked at two particular headset devices, the study is intended to bring attention to the change in the acoustic field that derives from wearable rigid AR/MR headsets placed on the head. Generally, both headsets will highly interact with the wavelengths that match the dimensions of the central display and rigid head-mount, reflecting, diffracting and perhaps resonating with the incident waveforms. The specific frequencies and locations of the heavier perturbations may change accordingly to the physical dimensions of the device components, but the general behaviour is expected to be similar for any HMD following the same kind of design concept which implies transparent sound-reproduction methods.

It is unlikely that the ITD cue modifications introduced by the HMDs are severe enough to be perceptually relevant (fig: 2). Although the peak divergence amount for the Meta-2 is slightly beyond the JND values for pure tones > 1 kHz [7], the same divergence does not occur below 1.5 kHz where the ITDs are the dominant cue for spatial localization. Even if the incoming waveforms were to present high energy around the sensitive frequencies and locations, the modifications would not

consist in more than few degrees in areas where the perceptual resolution of localization is already coarse due to *localization blur* effects [24, 25].

In agreement with previous studies about the effect of near-head obstructive materials on the acoustic field [18, 16, 17], the HMDs start to introduce acoustic perturbations to waveforms starting from frequencies > 3 kHz. Although other authors methodologies are not directly comparable, results suggest that much higher near-head acoustic perturbations were caused by MR HMDs as opposed to hair [17] and clothing items [16]. This is not unexpected given the rigid body materials of HMDs compared to hair or hats. The most affected identified areas are the anterior contralateral range $270^\circ - 360^\circ$ and, to a lesser degree, the lateral-posterior ipsilateral range $90^\circ - 150$. Although the elevation resolution is not refined enough to derive clear patterns or relationships, it looks like lower elevations cause a boost from ipsilateral posterior source-incidence angles, possibly because of constructive reflections bouncing from the headset headband or perhaps resonance from the front glass piece. Higher elevations do instead cause higher attenuation at the contralateral anterior angles. This finding was consistent with the expectations given that the direct path at those incidence angles from the source to the contralateral ear is considerably physically obstructed by the frontal kernel of the HMDs. This effect is accentuated on the Meta-2, which possesses a bulkier and heavier body compared to the HoloLens HMD.

The most affected octave bands were those spanning from 2 to 3 kHz and 4 to 8 kHz. Close inspection of the HRTFD revealed that for some incidence angles, frequencies between 3 to 6 kHz would lead to Spectral Distortion levels up to 10 dB, which is a non-negligible change for spatial perception. The 7-8 kHz range is consistently affected although the sharp swing behaviour (see fig 4) may result from a simple notch misalignment. By looking at fig. 1, the main HRTF frequency notch falls exactly into the range where high HRTFD is observed. A small notch misalignment or a less pronounced attenuation would result in a high HRTFD as that seen in the graphs. Since the 7 kHz peak was observed to be azimuth-independent (fig 3) it is possible that the HMDs' headband around the head might cause some sort of constructive diffraction which could reduce or shift the frequency notch (a 7.5 kHz wave would result in a 4.5 cm wavelength). The above results, when compared to those reported in [26]

for repeated measurements of the same dummy head within the same space over the span of 10 years, indicate that the HMD acoustic perturbations are not due to random acoustic fluctuations. The reported range of variation for that situation was of the magnitude of $\leq 3\text{dB}$ (and $< 0.35\text{dB}$ s for consecutive measurements in the same space over a week), much lower than the values reported here.

The main discussion that needs to follow up from this objective study is whether the acoustic modifications are perceptually important for spatial localization of real sound sources. Localization of high-frequency content, vertical localization and front/back discrimination are all perceptual tasks which rely on spectral cues present at the affected ranges [2]. Nevertheless, the findings have highlighted non-negligible changes with local peaks up to 10dB and 23dB for the SD and HRTFD respectively. Several of the most sensitive locations of incidence show sub-band spectral distortions in the 3-6 kHz range to be higher than the ILD JND values reported in [11, 12], indicating possible issues in high-frequency content localization and consequent lateralization issues. The notch location within the $\sim 7\text{-}8\text{ kHz}$ range has been previously found to affect vertical perception [27], thus its reduced intensity and location shift may degrade the localization accuracy of elevated sound sources. An argument of auditory plasticity could be made in favour of trained listeners but it may be not strong enough to dismiss the factor as it would imply a training stage and the assumption of linear cue transformations [28].

A more difficult question to answer is whether virtual sources in mixed reality applications should apply acoustic compensation to match the response of the real waveforms as they get modified by the MR headset. The challenge is to correctly acquire insights on the impact and importance of the factor within complex mixed scene environments and human sensitivity to the HRTF fine structure within multi-modal environments. Such investigations should be careful in isolating the contributions of the HMD from the other potential sources of degradation of spatial audio quality. It is expected that the reported perturbations are likely to be a less disruptive issue than sub-optimal HRTFs, issues with sound reproduction methods, and room divergence effects [3] which highly affect plausibility of virtual sources.

6 Conclusions

This work illustrates an objective investigation of the acoustic perturbations caused by wearable AR/VR HMD devices on HRTFs. The effects of two commercially-available MR HMDs were measured on a dummy head mannequin over a fine grid of elevation and azimuth positions. Spectral distortions and binaural cue changes occur due to the physical occlusion effects in the near-head acoustic field and may affect localization at specific locations, potentially leading to lateralization effects or vertical localization errors in mixed reality auditory settings.

Results computed on two metrics found in literature (*Spectral Distortion (SD)* and *HRTFD*) show that non-negligible relevant distortions are mostly present at the contralateral anterior (CA) quadrant ($\phi = 270 - 360$) and ipsilateral posterior (IP) ($\phi = 90^\circ - 180^\circ$), starting from approximately 3 kHz up to 6 kHz, followed by notch deviations between 7-8 kHz. Within these ranges, the SD was found to be as high as 10dB and the HRTFD occasionally reached local variations $> 20\text{dB}$. Source incidence from higher elevations seem to influence perturbations via attenuation in the CA quadrant, while lower elevations were responsible for boosts at the IP quadrant. Differences between the two HMDs are likely caused by specific geometry and materials. Neither the ITDs nor the ITD frequency range were significantly affected by HMD devices.

The quantitative observations are deemed to be non-negligible, enough to motivate further perceptual study. The high-frequency distortions are found to be beyond JND levels for spatial localization tasks, although it is hypothesized that, given optimal conditions, localization accuracy might be impacted only for content and directions that are sensitive to high-frequency energy, such as vertical localization. Such investigations are relevant for mixed reality scenarios where optimal reproduction and realism of virtual sources is sought.

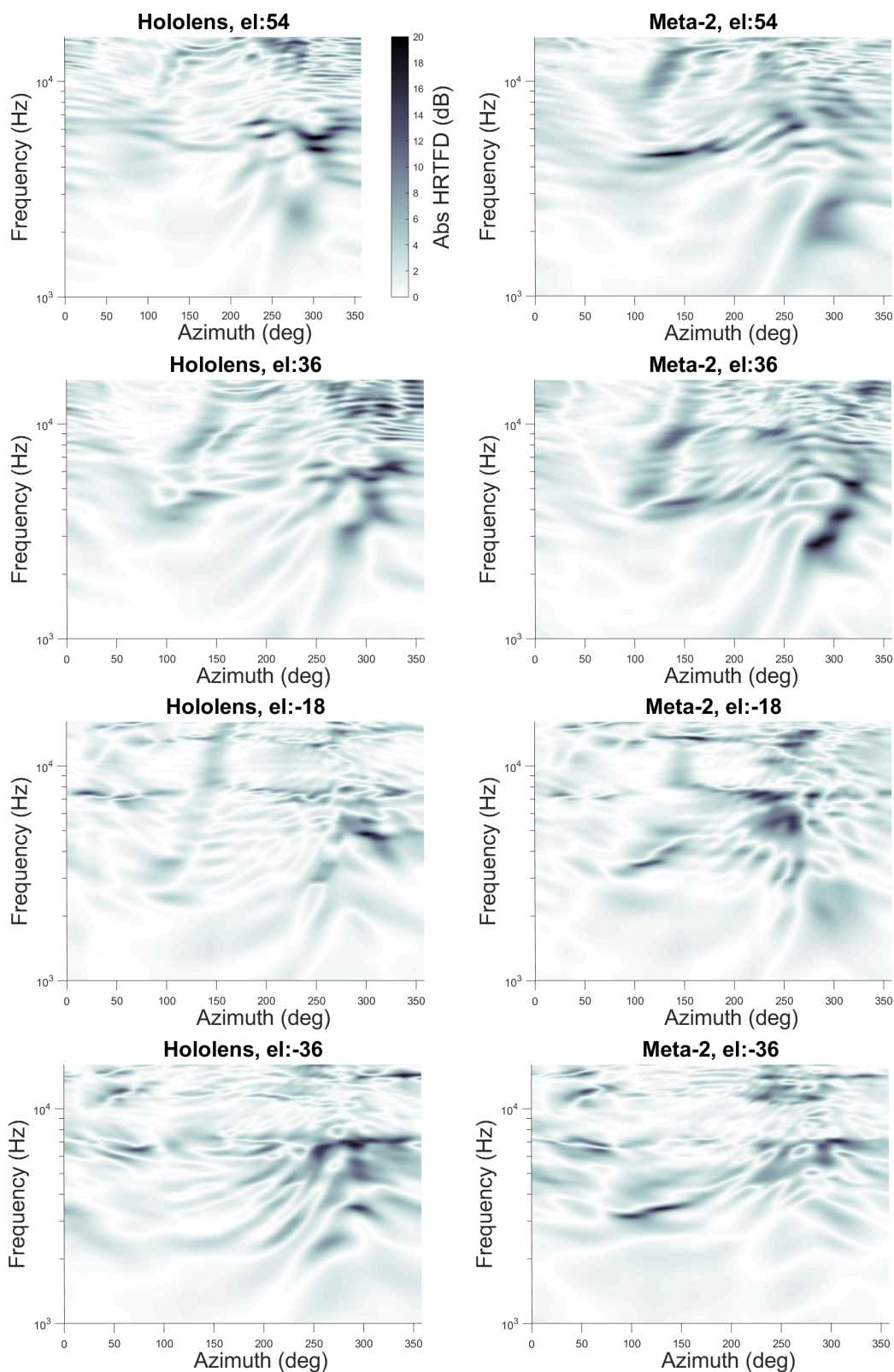
References

- [1] Ohta, Y. and Tamura, H., *Mixed reality: merging real and virtual worlds*, Springer Publishing Company, Incorporated, 2014.
- [2] Begault, D. R. and Trejo, L. J., *3-D sound for virtual reality and multimedia*, NASA, 2000.

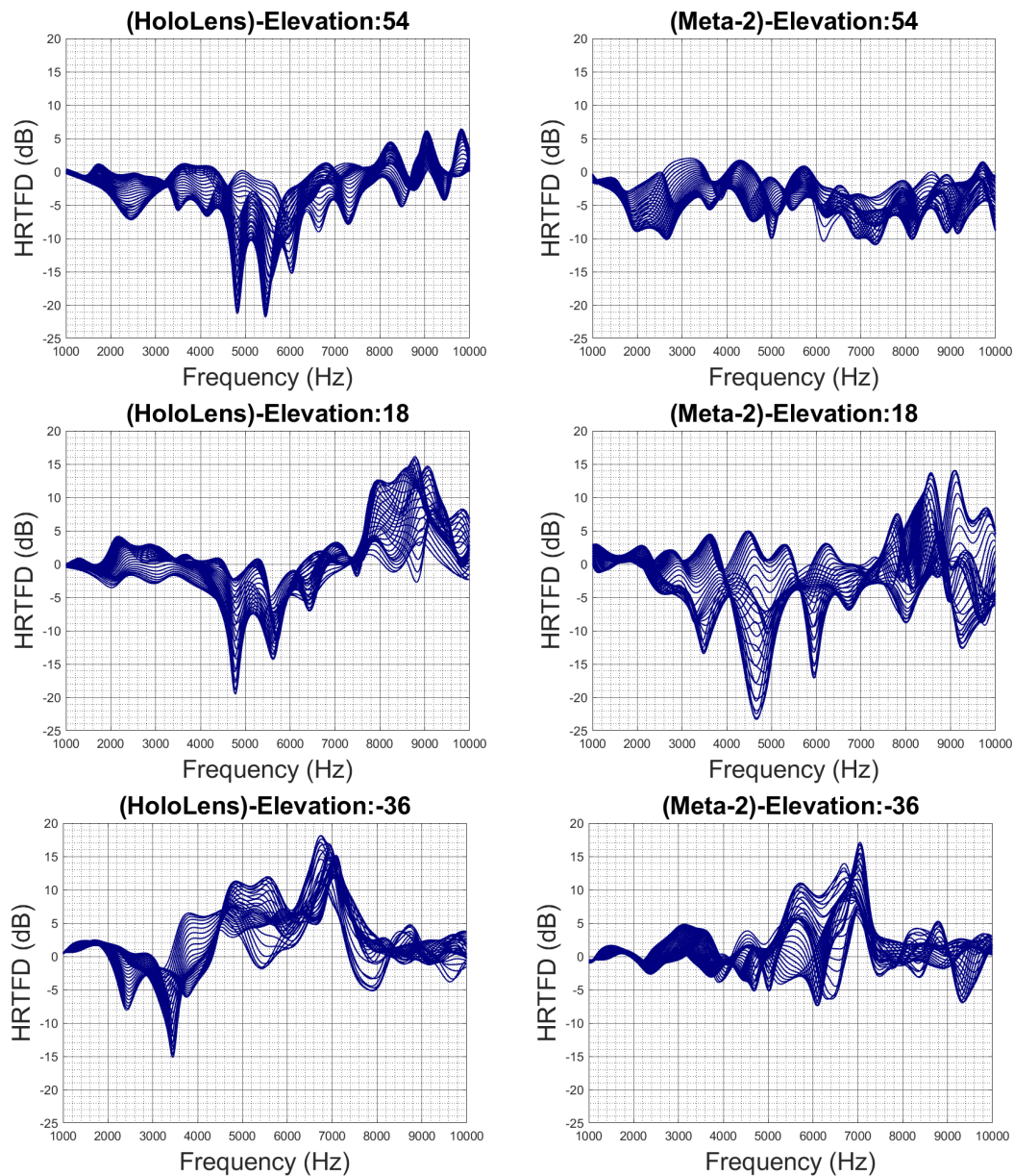
- [3] Werner, S., Klein, F., Mayenfels, T., and Brandenburg, K., "A summary on acoustic room divergence and its effect on externalization of auditory events," in *Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference on*, pp. 1–6, IEEE, 2016.
- [4] Freeman, J. and Lessiter, J., "Here, there and everywhere: the effects of multichannel audio on presence," Georgia Institute of Technology, 2001.
- [5] Pike, C., Melchior, F., and Tew, T., "Assessing the plausibility of non-individualised dynamic binaural synthesis in a small room," in *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*, Audio Engineering Society, 2014.
- [6] Brungart, D. S. and Rabinowitz, W. M., "Auditory localization of nearby sources. Head-related transfer functions," *The Journal of the Acoustical Society of America*, 106(3), pp. 1465–1479, 1999.
- [7] Klumpp, R. and Eady, H., "Some measurements of interaural time difference thresholds," *The Journal of the Acoustical Society of America*, 28(5), pp. 859–860, 1956.
- [8] Domnitz, R., "The interaural time jnd as a simultaneous function of interaural time and interaural amplitude," *The Journal of the Acoustical Society of America*, 53(6), pp. 1549–1552, 1973.
- [9] McFadden, D. and Pasanen, E. G., "Lateralization at high frequencies based on interaural time differences," *The Journal of the Acoustical Society of America*, 59(3), pp. 634–639, 1976.
- [10] Akeroyd, M. A., "The psychoacoustics of binaural hearing: La psicoacústica de la audición binaural," *International journal of audiology*, 45(sup1), pp. 25–33, 2006.
- [11] Domnitz, R. and Colburn, H., "Lateral position and interaural discrimination," *The Journal of the Acoustical Society of America*, 61(6), pp. 1586–1598, 1977.
- [12] Grantham, D. W. and Wightman, F. L., "Detectability of varying interaural temporal differences," *The Journal of the Acoustical Society of America*, 63(2), pp. 511–523, 1978.
- [13] Wersényi, G., "HRTFs in human localization: measurement, spectral evaluation and practical use in virtual audio environment," 2002.
- [14] Nishino, T., Inoue, N., Takeda, K., and Itakura, F., "Estimation of HRTFs on the horizontal plane using physical features," *Applied Acoustics*, 68(8), pp. 897–908, 2007.
- [15] Jo, H., Park, Y., and Park, Y.-s., "Analysis of individual differences in head-related transfer functions by spectral distortion," in *ICCAS-SICE, 2009*, pp. 1769–1772, IEEE, 2009.
- [16] Wersényi, G. and Illényi, A., "Differences in dummy-head HRTFs caused by the acoustical environment near the head," *Electronic Journal of Technical Acoustics*, 1, pp. 1–15, 2005.
- [17] Treeby, B. E., Pan, J., and Paurobally, R. M., "The effect of hair on auditory localization cues," *The Journal of the Acoustical Society of America*, 122(6), pp. 3586–3597, 2007.
- [18] Riederer, K. A., "Effects of eye-glasses, hair, headgear, and clothing on measured head-related transfer functions Part Ib," *The Journal of the Acoustical Society of America*, 114(4), pp. 2388–2388, 2003.
- [19] Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., et al., "What's new in Psychtoolbox-3," *Perception*, 36(14), p. 1, 2007.
- [20] Boren, B. and Roginska, A., "Multichannel impulse response measurement in matlab," in *Audio Engineering Society Convention 131*, Audio Engineering Society, 2011.
- [21] Boren, B. and Genovese, A., "Acoustics of Virtually Coupled Performance Spaces," in *International Conference on Auditory Displays, ICAD*, 2018.
- [22] Tylka, J. G., Boren, B. B., and Choueiri, E. Y., "A Generalized Method for Fractional-Octave Smoothing of Transfer Functions that Preserves Log-Frequency Symmetry," *Journal of the Audio Engineering Society*, 65(3), pp. 239–245, 2017.
- [23] Andreopoulou, A. and Katz, B. F., "Identification of perceptually relevant methods of inter-aural time difference estimation," *The Journal of the*

- Acoustical Society of America*, 142(2), pp. 588–598, 2017.
- [24] Perrott, D. R. and Saberi, K., “Minimum audible angle thresholds for sources varying in both elevation and azimuth,” *The Journal of the Acoustical Society of America*, 87(4), pp. 1728–1731, 1990.
- [25] Roginska, A. and Geluso, P., *Immersive Sound: The Art and Science of Binaural and Multi-channel Audio*, Taylor & Francis, 2017.
- [26] Andreopoulou, A., Begault, D. R., and Katz, B. F., “Inter-Laboratory Round Robin HRTF Measurement Comparison.” *J. Sel. Topics Signal Processing*, 9(5), pp. 895–906, 2015.
- [27] Lopez Poveda, E. A., *The physical origin and physiological coding of pinna-based spectral cues*, Ph.D. thesis, © Enrique A. López Poveda, 1996.
- [28] Shinn-Cunningham, B. G., Streeter, T., and Gyss, J.-F., “Perceptual plasticity in spatial auditory displays,” *ACM Transactions on Applied Perception (TAP)*, 2(4), pp. 418–425, 2005.

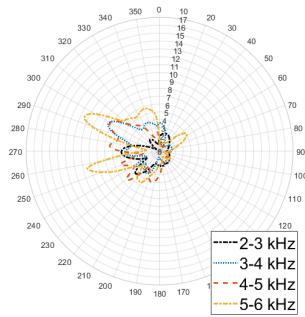
APPENDIX A. - Broadband HRTFDs at different elevations



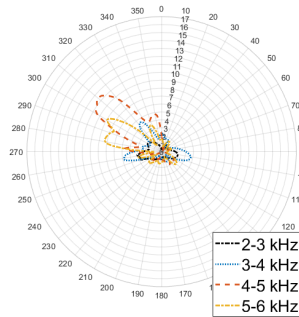
APPENDIX B. - Contralateral Anterior HRTFDs ranges at select elevations. Azimuth range: $\phi = 270^\circ - 330^\circ$, interval = 1.8° . Left column: HoloLens. Right column: Meta-2. Elevations go from higher to lower from the top row to the bottom.



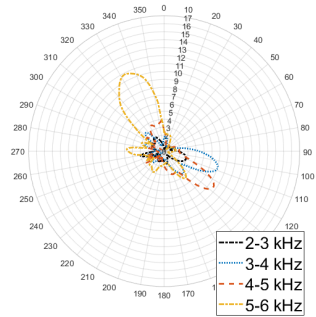
APPENDIX C. - Polar Spectral Distortion response for 1kHz-wide narrowbands across azimuths at each elevation level δ



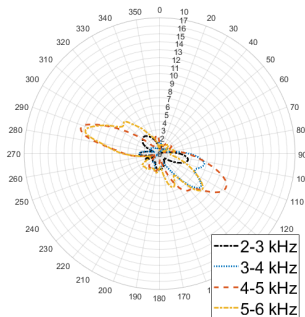
(a) Holens at $\delta = -36^\circ$



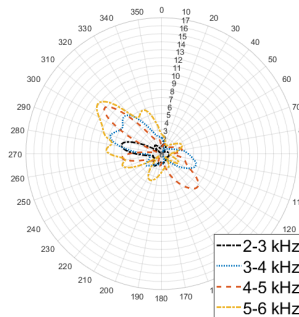
(b) Holens at $\delta = -18^\circ$



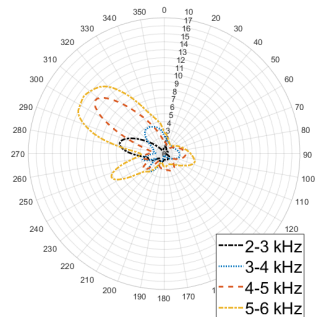
(c) Holens at $\delta = 0^\circ$



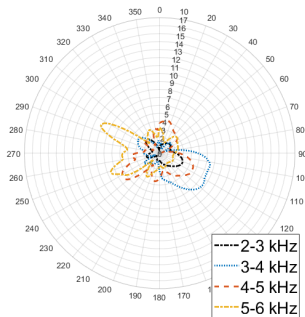
(d) Holens at $\delta = 18^\circ$



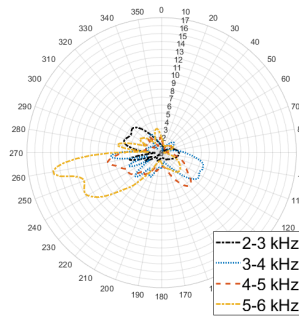
(e) Holens at $\delta = 36^\circ$



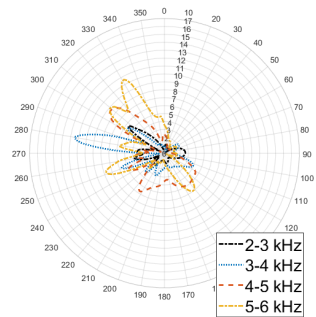
(f) Holens at $\delta = 54^\circ$



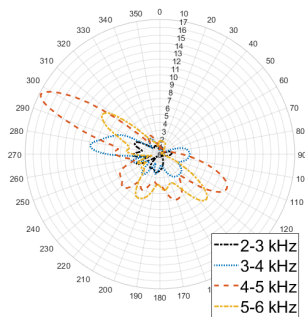
(g) Meta-2 at $\delta = -36^\circ$



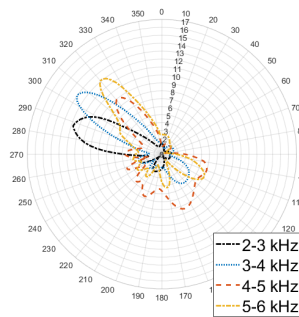
(h) Meta-2 at $\delta = -18^\circ$



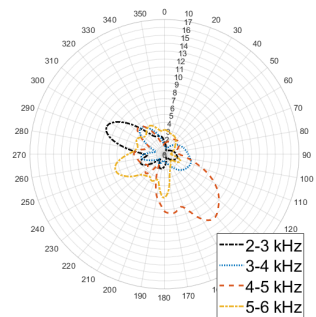
(i) Meta-2 at $\delta = 0^\circ$



(j) Meta-2 at $\delta = 18^\circ$



(k) Meta-2 at $\delta = 36^\circ$



(l) Meta-2 at $\delta = 54^\circ$

APPENDIX D. - Spectral Distortion (SD) boxplots across contralateral azimuths (180 to 360°) for Hololens (HL) and Metavision (META) for selected elevations and octaves)

