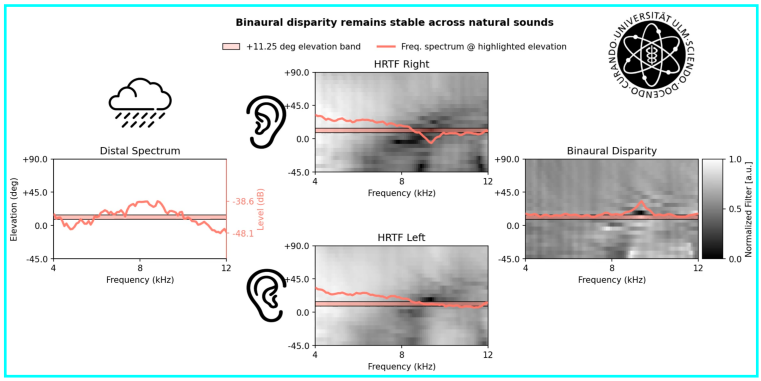
1 Supplementary



Video V1: Hearing Elevations https://10.6084/m9.figshare.33005588
This video must be watched with headphones.
Most people will be able to identify the sounds as externalized from their head, with roughly the correct elevation, as seen in the video. However, there is no guarantee that this demonstration (and thus also the following modifications) will work perfectly, as everyone's head and pinna shapes differ naturally. Achieving this reliably for everyone would require individualized recordings.
1 Left Ear / Right Ear, Intact: The sounds are in-ear recordings of natural, broadband noise-like signals presented predominantly in the median plane, with position changes largely restricted to elevation, such that ILD and ITD remain close to zero, as indicated by the bottom bar graphs.
2 Left Ear / Right Ear, Intact, with ILD and ITD Further Removed: To demonstrate that it is not ILD or ITD that is responsible for elevation hearing and externalization, we further removed ILDs and ITDs artificially, using a running window that set the interaural cross-correlation lag to zero and adjusted interaural level so that both ears received equal volume. This nullification method successfully removed persistent ILDs and ITDs, though some minor transients remained. Individuals who perceived the sounds in Condition 1 as externalized with the correct elevation perceived the elevation as unaltered by this manipulation, demonstrating the relevance of HRTFs – rather than ILD or ITD – for externalization and elevation hearing.
3 Monaural, Left Ear: Here, the right-ear channel was muted so that only the natural left-ear recording remained. Although this signal still contained the direction-dependent spectral filtering produced by the left pinna, the previously clear elevation percept, together with externalization, was largely lost. Sounds tended to be perceived near the left ear, inside the head, and their apparent position no longer reliably followed the elevation shown in the video. Thus, in this demonstration, spectral information from one ear alone was too spatially ambiguous to produce a stable elevation percept.
4 Left Ear / Left Ear, Mirrored: To test whether the loss of elevation in the monaural condition was caused merely by presenting sound to one side, the natural right-ear recording was replaced with a resampled copy of the left-ear recording (ensuring independent sampling noise in each channel). This guarantees that ITDs and ILDs are zero while preserving an intact HRTF at both ears, but without any binaural disparity (binaural disparity ratio = 1 throughout). Most people hear the sounds in this video as coming from close to the center of the head, without a clear sense of elevation or externalization. This demonstrates that binaural disparity – not monaural HRTF cues alone, nor their simple left/right summation – is responsible for elevation hearing and externalization.



Video V2: Binaural Disparity Remains Stable Across Natural Sounds https://10.6084/m9.figshare.33011876

This animation illustrates how changes in the sound source spectrum (left panel) strongly alter the monaural proximal spectra at both ears (upper and lower middle panels), even though the sound source remains fixed at the same location (red area, 11.25°). No clear structure that could be used for elevation hearing is visible in the proximal monaural signal across sounds, demonstrating the monaural ambiguity. In contrast, the binaural spectral disparity (right panel) remains largely invariant across sounds, demonstrating that it is largely independent of the variability introduced by the sound source spectrum. This invariance provides a stable cue for elevation localization that is unavailable from either ear alone. Turn on audio to hear the natural sounds associated with each frame.

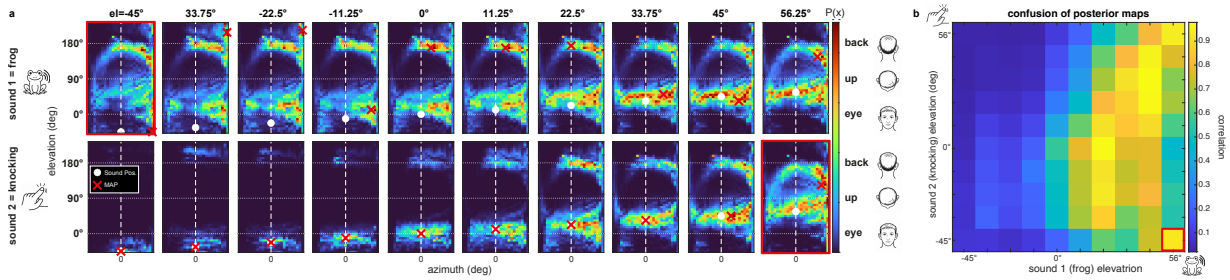


Figure S1. Supplementary Figure S1: Sound-dependent posterior localization maps a Posterior probability maps ($P(x)$) over candidate source locations ($x = (\text{azimuth}, \text{elevation})$) are shown for two example sounds (sound 1: frog vocalization, upper row, and sound 2: knocking sound, lower row) and 10 elevations (columns from el = -45° to 56.25°; az = 0°) using the monaural pathway of the spectral-template model. Warmer colors indicate higher posterior probability, i.e. higher similarity of the proximal monaural signal with the HRTF template map. White circle mark the true sound-source position x, and red crosses mark the maximum-a-posteriori estimate. Dashed guide lines indicate the reference azimuth of az = 0° and coarse elevation regions, labeled at right as eye: eye level, up: upward, and back: rearward locations. The maps illustrate that localization estimates are strongly shaped by the spectral structure of the stimulus. For the frog sound, posterior probability often forms broad elevation and front-back ambiguity ridges, causing MAP estimates to shift away from the true position. For the knocking sound, posterior mass is more concentrated in several panels but still shows elevation-dependent ambiguity. Red outlines highlight representative cases also used for Fig. 1 where the posterior structure leads to prominent correct or erroneous localization estimates. Overall, the figure demonstrates that the same HRTF template map can produce different inferred elevations depending on the source spectrum, emphasizing the stimulus dependence of spectral-template localization. **b** Confusion matrix of posterior maps over different elevations for azimuth=0°, for the two example sounds. X-axis sound 1: frog, y-axis sound 2: knocking. The bottom right, red rectangle indicates the high correlation of the two sounds at the example locations from Fig. 3.

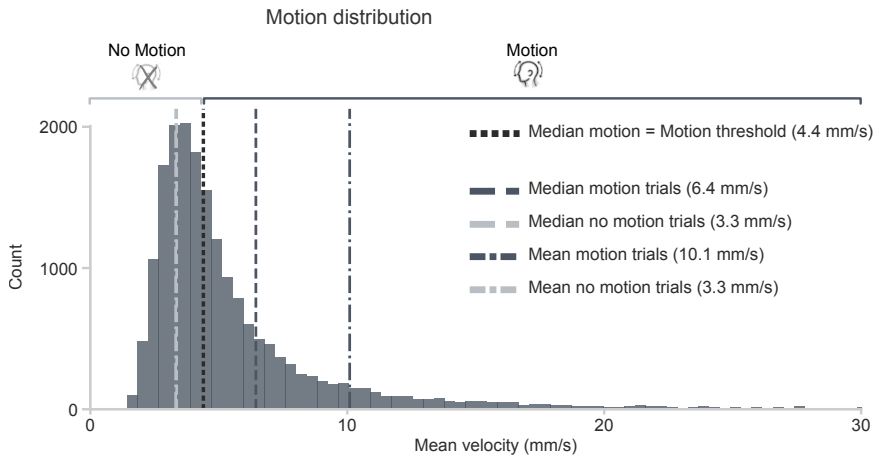


Figure S2. Supplementary Figure S2: Head motion analysis: split into motion and no motion trials Distribution of mean head velocity for all trials across all participants of the main localization experiment. Velocity was computed from OptiTrack position data as the mean absolute frame-to-frame displacement, converted to mm/s. Dotted vertical line marks the median split (4.4 mm/s), i.e. the threshold above which trials were classified a *motion* trials and below as *no-motion* trials. Dashed vertical lines mark the mean of each category (no motion: 3.3 mm/s; motion: 10.1 mm/s). Dash-dotted line vertical lines mark the median of each category (no motion: 3.3 mm/s; motion: 6.4.1 mm/s)

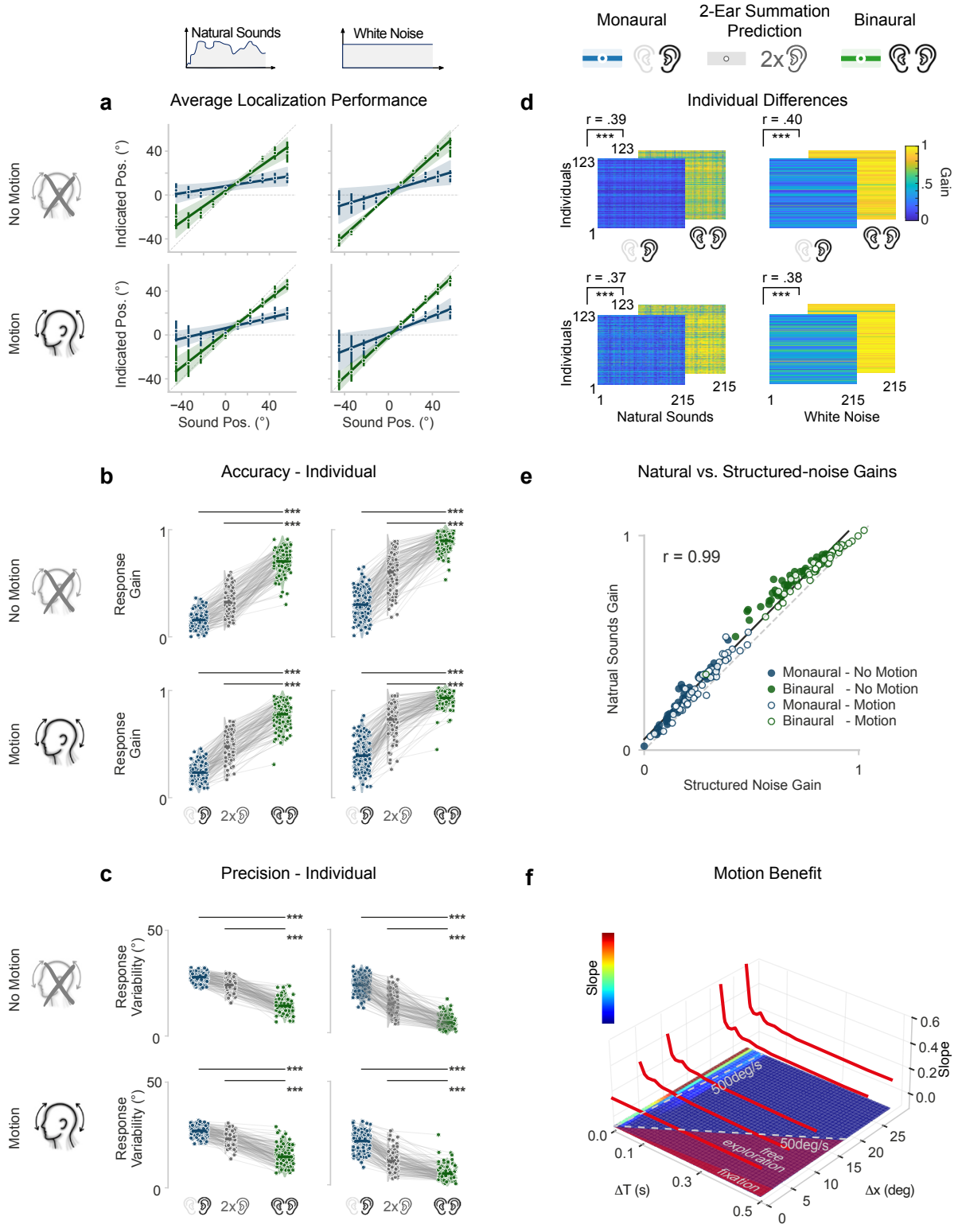

Natural Sounds
White Noise
Monaural
2-Ear Summation Prediction
Binaural
2x
a
Average Localization Performance
No Motion
Motion
Indicated Pos. (°)
Sound Pos. (°)
d
Individual Differences
r = .39
r = .40
r = .37
r = .38
Individuals
Gain
Natural Sounds
White Noise
b
Accuracy - Individual
Response Gain
2x
e
Natural vs. Structured-noise Gains
r = 0.99
Natrual Sounds Gain
Structured Noise Gain
Monaural - No Motion
Binaural - No Motion
Monaural - Motion
Binaural - Motion
c
Precision - Individual
Response Variability (°)
f
Motion Benefit
Slope
500deg/s
50deg/s
free exploration
fixation
ΔT (s)
Δx (deg)

Figure S3 ***(preceding page).*** **Supplementary Figure S3: Model analysis for Natural Sounds a–c Localization performance with natural sounds vs. flat-spectrum sounds.** Panels compare localization performance of the spectral template model using an extended set of 123 simulated listeners (combining CIPIC [3] and SoundSphere 2.0 HRTFs [49] databases), and for 215 natural and man-made sounds from the MIT [35] and ESC-50 databases [38] against flat spectrum sounds (white noise). Natural sounds are shown in the left sub-columns and white noise (flat spectrum) sounds in the right sub-columns; upper rows of each panel show the no motion conditions, with motion in the lower rows. **a** Localization performance: This panel shows average localization performance for the true source elevation against the model-estimated elevation. Blue regressions indicate monaural localization, binaural localization performance is shown in green. Dots indicate mean responses for each elevation, solid lines show population regression fits, shaded ribbons indicate uncertainty around the fit. Dashed gray lines mark veridical and chance performance. **b** Accuracy: Distribution of individual localization slopes for all 8 conditions (monaural: blue, binaural: green) and the 2-ear-summation prediction (gray). **c** Precision: Distribution of individual localization variability for all 8 conditions and the 2-ear-summation prediction (same color code). The variability is quantified as posterior standard deviation in degrees. Performance with natural sounds is almost identical to that with structured-noise sounds from Fig. 3 (c.f. panel d). **d Individual differences in localization gain for natural sounds across conditions**. Panel d shows individual differences in model gain across virtual listeners and sounds. Heatmaps show localization gain for natural sounds and white noise, separately for stationary and motion conditions; warmer colours indicate higher gain. Correlation coefficients quantify the consistency of individual model differences across listening conditions. **e Comparison of natural sounds vs. structured noise.** Model localization gain using natural sounds against model localization gain for structured-noise stimuli for any simulated listener (dots: HRTF sets from the databases). Data are shown for all 4 listening-condition (motion: open symbols / no motion: filled symbols; monaural: blue / binaural: green). Solid black line gives the least-squares regression fit. Localization slopes were nearly identical as indicated by the correlation of r=0.99, (p< .001), (n=180). This emphasizes and verifies the similarities in using structured-noise sounds and natural sounds. **f Added motion benefit as a function of time window width (ΔT) and change in head orientation (Δx)** Benefit for natural sounds from adding motion over the no motion asymptote for monaural localization. The benefit is shown as a function of time window width $\Delta T = t_2 - t_1(s)$ and amount of head nodding $\Delta x(^\circ)$. We chose to show the monaural localization condition, as for this condition the room for improvement is largest. Qualitatively the effects are identical for the binaural condition. The heatmap marginals (floor) as well as the red lines illustrate the improvement of the regression gain over the asymptote (no motion) from adding motion (angular head speed ($v = \Delta x/\Delta t$) with Δt=(0–0.5 s) and Δx=(0–25°)). Gray dashed lines on the graph floor indicate typical head speeds for fixation ($\lesssim 3^\circ/s$), typically occurring head velocities for free exploration ($\lesssim 50^\circ/s$) and a threshold head velocity potentially resulting in a marked motion effect size ($\gtrsim 500^\circ/s$), illustrating that such high velocities are implausible. Significance labels denote paired comparisons (*, ($p < .05$); **, ($p < .01$); ***, ($p < .001$)).

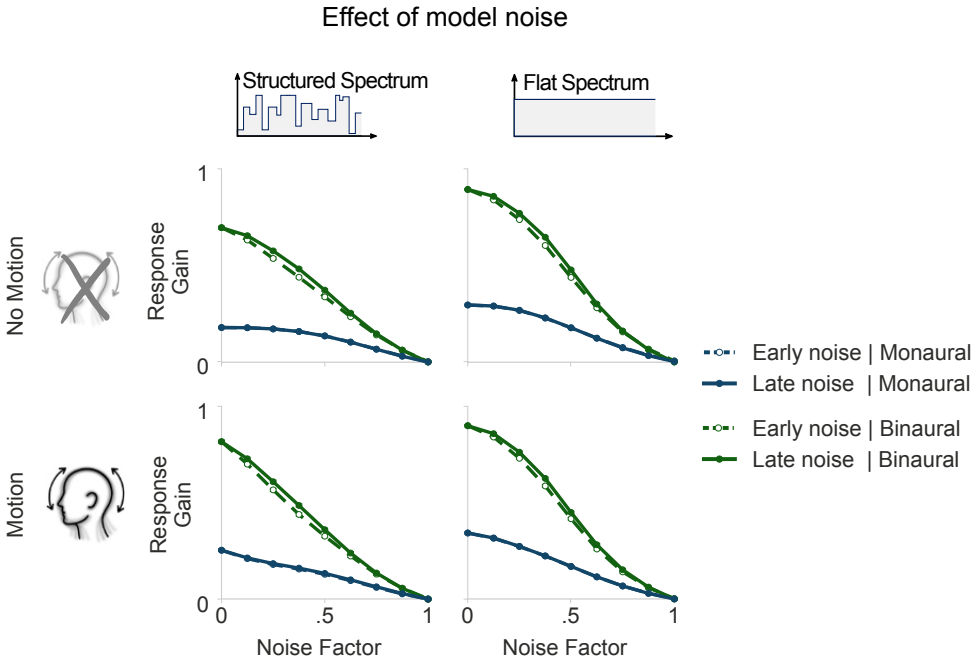


Figure S4. Supplementary Figure S4: Effect of noise on the spectral-template model. Localization gain as a function of the noise factor for the monaural (blue) and binaural (green) model conditions. The noise factor specifies the relative contribution of stimulus and noise: a noise factor of 0 corresponds to 100% stimulus and 0% noise, whereas a noise factor of 1 corresponds to 0% stimulus and 100% noise; intermediate values represent the corresponding linear mixture of the two components. Open symbols indicate early noise, and filled symbols indicate late noise. Performance for structured-noise sounds is shown in the left columns and performance for white noise (flat spectrum) in the right columns. The upper row shows the no-motion condition, and the lower row shows the motion condition. In the early-noise condition, independent normally distributed noise was added to each frequency channel before computation of the binaural log-spectral difference. In the late-noise condition, noise was added after the binaural computation, during the template-matching stage. For the monaural model, early and late noise are equivalent. The difference between early and late noise was small, while the binaural advantage over monaural localization remained robust across noise levels, indicating that the binaural benefit was largely insensitive to the stage at which noise was introduced. Structured-noise sounds behaved essentially identical to the natural sounds shown here when noise was added.

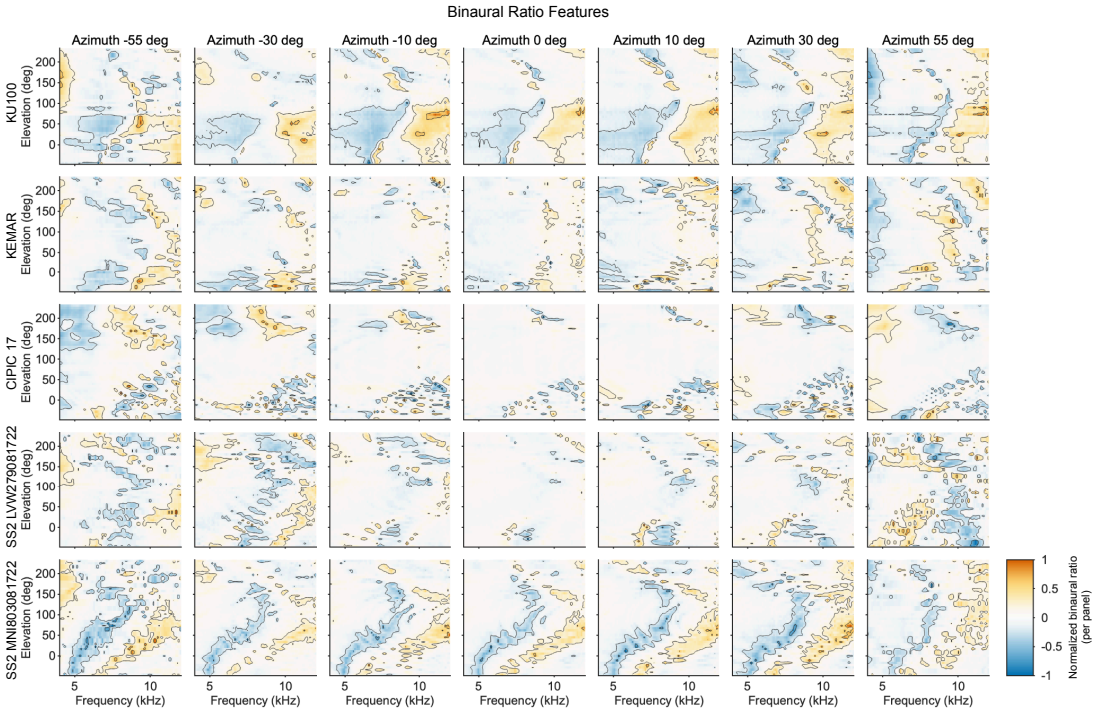


Figure S5. Supplementary Figure S5: Binaural HRTF ratio features. Rows show the binaural spectral ratio maps for frequencies from 4 to 12 kHz against elevation derived from dummy heads recordings (row 1: KU100 - Neumann; row 2: KEMAR mannequin) and from example human individuals of the CIPIC (row 3: Nr. 017 - individual used for Fig 1&2) and the SoundSphere2 database (row 4: Nr. LVW279081722 ; row 5: Nr. MNI803081722). Columns show azimuths of $[-55, -30, -10, 0, +10, +30, +55]$. Plotted is the difference between the processed left- and right-ear monaural HRTF feature maps (L-R in the log-magnitude spectral domain). Positive values (warm colors) indicate relatively stronger left-ear spectral energy/features, and vice versa for negative values (cool colors), neutral colors (white) indicate weak interaural spectral differences. Contour lines mark iso-value structure within each map. Each panel was normalized by its own maximum absolute value before plotting. Features are weakest at 0° azimuth and get stronger with azimuth, left and right. Some individuals have stronger features (row 5) than others (row 3 & 4). Even mannequin heads show binaural ratio features at 0 azimuth, perfect symmetry would predict no features (plane of confusion).