Anticipated ITD statistics are built into human sound localization

Rodrigo Pavão[1,2*], Elyse S. Sussman[1], Brian J. Fischer[3], José L. Peña[1]

[1] Dominick P. Purpura Department of Neuroscience - Albert Einstein College of Medicine, NY, USA
[2] Centro de Matemática, Computação e Cognição - Universidade Federal do ABC, SP, Brazil
[3] Department of Mathematics - Seattle University, WA, USA
* rodrigo.pavao@einstein.yu.edu

The variability of natural scenes places perceptual processes in the realm of statistical inference. Perceptual tasks may be optimized if the invariant statistical structure of sensory cues is built into the neural processing. We investigated this question in human sound localization. Localizing sounds in the horizontal plane relies on interaural time differences (ITD). We estimated ITD statistics from human head-related transfer functions (HRTFs) and properties of cochlear filters. ITD varied with azimuth following a sigmoid relationship, whose slope was steepest at the center in most frequencies. In addition, ITD was more variable over time for sounds located in the periphery compared to the center, in a frequency-dependent manner. We tested the hypothesis that these statistics (ITD slope and ITD variability) are anticipated by the human brain, influencing spatial discriminability and novelty detection. Previously reported thresholds for discriminating ITD changes could be better predicted by a model relying on both ITD slope and variability than on ITD slope alone. Furthermore, mismatch negativity (MMN) brain signals, a pre-attentive index of deviance detection recorded in subjects undergoing a spatial oddball paradigm, were weighted by ITD slope and ITD variability with significantly stronger influence of statistics expected for repeated (standard) over sporadic (deviant) locations. These results show that spatial discriminability thresholds and novelty detection are consistent with a representation of anticipated ITD statistics in the brain, supporting the hypothesis that high-order statistics are built into human perceptual processes biasing behavior.

Keywords: Sound localization; ITD; HRTF; Natural statistics; Variability; Detectability; JND; MMN.

Introduction

Humans and other species localize sound sources in the horizontal plane using the sub-millisecond interaural time difference (ITD) between signals arriving at the two ears (Middlebrooks & Green 1991). ITD is detected by auditory brainstem neurons within narrow frequency bands (Goldberg & Brown 1969; Yin & Chan 1990; Carr & Konishi 1990; McAlpine et al. 2001).

Classical psychophysical studies demonstrated that humans detect sound location better in the front than in the periphery (Mills 1958; Yost 1974; Makous & Middlebrooks 1990). The enhanced capability in frontal locations could be efficient for hunting and foraging, as analogously proposed for vision (Collins et al. 1922; Cartmill 1974; Changizi & Shimojo 2008). Physiological evidence indicates that coding of binaural spatial cues could support finer discriminability in the front (van Bergeijk 1962; Feddersen et al. 1957; McAlpine et al. 2001; Grothe et al. 2010). Better sound discrimination and localization in frontal locations can also be predicted from the geometry of the head and the placing of the ears, causing larger ITD changes as a function of azimuth in the front (Woodworth 1938; Feddersen et al. 1957; Gelfand 2016).

For azimuth detection based on ITD, sound diffraction by structures surrounding the ear modulates the ITD vs azimuth relationship (Aaronson & Hartmann 2014). In addition, because coincidence detector neurons compute ITD in narrow frequency bands (Yin & Chan 1990; Carr & Konishi 1990), the interaction of nearby frequencies may be an important source of ITD variability. Sensory variability is known to be fundamental in change detection. It has been shown that stimulus discrimination depends not only on the mean difference between standard and deviant stimuli but also on the variability of the sensory evidence (Green & Swets 1966).

This study tested the hypothesis that there is a built-in representation of environmental statistics in the brain. Using human HRTF databases, we estimated ITD rate of change across azimuth and ITD variability over time. We tested whether these statistics predicted spatial discrimination thresholds and deviant detection. Using sound stimuli that removed ITD variability (tones delivered by earphones inserted in the ear canal), allowed us to test whether anticipated, rather than actual, statistics predicted deviant detection. Our results indicate that expected variability determines sound localization thresholds and novelty detection, supporting the notion that high-order sensory statistics are built into human brain processing.

Results

ITD statistics (specifically, the derivative of the mean ITD over azimuth and the standard deviation of ITD) were estimated from human HRTFs. We first tested whether the ITD statistics predict human spatial discrimination thresholds (Mills 1958). Next, we used EEG and mismatch negativity signals (MMN) to address the question of whether natural ITD statistics are built into spatial deviant detection.

ITD statistics estimated from human HRTFs

To test whether ITD statistics are built into brain processing, we estimated the ITD rate of change across locations and ITD variability of sounds reaching the ears of humans subjects. Convolving acoustic signals by impulse responses reported in publicly available human HRTF databases (Listen HRTF database, http://recherche.ircam.fr/equipes/salles/listen; 51 subjects), results in modulation of ongoing phase and gain that depend on sound direction. We estimated the effect of these modulations on the mean and standard deviation of instantaneous ITD in sounds

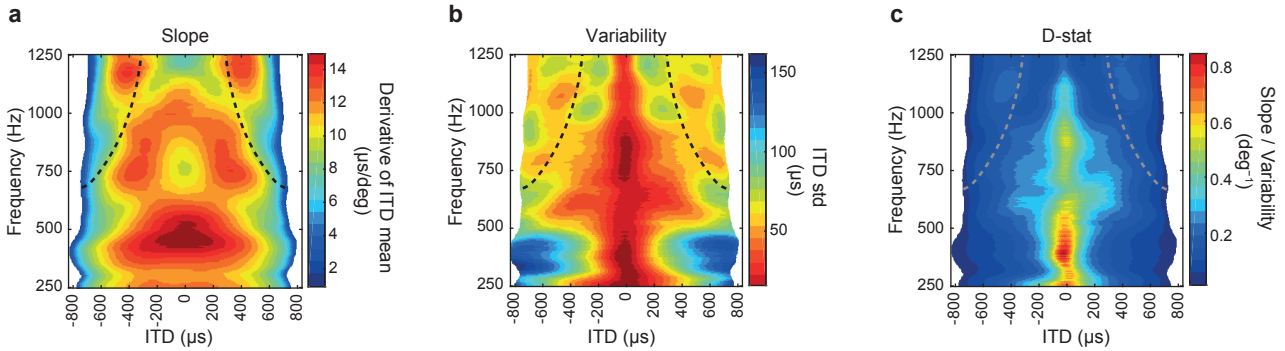

Figure 1. Natural ITD statistics from human HRTFs. a) ITD slope across azimuth and frequency, plotted as a function of the ITD corresponding to each azimuth. As previously reported (Mills et al. 1972; Gelfand 2016), the slope is higher mostly in the front, although the pattern is complex. **b)** ITD variability (standard deviation over time) for different ITDs and frequencies. Variability is lower in the front and higher in the periphery across frequency, with a more pronounced pattern in low frequencies. **c)** ITD discriminability (D-stat). Dashed lines indicate the pi-limit, beyond which ITD information becomes ambiguous within narrow frequency bands. The color map in b is opposite to the color maps in a and c in order to highlight intervals of highest discriminability with red.

filtered using parameters analogous to human cochlear filters (Glasberg & Moore 1990). ITD statistics (rate of change and variability) were derived from these measures.

The rate of change of ITD (ITD slope) across location is shown as a function of ITD and frequency in Figure 1a (see Suppl. Figure S1 for a details about the method). We found that the ITD slope was steeper in the midline for most frequencies. However, ITD slope varied across frequency, including cases where the steepest slope occurred at locations distant from the midline. Although this pattern was not highlighted in previous studies (Kuhn 1977; Benichoux et al. 2016), it was already observable (Suppl. Figure S2), indicating that it is consistent across approaches.

Next, we estimated the ITD variability induced by phase modulation over the stimulus duration (ITD variability; Figure 1b and Suppl. Figure S1), calculated from the standard deviation of IPD along broadband signals convolved with impulse responses. We confirmed that this variability results from the interaction between neighboring frequencies, within the critical band of single cochlear filters (Suppl. Figure S3a-c). ITD standard deviation was consistent across signals with different frequency spectra, demonstrating that the ITD statistics we estimated are a property of the HRIR, and not of the signal we convolved it with (Suppl. Figure S3d). The estimated variability over time was equivalent (r = 0.99) to the variability over trials (see Methods), indicating that ITD variability exhibits ergodicity.

Finally, azimuth discriminability based on ITD was estimated by computing a D-stat index, dividing ITD slope by ITD variability at each azimuth location (Figure 1c). From these data, frequencies and azimuth ranges of interest were selected for subsequent analyses and experimental design.

We also tested consistency of ITD statistics across environments, comparing ITD statistics from HRTFs recorded in anechoic and reverberant rooms (database available in http://medi.uni-oldenburg.de/hrir; Kayser et al. 2009). We found that echoes significantly disrupted ITD statistics at time latencies larger than 5 ms from the beginning of the signal (Suppl. Figure S3e). However, at latencies less than 5 ms, where the precedence effect is expected to segregate leading signals from their lagging echoes (Brown et al. 2015), ITD statistics estimated in anechoic conditions and reverberant environments were similar, indicating that ITD statistics are preserved in transient portions of sounds in reverberant environments.

Prediction of spatial discrimination thresholds from ITD statistics

ITD just-noticeable difference (JND) as a function of azimuth and ITD was obtained from classic published data on human sound localization (Mills 1958) (Figure 2a,b).

We tested whether ITD slope, ITD variability or their combination (ITD D-stat) predicted the JND. Figure 2c shows the JND as a function of ITD slope, ITD variability and ITD D-stat across the full frequency and ITD ranges (Figure 2c, top), for the ITD range within the pi-limit (removing the phase-ambiguous range; Figure 2c, middle) and for the low frequency range (Figure 2c, bottom). JND was significantly correlated with ITD slope (Figure 2c, left column) and ITD variability (Figure 2c, middle column) but the highest correlation coefficients were found for ITD D-stat (Figure 2c, right column). Correlation coefficients for each plot are shown in Figure 2d. ITD D-stat was also a better predictor of JND than ITD slope measurements from other studies (Kuhn 1977; Benichoux et al. 2016) (Suppl. Figure S4a). These results suggest that ITD statistics are built into the brain processing underlying discrimination thresholds.

To take into account mechanisms supporting greater ITD sensitivity at the midline, we considered previously reported brainstem models predicting greater ITD sensitivity in the front. Specifically, Stern & Colburn (1978) proposed an exponential increase in the density of ITD sensitive cells towards the midline (Suppl. Figure S5a), consistent with the increased sensitivity in the front. Similarly, Harper & McAlpine (2004) showed that a brainstem ITD coding maximizing Fisher information (Suppl. Figure S5b) could also predict differential sensitivity across locations and frequencies. These models showed varied levels of predictive power

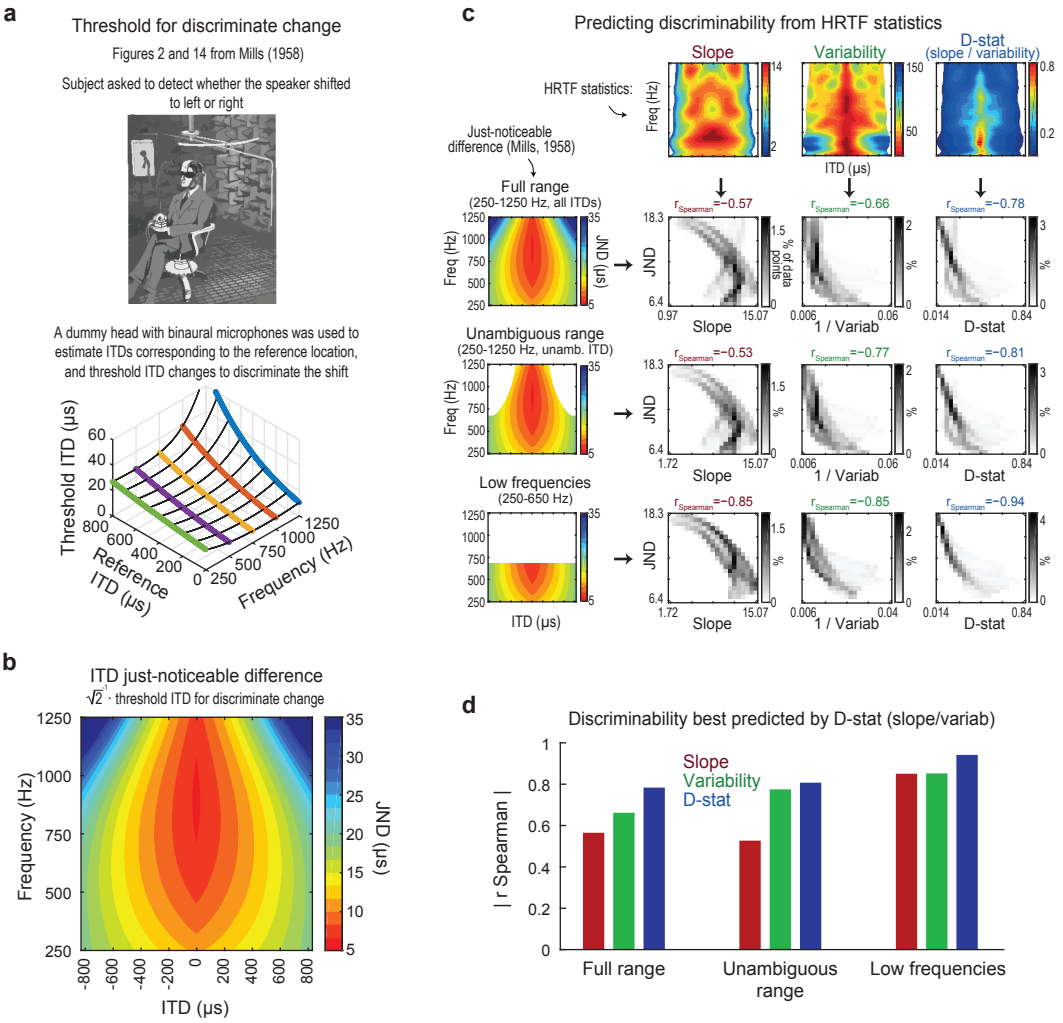

Figure 2. Human detection thresholds of ITD changes match expected ITD statistics. a) Mills (1958) estimated the minimum ITD change (threshold ITD, proportional to JND) that could be detected across frequency and location (top). Azimuth was converted to ITD using a dummy head with binaural microphones (bottom). Reproduced from Mills (1958); doi: 10.1121/1.1909553, with the permission of the Acoustical Society of America. **b)** JND across frequency and ITD. JND was calculated by dividing ITD change thresholds estimated by Mills (1958) by the square root of 2, to take into account reported underestimation of JND by minimum audible angle measurements (Hartmann & Rakerd 1989). JND is proportional to the inverse of discriminability. **c)** Grayscale plots show JND as a function of ITD slope (left), ITD variability (middle) and their ratio (i.e., slope/variability; right). Darker gray areas indicate high data point density. Each row shows three different frequency and ITD ranges - whole range (top), unambiguous ITD range (middle) and fully unambiguous ITD and low-frequency range (bottom). Ranked correlation coefficients are presented on the top of each plot. **d)** Bar plot of correlation coefficients. While discriminability is significantly correlated with ITD variability and ITD slope, the ITD D-stat combining both slope and variability yields the highest correlation coefficients.

of Mills (1958) JND data (Suppl. Figure S5c), consistent with the idea that JND can be linked to properties of the network supporting sound localization. Taken together, the two best predictors of JND were the exponential model (Stern and Colburn 1978) and the D-stat. This can be explained by the analogous monotonic relationship between population density and ITD discriminability in Stern and Colburn (1978) and between D-stat and ITD within each frequency in our model; in fact, modeling ITD discriminability as a linear function of ITD carries similar predictive power (Suppl. Figure S5c). These results hint at the still unanswered question of whether tuning properties of the neural population representing ITD can capture ITD statistics.

The representation of ITD statistics in tuning properties has been proposed for barn owls, where ITD-selective midbrain neurons are tuned to frequencies that display lower interaural phase difference (IPD) variability in concurrent-sound environments (Cazettes et al. 2014; Fischer & Peña 2017). Due to the difference in frequency range over which owls and humans detect ITD (Konishi 2003), and the strong frequency dependence of gain modulation exerted by the head (Gelfand 2016), it is expected that ITD variability induced by concurrent sounds will vary with frequency. In particular, because low frequencies are more weakly filtered, concurrent sounds within the frequency ranges of human and owl ITD detection are expected to differ. To test this prediction, we estimated the variability of ITD with concurrent sound for the human data (Suppl. Figure S6a). As predicted, this statistic was strongly dependent on the amplitude of concurrent sounds, indicating that ITD variability induced by concurrent sounds is not a reliable statistic to be anticipated by the human brain. Consistently, ITD

variability estimated by this method did not predict human JND (Suppl. Figure S6b). These results indicate that across all ITD statistics investigated (i. ITD slope across azimuth; ii. ITD variation over time; and iii. ITD variability by concurrent sounds), the slope and variation are the most reliable across contexts. Furthermore, the observation that a ratio between ITD slope and ITD variation over time (D-stat) was the best descriptor of human JND, supports the notion that both statistics are relevant for human sound localization. To further test the hypothesis that ITD variability determines discriminability, a novelty detection paradigm was used.

ITD statistics built into deviance detection

To further test if statistics are built into the human brain processing, we evaluated novelty detection EEG responses for sounds coming from standard (repeated) and sporadic (deviant) locations. The amplitude of the MMN indexes discriminability between standard and deviant sounds without the listener making an overt response. Thus, we addressed the question of whether the amplitude of the MMN was modulated by anticipated ITD statistics.

We recorded EEG mismatch negativity responses (Näätänen et al. 1978) for spatial deviants in tones presented through earphones. Using tonal stimuli delivered through insert earphones removes the ITD variability induced by frequency interference within the critical band of cochlear filters and gain modulation by the filtering effect of outer ear and surrounding structures. Thus, differences in deviant detection could be assigned to expected rather than current stimulus statistics. MMNs allowed for a quantitative indexing of the effect of reliability on both standard and deviant stimuli. In addition, the pre-attentive nature of MMN signals did not require training subjects to perform behavioral tasks, which might introduce confounding effects.

MMNs are observable when a sequence of standard (repeated) stimuli is unexpectedly interrupted by a low probability deviant stimulus (Näätänen et al. 1978; Pakarinen et al. 2007; Sussman 2007; Sussman et al. 2013) (Figure 3a). The MMN signal is obtained by subtracting event-related potentials (ERPs) evoked by deviant stimuli from ERPs evoked by the standard. The larger the separation between standard and deviant stimuli, the more negative the MMN amplitudes, which has led researchers to use MMN amplitude as an index of deviant discriminability (Deouell et al. 2006; Sams et al. 1985; Pakarinen et al. 2007; Tiitinen et al. 1994).

We measured MMN responses in 33 subjects. A set of ITD and frequency conditions was selected in order to sample critical ranges drawn from the HRTF analysis (Figure 3b). In particular, we selected 400, 550, 600 and 650 Hz because ITD slope and variability changed as a function of azimuth in a manner that could maximize the difference in predictive power across models; frequencies above 650 Hz were excluded to avoid phase ambiguity confounds. MMN signals were measured across subjects and conditions. The data showed the characteristic frontal topography of MMN responses (Giard et al. 1990) (Figure 3c). The peak amplitude of MMN was used to quantify the effect of stimulus manipulations on novelty detection.

The MMN peak amplitude averaged across subjects showed negative correlation to the ITD difference (absolute difference between the ITDs of standard and deviant stimuli) (Figure 3d), consistent with previous reports (Pakarinen et al. 2007). However, we observed stronger correlation between ITD difference and MMN amplitude when the ITD of the standard was frontal (Figure 3d, right) and no correlation with the standard in the most peripheral locations. This was due to an absence of MMN when the standard was at the most peripheral locations, suggesting that the relationship depends not only on the ITD difference but on the location of the standard. Consistently, standard stimuli located in the front led to larger MMNs than deviants in the front (Figure 3e). This comparison was performed between pairs of standard and deviants of exactly the same frequency and ITDs, where only the probability of either stimulus was switched. These results indicate that MMNs were mainly elicited when the standard stimuli were located in the front and that deviant detection signals diminished for standards in the periphery.

Finally, we selected the statistical model that best described the MMN amplitude. Three groups of models were tested, which normalized the ITD difference between standard and deviant by the ITD slope (Figure 3f, red), ITD variability (Figure 3f, green) or ITD D-stat (Figure 3f, blue). For each model, the effect of changing the relative weights of standard and deviant was examined. We found that normalization by ITD variability was required for a good predictive power. Interestingly, models that considered only the ITD statistics of the standard performed better (model "m_1" and "m_2"; Figure 3g, left and middle). Specifically, the best prediction of MMN peak was obtained by normalizing the ITD difference between standard and deviant by the ITD D-stat, and assigning 74% weight to the standard and 26% to the deviant (model "m_3"; Figure 3g, right). The two sets of subjects, represented in red and blue, showed consistent results, indicating that these findings were robust across conditions.

Additionally, we examined the predictive power of MMN amplitude by models taking into account only ITD slope measurements described in previous studies (Kuhn 1977; Benichoux et al. 2016). These models showed performance similar to the ITD slope data used in our study but poorer prediction compared to models that accounted for ITD variability (Suppl. Figure S4b).

We also tested the prediction of MMN peak amplitude modeling ITD sensitivity by (1) an exponential function of cell density across ITD reported by Stern & Colburn (1978) (Suppl. Figure S5a), and (2) the firing rate changes across ITD weighted by the distribution of cells reported in Harper & McAlpine (2004) (Suppl. Figure S5b). Normalizing the ITD difference between standard and deviant by either of these ITD sensitivity metrics lowered the predictive power relative to the models relying

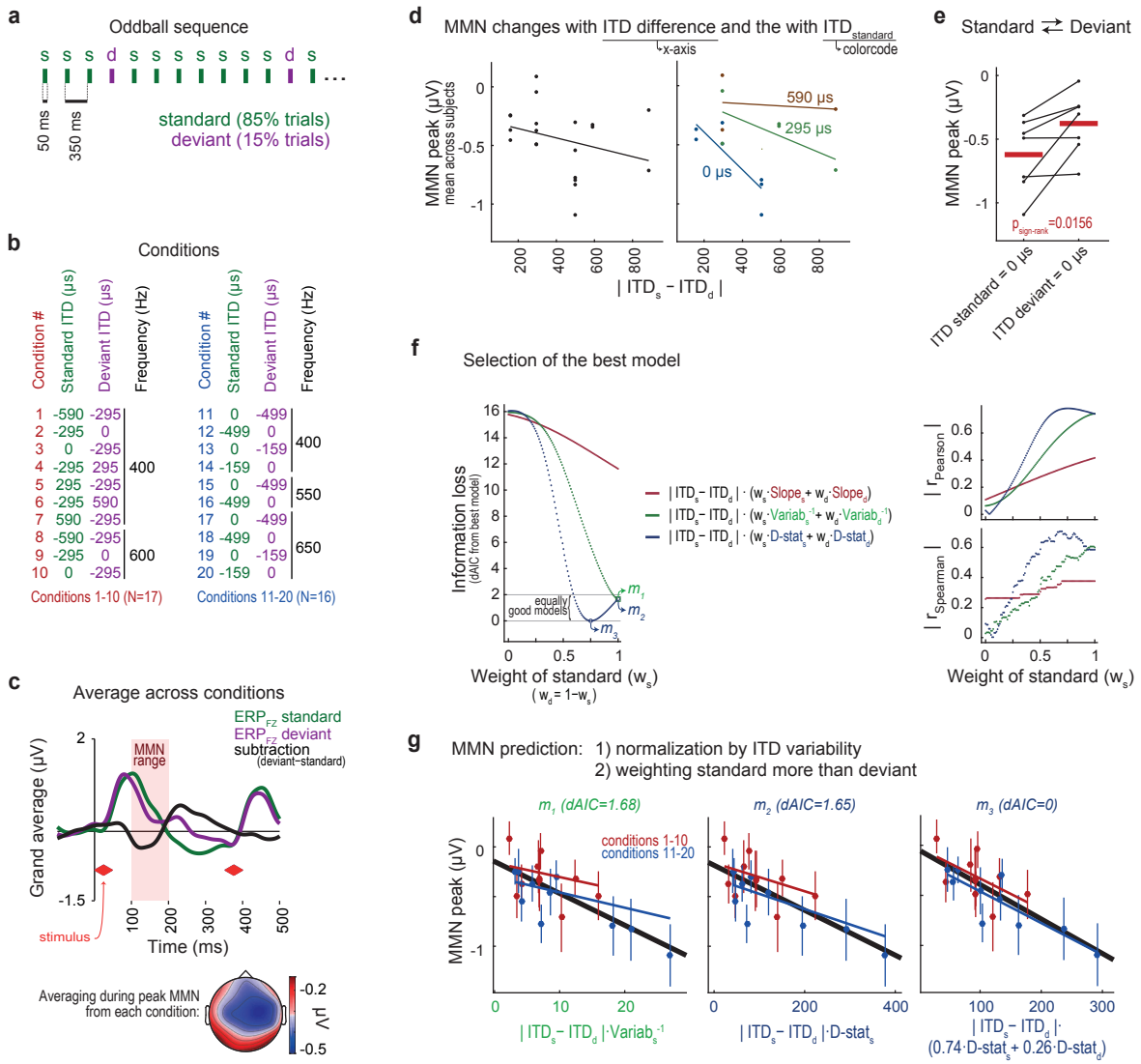

Figure 3. Discriminability of spatial deviants indexed by MMN responses is weighted by expected ITD variability at the standard location. **a)** Passive oddball sequence protocol, in which subjects listened to frequent "standard" stimuli embedded with rare "deviants." **b)** In each condition, two tones of same frequency and distinct ITDs were presented. Two sets of 10 conditions were tested in two different groups of subjects. **c)** Standard (green) and deviant (purple) event related potentials (ERP), recorded at the midline frontal electrode (FZ), were averaged across conditions and subjects. Note the negative deflection of the subtraction (deviant-standard) ERP within the 100-200 ms range (black). MMN signals are characteristically more negative within frontal locations. **d)** While MMN amplitude was weakly related to ITD difference when all data were included (left), grouping the data by the ITD of the standard (0 µs, 295 µs and 590 µs), exposed different degrees of correlation (right). In particular, the stronger relationship occurred when the standard was frontal (0 µs, blue). **e)** Paired comparison interchanging the ITD of the standard and deviant showed significantly more negative MMNs when the standard was in the front (Wilcoxon signed rank test). **f)** AIC statistical test for selecting the model with best predictive power of MMN amplitude. The models normalized ITD difference by anticipated ITD slope and/or ITD variability, weighting standard (w_s) and deviant (w_d) differently (left). The good models show also high correlation coefficients with MMN peak (right). **g)** Prediction of MMN amplitude by three models where ITD difference was normalized by ITD variability of the standard alone (m_1, left), ITD D-stat of standard alone (m_2, middle) and D-stat of standard and deviant with differential weights (m_3, right). The best of these models (dAIC=0) was m_3. dAIC indices are shown on top of each plot.

on ITD variability (Suppl. Fig S5d). This observation shows that models taking into account only ITD slope have lower predictive power of MMN amplitude but does not exclude cell density and tuning properties as mechanisms to represent ITD statistics (Fischer & Peña 2011; Cazettes et al. 2016). Finally, we examined the predictive power of MMN amplitude by ITD variability over concurrent sounds (as described in Cazettes et al. 2014), which was poor compared to D-stat (Suppl. Fig S6c).

Discussion

Human HRTFs were used to estimate the ITD statistics across subjects, namely the rate of change of ITD over azimuth (ITD slope) and standard deviation of ITD over time and trials (ITD variability), across frequency and ITD. ITD slope and ITD variability, and their ratio (ITD D-stat) were used to show that natural statistics are built into the brain processing underlying human sound localization.

Different explanations for the greater discriminability in the frontal region have been proposed, such as variation in density of brainstem ITD-sensitive neurons (Colburn 1977; Stern & Colburn 1978) or greater variation in firing rate as a function of azimuth in the front compared to the periphery (McAlpine et al. 2001; Harper & McAlpine 2004). These are valid explanations of biological mechanisms underlying discriminability. In this context, other mechanisms based on the amount of spatial information carried by auditory stimuli have also been invoked, such as the rate of variation of ITD as a function of azimuth (Mills et al. 2018; Gelfand 2016). Our study does not negate the role of these mechanisms in ITD discriminability but specifically tests whether the inherent variability of interaural cues from different positions is a significant causal parameter determining discriminability as well. In fact, our results further support the notion that network properties supporting discriminability are evolutionarily adapted to capture sensory statistics.

Previous reports have investigated the selectivity of midbrain neurons to the variability of spatial cues in the owl's auditory system (Cazettes et al. 2014; Fischer & Peña 2017). These studies provided evidence of how sensory reliability could be represented (Fischer & Peña 2011; Rich et al. 2015; Cazettes et al. 2016) and integrated into an adaptive behavioral command (Cazettes et al. 2017). Although the ITD statistics relevant to owls and humans differ based on the frequency range over which ITD is detected, and the coding of ITD cannot be assumed identical across species (Schnupp and Carr 2009), the present study shows that specific ITD statistics, ITD slope and ITD variability, determine human sound localization discriminability and novelty detection of acoustic spatial deviants. The present work does not make any assumption on how ITD and ITD statistics are encoded in the human brain. The code underlying the causal link between inherent variability of interaural cues and discriminability in the human brain is an open question for future studies.

Natural ITD statistics predict spatial discriminability thresholds

Discrimination thresholds of spatial changes across frequency and location have been predicted by the rate of change of ITD in azimuth (Mills et al. 1972; Gelfand 2016). This interpretation does not consider ITD variability. Here we show that ITD variability improves the prediction of these thresholds. The correlation between discriminability and each of the ITD statistics shows that the ratio between ITD slope and ITD variability (D-stat) is the best predictor of discriminability thresholds. Yet some specific patterns of the JND data are not captured by ours and others' models, such as the unexpected higher ITD slope or ITD D-stat in the midline for 450 Hz compared to 700 Hz. Besides these limitations, the D-stat model is a better predictor of ITD discriminability than models considering ITD slope alone, supporting the idea that the brain anticipates, and therefore represents, the spatial discriminability as a function of frequency, location and reliability of the sensory cue.

Because sound stimulation in Mills (1958) consisted of tones, the across-frequency interaction was not a source of ITD variability. On the other hand, Mills (1958) used sounds presented through external speakers (free-field stimulation), which allows for an effect of head and ear-canal filtering on the acoustic signals reaching the eardrum. Thus, the Mills' (1958) data do not allow us to discern other information-limiting effects (e.g. changes in gain or actual ITD variability) from anticipated ITD variability on discriminability thresholds. However, the better predictive power of Mills' (1958) data by models including anticipated ITD variability suggests that these statistics play a role.

Novelty detection weighted by anticipated ITD statistics

The novelty detection studies conducted with tonal sounds delivered by insert earphones were designed to more directly approach the question of whether anticipated ITD variability is built into for spatial perception. Consistent with observations on the discriminability thresholds reported in Mills (1958), the ratio between ITD slope and ITD variability (D-stat) was a better predictor of deviant detection responses.

Although ITD-change detection thresholds, as probed in Mills (1958), and our novelty detection paradigm required discriminating a change in ITD, the latter also involves identification of a repetitive (standard) pattern. The internal representation of standard patterns may be optimized by integrating predicted variability. Parras et al. (2017) developed an approach designed to isolate the relative contribution of prediction errors and repetition suppression in novelty detection. In the present study, anticipated ITD variability might modulate novelty detection at both levels. Future work may address this question. Consistently, the models that best described the MMN signals weighted the ITD difference between standard and deviant primarily by the ITD statistics of the standard location. Detecting ITD-changes in a reference location, as in Mills (1958), is influenced by attention and training. The novelty detection protocol allowed for control of these factors because the MMN provides a brain response to detected deviations irrespective of attention or training. In addition, the novelty detection paradigm showed that the prediction is based mostly on the ITD statistics of the standard stimulus, consistent with the interpretation that anticipated ITD variability is critical for pattern detection. Furthermore, we speculate that the different weights for standard and deviant is the result of the mechanism underlying the standard formation, the building of a standard, which requires repetitive stimulation. We hypothesize that, when the standard is in a location anticipated to produce reliable cues, a 'stronger' standard is built.

In sum, we found evidence that anticipated ITD statistics correlate with auditory spatial perception. These statistics of the natural auditory scene could be learned through individual experience. However, the consistency of these statistics across subjects indicates that this

information may be invariant enough to be genetically encoded and conserved. Thus, the ability to predict these statistics by a built-in representation is a potentially adaptive evolutionary mechanism for approaching optimal performance. This study provides evidence supporting the hypothesis that high-order statistics important for sound localization are built into sensory processing mechanisms of the human brain. Our results also identify stimulus parameters relevant to spatial discrimination and novelty detection that could lead to the development of more efficient hearing-aid devices.

Methods

HRTF measurement

The dataset used in this study consisted of head-related impulse responses collected at the Institute for Research and Coordination in Acoustics/Music (IRCAM) from 2002 to 2003, available to the public at the LISTEN HRTF website http://recherche.ircam.fr/equipes/salles/listen. The procedure was performed inside an anechoic room with walls that absorbed sound waves above 75 Hz. The pulse signals were played by TANNOY 600 speakers facing the subjects, placed 1.95 m of distance of the head center. Subjects were seated on an adjustable rotating chair with a position sensor placed on the top of the head, allowing recording only when the position was correct. The impulse sounds were collected with a pair of small Knowles FG3329 microphones, calibrated using a B&K 4149 microphone. These small microphones were mounted in a silicon structure which occluded the ear canals of the subjects, avoiding resonance and placing the microphone sensor at the entrance of the ear canal. The signal captured by the microphones was driven to a custom amplifier with 40 dB gain, recorded through a RME sound card interface with Max/MSP real time library which deconvolved the microphone signal.

HRTF analysis

Data from 51 subjects were included in the HRTF analysis (Suppl. Figure S1). We selected impulse responses corresponding to 0-degree in elevation and azimuth from -90 to 90 degrees. The data in the database was sampled in azimuth with 15-degree steps. We convolved the impulse responses with a white noise signal of 1-second duration. This procedure transfers the temporal and level cues to the convolved signal. The convolved signal was filtered in multiple frequency bands, with center frequencies ranging from 250 to 1250 Hz with 5 Hz steps (range that ITD is an effective cue (Middlebrooks & Green 1991) and that corresponds to range that have the thresholds estimated (Mills 1958)), and bandwidth equivalent to the estimated bandwidth for human cochlear filters (Glasberg & Moore 1990). We used the gammatone filter bank from Malcolm Slaney's Auditory Toolbox, available in https://engineering.purdue.edu/~malcolm/interval/1998-010). The filtered signal was represented in the Hilbert space for extracting its instantaneous phase, using the Signal Processing Toolbox (Mathworks). For each frequency and azimuth, we calculated the circular mean and standard deviation of the interaural phase difference (IPD, in radians). Next, we unwrapped the IPD data, to avoid the ITD slope be corrupted by artificial negative values caused by phase ambiguity. We also selected IPD values within the -pi to pi range around the midline (thus after the conversion, ITD values were close to zero). Finally, we converted to ITD (in μs) by the following equation: $ITD=10^6 IPD(2\pi Freq)^{-1}$.

For each azimuth and frequency we computed the median ITD across subjects. We calculated the rate of change of ITD across azimuth (ITD slope) as the ITD difference between two consecutive positions divided by their azimuth difference (15 degree). The ITD slope values were assigned to the intermediate position between each pair of consecutive azimuths. We interpolated these measures using the cubic spline method, then converted azimuth to ITD (using the original relationship between azimuth vs ITD), obtaining an estimate of ITD slope across frequency and ITD (Figure 1a).

We computed the median across subjects of the ITD variability for each azimuth and frequency sampled (Suppl. Figure S1). These data were then interpolated across azimuth and represented as a function of ITD (Figure 1b). To confirm that this metric of ITD variability is ergodic, i.e., variability across time and trials is equivalent, we also computed the ITD variability over trials, by selecting ITDs at the midpoint of 1-second signals, across 200 different broadband sounds.

Although both ITD slope and ITD variability are likely affected by diffraction and interference between neighboring frequencies, their correlation was not strong (rPearson=-0.41; rSpearman=-0.42; data not shown), indicating that they are partially independent. The combined effect of slope and variability was assessed by their ratio, a metric consistent with the Signal Detection Theory's d-prime (difference in means divided by square root of mean variance). This ratio was referred to as ITD D-stat (Figure 1c). Note that d-prime and D-stat are different quantities.

Estimation of spatial discriminability thresholds

Human spatial discriminability thresholds were estimated on the classic Mills (1958) study. Data collection for this study was performed inside an anechoic room with a movable speaker which delivered tones in different frequencies. The 3 participants were blindfolded and had their heads fixed by a clamp mounted on the chair on which they were sitting (Figure 2a, top). In each trial, a 1-second duration 'reference azimuth' tone was played first, and 1 second after, the same signal played again after the speaker was moved slightly to the left or to the right. Subjects reported the left or right change using an interface box. Psychometric functions were obtained plotting the proportion of judgements to the left and to the right against the angle between reference and test locations. The azimuth changes leading to 75% responses to the right and to the left were estimated by linear interpolation. The threshold angle for discriminating a change was estimated by dividing the distance between these values by 2.

To convert azimuth to ITD, Mills (1958) used binaural microphones placed inside a dummy-head ears. The ITD corresponding to the reference azimuth and the IPD corresponding to the threshold azimuth change were measured visually in an oscilloscope. The threshold IPD vs. reference ITD were plotted in a logarithmic Y-axis vs linear X-axis scale and linear curves were fit to the data.

For the current study, we extracted the scales, intercept and slopes from the linear fit of threshold IPD against reference ITD across frequency of Mills (1958). The first conversion we did was threshold IPD to threshold ITD, followed by interpolation using the cubic spline method. This conversion was also performed by Mills (1958) and allowed us to confirm that our value extraction was correct (Figure 2a, bottom). Hartmann & Rakerd (1989) demonstrated that the minimum audible angle (MAA) method underestimates the just-noticeable difference (JND) for human listeners; we thus divided the original data from Mills by the constant factor predicted by the decision theory, square root of 2. Using the original Mill's data or corrected JND led to the same correlation coefficients, since this analysis is unaffected by scaling constants. Finally, we represented the JND of ITD (an inverse measure of discriminability measure) in the same format in which we represented the HRTF metrics (Figure 2b).

Prediction of spatial discriminability thresholds by ITD statistics

We calculated the ranked (Spearman) correlation coefficients between JND and each of the HRTF measures. We tested this relationship in three overlapping but different ranges of ITD and frequency: (1) the full frequency (250 to 1250 Hz) and ITD range described by Mills (1958); (2) excluding the range where ITD becomes ambiguous at high frequencies, i.e., outside pi-limit range; (3) including only the frequencies that do not show phase ambiguity within the human ITD physiological range (250 to 650 Hz) (Figure 2c, left panels).

Collection and analysis of the Mismatch Negativity (MMN) component

Healthy adult subjects were included in the sample (N=33, 16 females; 17 males; mean age 29.5 ± 4.8; all right-handed). After the procedure was described to the subjects, they provided written informed consent. The protocol was approved by the Internal Review Board of the Albert Einstein College of Medicine, where the study was conducted. All subjects had no reported history of neurological disorders and passed a bilateral hearing screening (20 dB HL at 500, 1000, 2000, and 4000 Hz).

The subjects were seated inside a sound-proof booth (IAC Acoustics, Bronx, NY) in front of a TV screen where a subtitled muted movie was presented. Sound was delivered by a Neuroscan StimAudio system through insert earphones (3M Eartone 3A) calibrated using a B&K 4947 microphone with an artificial ear. Sound signals consisted of 50-ms tones (25-ms rise-fall time) of different frequencies and ITDs synthesized with Matlab software.

Subjects listened to oddball sequences consisting of repetitive ("standard") tones embedded with sporadic ("deviant", 15% of the trials) tones, with an inter-stimulus interval of 350 ms (Figure 3a). Each subject was presented with 10 conditions, which differed in frequency and ITD (Figure 3b). Subjects 1 to 17 performed the conditions 1 to 10, while subjects 18 to 33 performed the conditions 11 to 20. Each condition was presented in 3 blocks of 474 trials; the block order was randomized for each subject. Each of the conditions was presented 3 times during the experimental session; the order the blocks was randomized for each subject. Sessions lasted approximately 2.5 hours, including placement of electrode caps and breaks during the EEG recording.

The EEG signals were recorded using a Neuroscan SynAmps system connected to a 32-channel electrode cap following the 10-20 System, including electrodes on the nose (reference) and left and right mastoids (LM and RM, used for offline analysis). A bipolar configuration was used between an external electrode below the left eye and the FP1-electrode position for measuring vertical electro-oculogram (EOG). The signal was recorded at 500 Hz sampling rate, and subsequently band-pass filtered from 0.05 to 100 Hz.

To measure the MMN, EEG signals from each subject were processed as follows: (1) 20-Hz low-pass filtering; (2) pooling signals obtained during sound stimulation; (3) removal of eye-blink and saccade artifacts by performing Independent Component Analysis and reconstructing the signal without components correlated to the EOG; (4) selection of 600-millisecond epochs around sound presentation (-100 to 500 ms from sound onset); (5) removal of linear portions of EEG signals and exclusion of trials containing voltage deflections larger than 75 mV; (6) re-reference of the EEG recordings to the mastoid recordings by subtracting the average mastoid signal from the other channels; (7) ERP measurement, averaging separately signals evoked by standard and deviant trials (the first 19 trials were used for subject training and excluded from the analysis); (8) subtraction of standard from the deviant ERPs to compute the MMN; (9) identification of time bin containing the MMN peak in average signals from the FZ channel across subjects, for each condition; (10) measurement of MMN peak for each subject and condition within this time bin.

Grand average of ERPs recorded on FZ electrodes were computed for standard and deviant trials across all subjects and conditions (Figure 3c, top). For estimating the MMN topography, the signal from each electrode in the time bin of the peak MMN was averaged across subjects and conditions (Figure 3c, bottom).

Prediction of MMN by ITD statistics

An initial analysis of the linear (Pearson) and ranked (Spearman) correlation coefficients was performed for the relationship between the mean MMN peak and absolute difference between ITDs of the standard and deviant (Figure 3d).

Additional analyses were then performed

normalizing the ITD difference by the ITD variability, ITD slope and their combination, from both standard and deviant locations. The selection of which normalization was the best predictor of MMN peaks was performed with linear mixed-effect models (Matlab function fitlme, equation 'MMNpeak ~ 1 + model + (1|subject)'). The Akaike Information criterion (AIC) was used to compare the performance of each model, where the lowest AIC corresponds to the best model. Since AIC is a relative quantity (i.e., the actual value brings no information alone), we normalized the AIC values by subtracting the minimum AIC observed for that behavioral measure (dAIC), which corresponds to the information loss compared to the best available model. Models with dAIC between 0 and 2 were considered equally good in their prediction capacity.

All data processing was performed in Matlab (Mathworks) using built-in or custom made routines. The datasets generated during and analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgements

This work was supported by NIH BRAIN Initiative (Grant NS104911) and by the NIDCD (Grant R01 DC004263). We thank Fanny Cazettes, Gervasio Batista, Michael Beckert and Ruben Coen Cagli for comments on the manuscript, and Renee Symonds, Kelin Brace and Huizhen (Joann) Tang for support with troubleshooting during early stages of data collection.

Author contributions

RP, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; BF, ES, JP, Conception and design, Drafting or revising the article.

Ethics

This study was performed in accordance with NIH Human Subjects Policies and Guidance, and it was approved by the Internal Review Board of the Albert Einstein College of Medicine (#1999-023).

References

Aaronson, N. L. & Hartmann, W. M. Testing, correcting, and extending the Woodworth model for interaural time difference. *J. Acoust. Soc. Am.* **135,** 817–823 (2014).

Benichoux, V., Rébillat, M. & Brette, R. On the variation of interaural time differences with frequency. *J. Acoust. Soc. Am.* **139,** 1810 (2016).

Brown, A. D., Stecker, C., Tollin, D. J. The precedence effect in sound localization. *J. Acoust. Soc. America.* **16,** 1-28 (2015).

Carr, C. E. & Konishi, M. A circuit for detection of interaural time differences in the brain stem of the barn owl. *J. Neurosci.* **10,** 3227–3246 (1990).

Cartmill, M. Pads and Claws in Arboreal Locomotion. in *Primate Locomotion* 45–83 (1974).

Cazettes, F., Fischer, B. J., Beckert, M. V., & Peña, J. L. Emergence of an Adaptive Command for Orienting Behavior in Premotor Brainstem Neurons of Barn Owls. *J. Neurosci.* **38,** 7270–79 (2018).

Cazettes, F., Fischer, B. J. & Peña, J. L. Cue Reliability Represented in the Shape of Tuning Curves in the Owl's Sound Localization System. *J. Neurosci.* **36,** 2101–2110 (2016).

Cazettes, F., Fischer, B. J. & Pena, J. L. Spatial cue reliability drives frequency tuning in the barn Owl's midbrain. *Elife* **3,** e04854 (2014).

Changizi, M. A. & Shimojo, S. 'X-ray vision' and the evolution of forward-facing eyes. *J. Theor. Biol.* **254,** 756–767 (2008).

Colburn, H. S. Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. *J. Acoust. Soc. Am.* **61,** 525–533 (1977).

Collins, E. T., & Opthalmological Society of the United Kingdom. *Arboreal life and the evolution of the human eye: a revised publication of the Bowman lecture delivered before the Ophthalmological Society of the United Kingdom in May, 1921 / by E. Treacher Collins*. (1922).

Deouell, L. Y., Parnes, A., Pickard, N. & Knight, R. T. Spatial location is accurately tracked by human auditory sensory memory: evidence from the mismatch negativity. *Eur. J. Neurosci.* **24,** 1488–1494 (2006).

Feddersen, W. E., Sandel, T. T., Teas, D. C. & Jeffress, L. A. Localization of High-Frequency Tones. *J. Acoust. Soc. Am.* **29,** 988–991 (1957).

Fischer, B. J. & Peña, J. L. Optimal nonlinear cue integration for sound localization. *J. Comput. Neurosci.* **42,** 37–52 (2017).

Fischer, B. J. & Peña, J. L. Owl's behavior and neural representation predicted by Bayesian inference. *Nat. Neurosci.* **14,** 1061–1066 (2011).

Gelfand, S. A. *Hearing: An Introduction to Psychological and Physiological Acoustics*. (CRC Press, 2016).

Giard, M. H., Perrin, F., Pernier, J. & Bouchet, P. Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. *Psychophysiology* **27,** 627–640 (1990).

Glasberg, B. R. & Moore, B. C. Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* **47,** 103–138 (1990).

Goldberg, J. M. & Brown, P. B. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *J. Neurophysiol.* **32,** 613–636 (1969).

Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics*. (1966).

Grothe, B., Pecka, M. & McAlpine, D. Mechanisms of Sound Localization in Mammals. *Physiol. Rev.* **90,** 983–1012 (2010).

Harper, N. S. & McAlpine, D. Optimal neural population coding of an auditory spatial cue. *Nature* **430,** 682–686 (2004).

Hartmann, W. M. & Rakerd, B. On the minimum audible angle—A decision theory approach. *J. Acoust. Soc. Am.* **85,** 2031–2041 (1989).

Konishi, M. Coding of auditory space. *Annu. Rev. Neurosci.* **26**, 31-55 (2003).

Kuhn, G. F. Model for the interaural time differences in the azimuthal plane. *J. Acoust. Soc. Am.* **62,** 157–167 (1977).

Makous, J. C. & Middlebrooks, J. C. Two-dimensional sound localization by human listeners. *J. Acoust. Soc. Am.* **87,** 2188–2200 (1990).

McAlpine, D., Jiang, D. & Palmer, A. R. A neural code for low-frequency sound localization in mammals. *Nat. Neurosci.* **4,** 396–401 (2001).

Middlebrooks, J. C. & Green, D. M. Sound Localization by Human Listeners. *Annu. Rev. Psychol.* **42,** 135–159 (1991).

Mills, A. W. Auditory localization (Binaural acoustic field sampling, head movement and echo effect in auditory localization of sound sources position, distance and orientation). *Foundations of Modern Auditory Theory*. **2,** 303–348 (1972)

Mills, A. W. On the Minimum Audible Angle. *J. Acoust. Soc. Am.* **30,** 237–246 (1958).

Näätänen, R., Gaillard, A. W. & Mäntysalo, S. Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.* **42,** 313–329 (1978).

Pakarinen, S., Takegata, R., Rinne, T., Huotilainen, M. & Näätänen, R. Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP).

Clin. Neurophysiol. **118,** 177–185 (2007).
Parras, G. G. *et al.* Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nat. Commun.* **8,** 2148 (2017).
Rich, D., Cazettes, F., Wang, Y., Peña, J. L. & Fischer, B. J. Neural representation of probabilities for Bayesian inference. *J. Comput. Neurosci.* **38,** 315–323 (2015).
Sams, M., Paavilainen, P., Alho, K. & Näätänen, R. Auditory frequency discrimination and event-related potentials. *Electroencephalogr. Clin. Neurophysiol.* **62,** 437–448 (1985).
Schnupp, J. W., & Carr, C. E. On hearing with more than one ear: lessons from evolution. *Nature Neurosci*, ***12***, 692-697. (2009)
Stern, R. M. & Steven Colburn, H. Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position. *J. Acoust. Soc. Am.* **64,** 127–140 (1978).
Sussman, E. S., Chen, S., Sussman-Fort, J. & Dinces, E. The Five Myths of MMN: Redefining How to Use MMN in Basic and Clinical Research. *Brain Topogr.* **27,** 553–564 (2013).
Sussman, E. S. A New View on the MMN and Attention Debate. *J. Psychophysiol.* **21,** 164–175 (2007).
Tiitinen, H., May, P., Reinikainen, K. & Näätänen, R. Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature* **372,** 90–92 (1994).
van Bergeijk, W. A. Variation on a Theme of Békésy: A Model of Binaural Interaction. *J. Acoust. Soc. Am.* **34,** 1431–1437 (1962).
Woodworth, R. *Experimental Psychology*. (New York: Holt, 1938)
Yin, T. C. & Chan, J. C. Interaural time sensitivity in medial superior olive of cat. *J. Neurophysiol.* **64,** 465–488 (1990).
Yost, W. A. Discriminations of interaural phase differences. *J. Acoust. Soc. Am.* **55,** 1299–1303 (1974).

Supplementary Figures

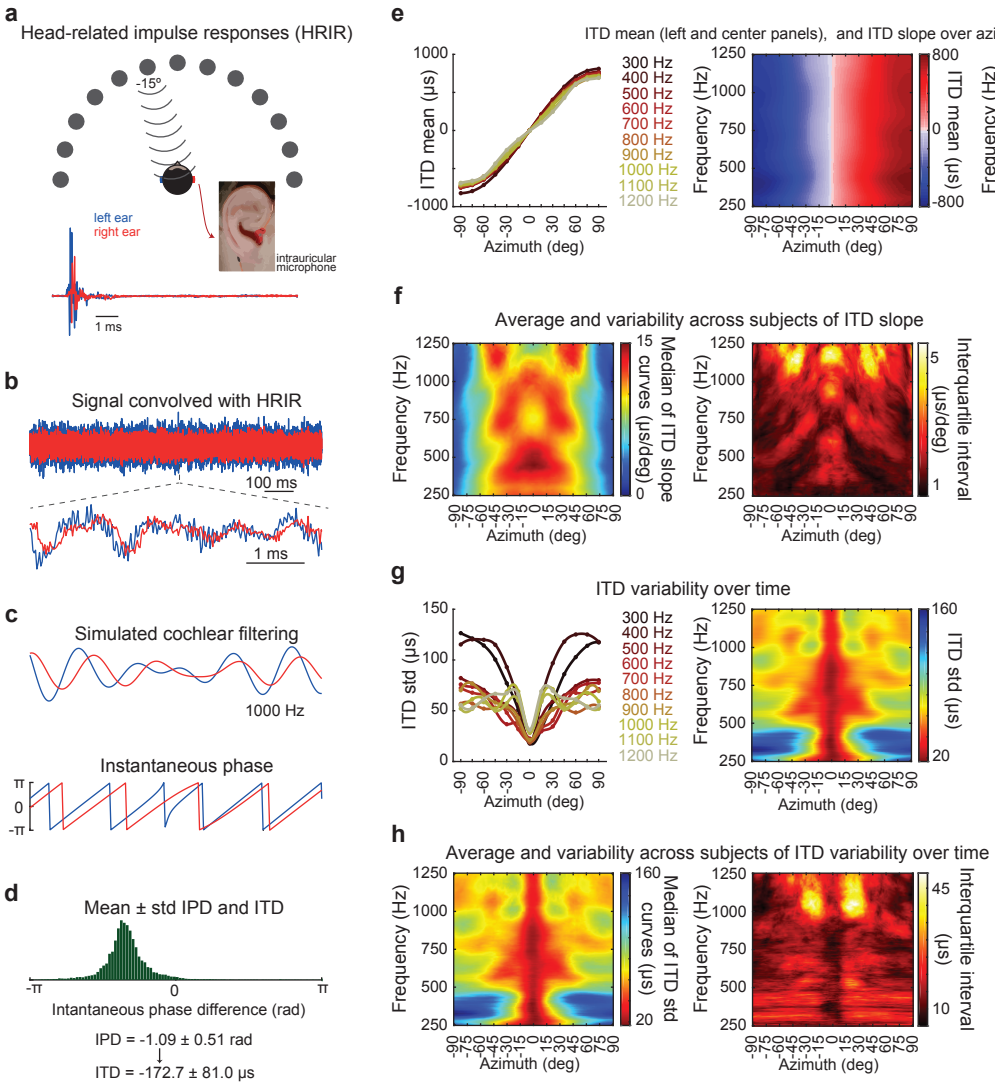

Figure S1. Step-by-step methods for estimating ITD slope and variability. a) Speaker azimuths and binaural microphone placement for collecting head-related impulse responses (top) and example impulse responses in right (red) and left (blue) ears elicited by a sound at azimuth -15º (bottom). **b)** Impulse responses in right (red) and left (blue) ears were convolved with broadband signals to simulate sounds arising from particular directions. **c)** Convolved signals were filtered using parameters analogous to human cochlear filters. Example of filtered signal in a frequency band centered on 1000 Hz (top) and its instantaneous phase (bottom). **d)** Histogram of instantaneous phase differences between left and right signals (IPD), depicting their variability over time. IPD (in radians) was converted to ITD (in microseconds). **e)** Mean ITD over time across subjects; median across subjects are fit by a spline curve, color coded for each frequency (left). These curves were used to calculate mean ITD across frequency (middle). Their derivative was the metric for ITD slope across azimuths (right). **f)** To further verify our results, we used an alternative procedure for computing ITD slope, by fitting the spline curves in panel e to the data from each subject. The median across subjects of the derivative curves (left) is very similar to the ITD slope shown in panel e. This analysis allowed us to assess the variability of the ITD slope across subjects (right). **g-h)** Same procedure described in panels e-f, but for ITD variability over time. The procedures for estimating ITD slope shown in panels e and f are equivalent; the same is valid for the procedures for estimating ITD variability of panels g and h. The second method for estimating ITD variability, also allowed to describe the small variability across subjects (panels f and h).

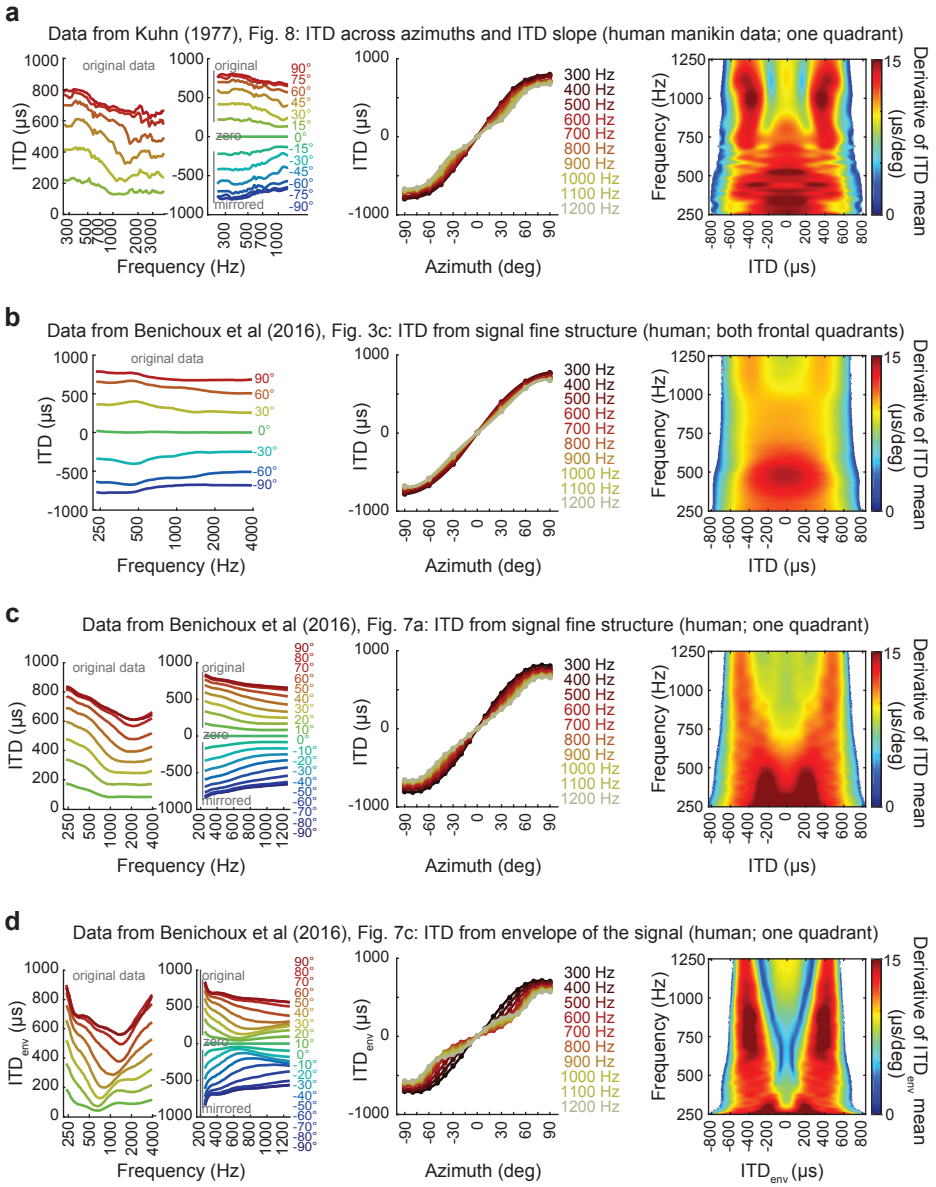

Figure S2. ITD slope across previous studies. a) Data of ITD across frequency and azimuth recorded in human manikins, reported in Kuhn (1977). Plots on the left: original figure and replot of these data adding an assumed straight line at ITD zero and mirrored data representing the contralateral hemifield. Mean ITD across azimuth for each frequency, fit by a spline curve (middle). ITD derivative across azimuths (right). Note the similarity of the pattern of the plot on the right to the ITD slope across location shown in Figures 1a and S1e-f. **b-d)** Same procedure as in panel a, but using the ITD across frequency and azimuth described in Benichoux et al. (2016). The panels depict ITD measurements estimated from interaural phase difference divided by frequency (i.e., the fine-structure of the signal; panels b-c) or envelope of the signal (panel d). Additionally, the panels depict the complete frontal azimuth range (both frontal quadrants; panel a) or collapsing left and right quadrants in a single one quadrant (panels c-d).

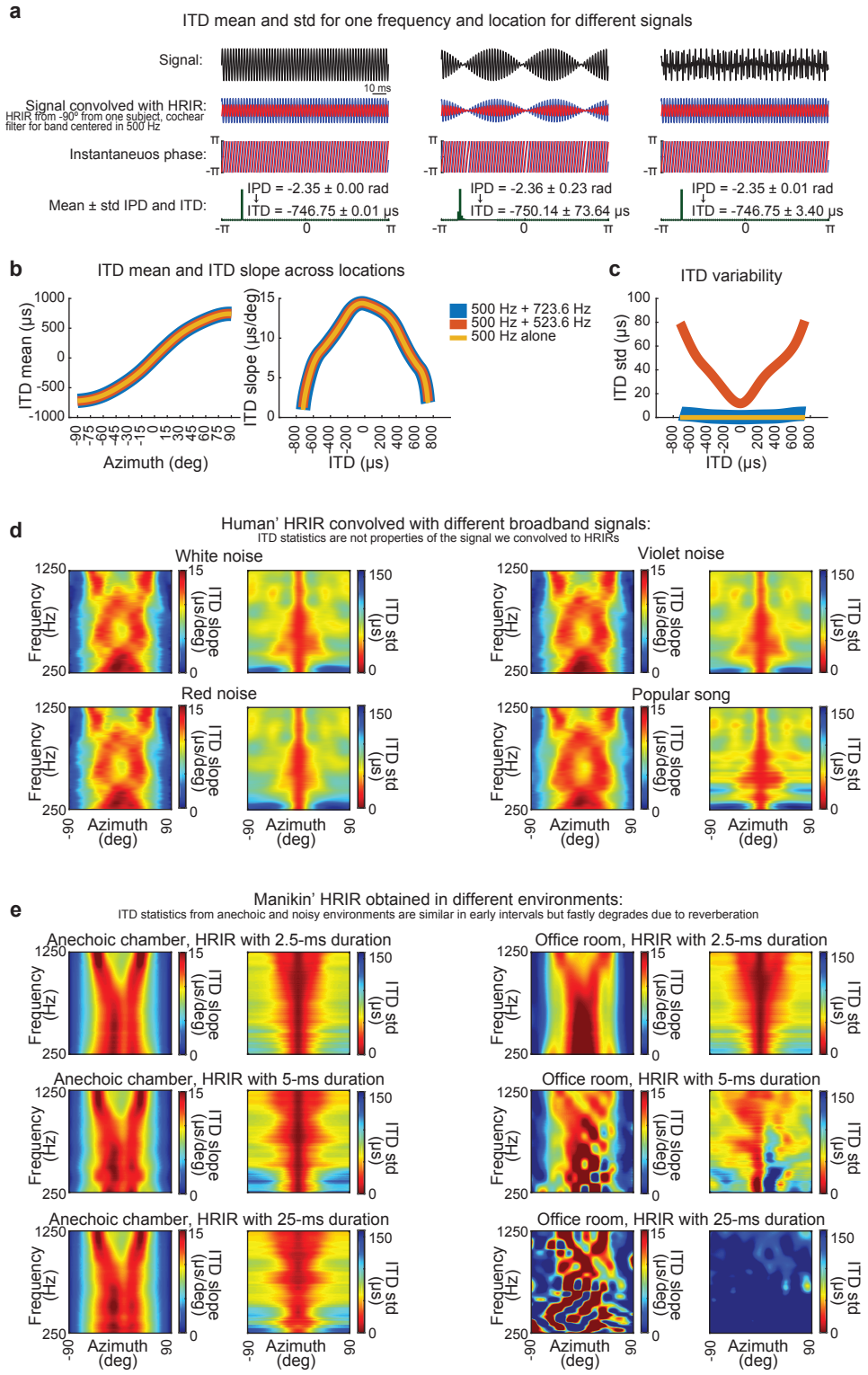

Figure S3. Bases of ITD statistics. a) ITD mean and standard deviation (std) for a sound at -90 deg in azimuth, calculated for three different signals: 500 Hz tone alone (left), 500 Hz combined with 523 Hz (middle), and 500 Hz combined with 723 Hz (right). 523 Hz and 723 Hz lie within and outside the cochlear filter range with center frequency 500 Hz, respectively. We applied the same procedure for calculating mean ITD and std described in Figure S1, by convolving the signals in the right (red) and left (blue) ears with the impulse response and filtering by the corresponding (500 Hz-centered) cochlear filter. Note that the mean ITD (bottom histogram) is virtually identical across conditions, while ITD std is different from zero only in the condition where 500 Hz and 523 Hz are combined. This procedure was applied for all other locations across the 51 subjects reported in the dataset. **b)** Mean ITD (left) and ITD slope (right) across azimuths for locations from -90° to 90° in azimuth, averaged across subjects, for the same signals shown in panel a. Note that the three signals yield the same mean ITD and ITD slope, indicating no effect of frequency interference. **c)** ITD variability over time averaged across subjects. Note that variability is near zero for 500 Hz alone (orange) and 500 + 723 Hz (blue), but increases from front to periphery when neighboring frequencies are mixed (500 + 523 Hz; red), consistently with the ITD variability shown in Figure 1b. This shows that ITD variability is caused by interference between neighboring frequencies. **d)** ITD variability (left) and ITD slope (right) computed for different types of broadband signals (from top to bottom: white noise, red noise, violet noise and a popular song (first second of the chorus of Alive, by Pearl Jam). **e)** ITD variability (left) and ITD slope (right) computed from impulse responses recorded in human-head shaped manikin inside different types of environment and stimulus durations. Anechoic chambers and reverberant office spaces are presented in the left and right columns, respectively. Durations of 2.5 ms (corresponding mostly to the direct click reaching the ears), 5 ms and 25 ms (with progressively larger echo components in the reverberant room) are presented in the top, middle and bottom rows, respectively.

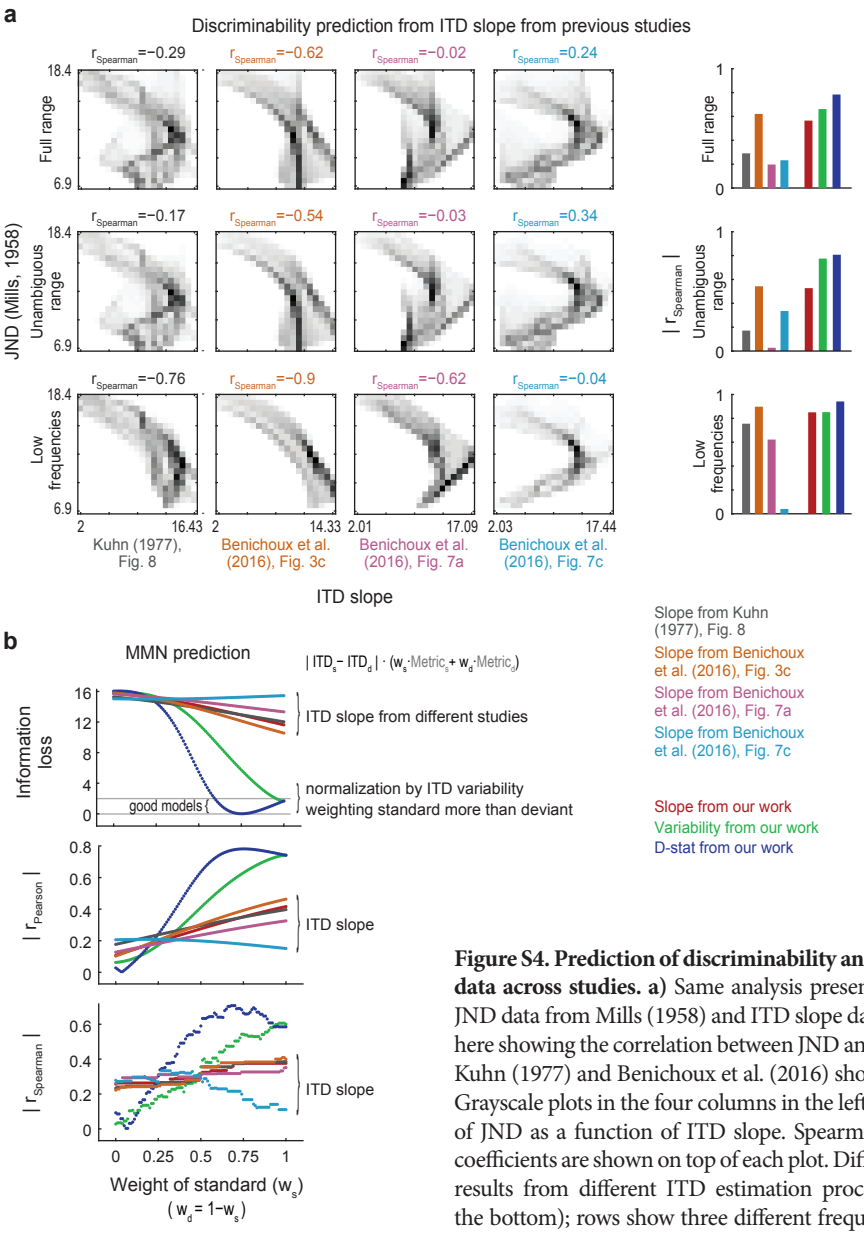

Figure S4. Prediction of discriminability and MMN by ITD slope data across studies. a) Same analysis presented in Figure 2 using JND data from Mills (1958) and ITD slope data from our study, but here showing the correlation between JND and ITD slope data from Kuhn (1977) and Benichoux et al. (2016) shown in Supl. Figure S2. Grayscale plots in the four columns in the left show the distribution of JND as a function of ITD slope. Spearman ranked correlation coefficients are shown on top of each plot. Different columns display results from different ITD estimation procedures (referenced in the bottom); rows show three different frequency and ITD ranges: whole range (top), unambiguous ITD range (middle) and fully unambiguous ITD and low-frequency range (bottom). Bar plots on the right column show correlation coefficients for each of the four datasets, compared to the correlation coefficients obtained for the HRTF metrics of the present study (color code indicated on the right). Note that the D-stat metric (dark blue), which used both ITD slope and ITD variability, shows the highest correlation coefficients for all the frequency and ITD ranges, across datasets. **b)** Same (AIC) information analysis shown in Figure 3 with MMN amplitude indexing discriminability of spatial deviants, but using ITD slope data from Kuhn (1977) and Benichoux et al. (2016). Top, AIC statistical test is used for selecting the model with best predictive power of MMN amplitude; the best models show the smaller dAIC, i.e., smaller information loss. Note that all slope datasets show similar AIC scores and that the best description of MMN amplitude is observed in models that normalize ITD difference by ITD variability and D-stat, weighting standard (w_s) and deviant (w_d) differently. In addition, predictions from all good models show high correlation with MMN amplitude, quantified by both linear and ranked correlation metrics (middle and bottom).

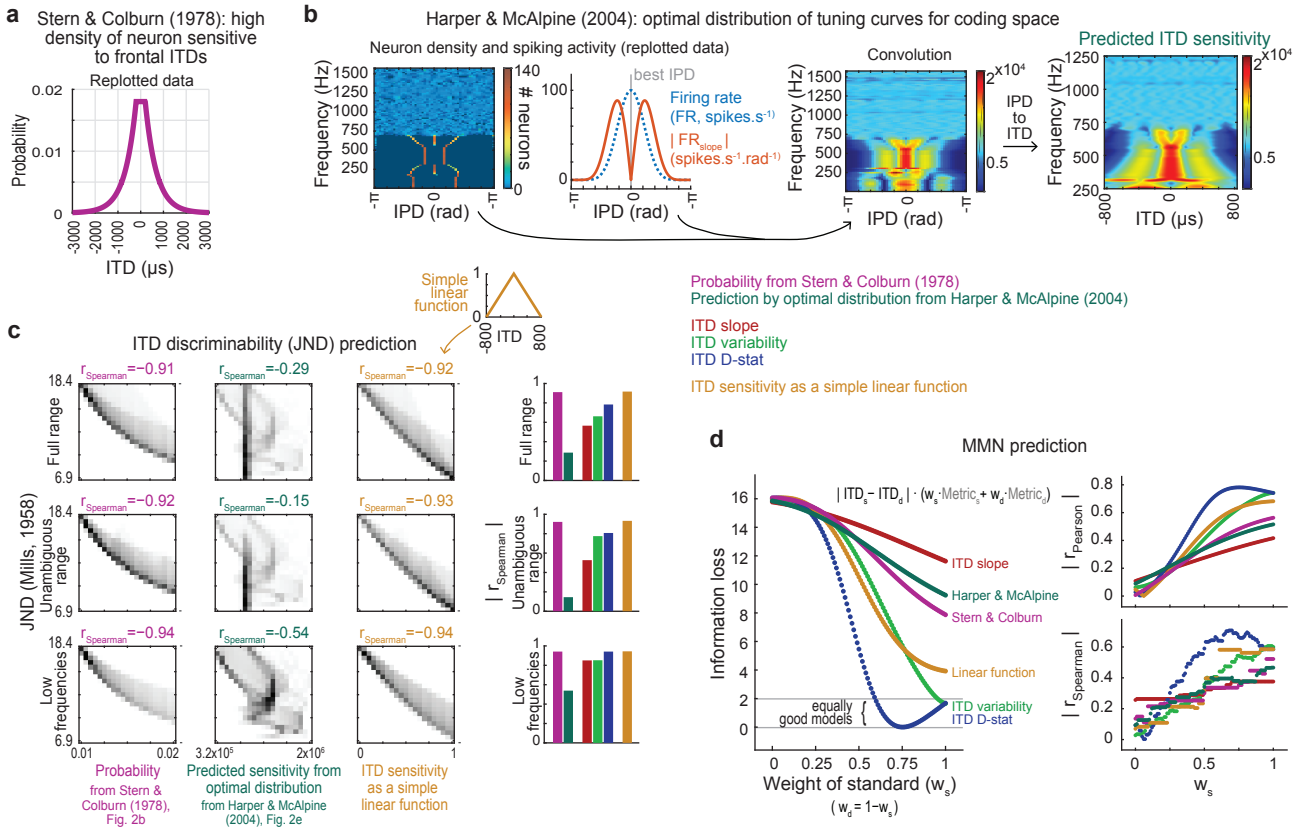

Figure S5. JND and MMN prediction by ITD discriminability predicted by models of brainstem-neurons response properties. a) Probability function replotted from Stern & Colburn (1978), which proposes higher density of ITD-sensitive cells in the MSO near the midline. **b)** IPD tuning distribution maximizing information carried by firing rate (FR) change across the midline, proposed by Harper & McAlpine (2004). Distribution of best IPD tunings (left-most panel) and firing rate slope as a function of IPD (second panel), replotted from the original article. The predicted IPD sensitivity was calculated by convolving the tuning distribution with the FR slope (third panel), which was converted to ITD (right-most panel). **c)** Same analysis presented in Figure 2, but testing the predictive power of JND data (Mills 1958) of the models shown in panels a and b, and by simple linear function describing ITD sensitivity, varying from 0 in the periphery to 1 in the midline. Bar plots on the right show correlation coefficients compared to the correlation coefficients obtained for the ITD statistics of the present study. **d)** Same analysis shown in Figure 3, but here testing the predictive power of MMN amplitude by the models presented in panels a and b, compared with the ITD statistics models of the present study.

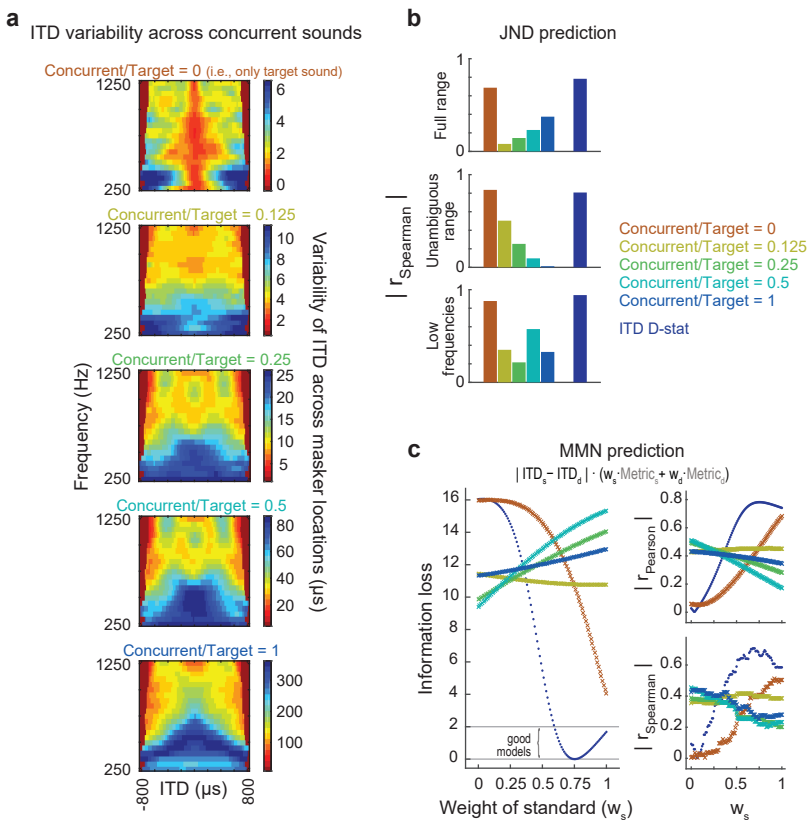

Figure S6. Estimation of ITD statistics across concurrent sound conditions. a) ITD variability caused by combining target sounds with a second sound. For each target location (from -90 to 90 degree in 15-deg intervals and zero elevation), we mixed the target signal with concurrent sounds from each of the 13 directions. From the mean ITD over time of these combinations, we calculated the standard deviation across concurrent sound locations, as the estimate of ITD variability across concurrent sounds. We used concurrent sounds of different amplitudes relative to the target, ranging from a ratio of 0 (absent concurrent sound, top) to 1 (equal amplitude of target and concurrent sounds, bottom). Variability estimates across locations are plotted as a function of the target ITD. Color scales are set to the range of mean ± 2*std. Note that both the pattern and scale of ITD variability changes with the amplitude of the concurrent sounds. As expected, the pattern of ITD variability without concurrent sounds (top) is equivalent to the ITD variability shown in Figure 1b. **b)** Predictive power of JND (Mills 1958) by a model using; ITD variability across concurrent sounds. Bar plots show correlation coefficients of statistics presented in panel a and JND, and between D-stat and JND (blue bars on the right). Except for the condition with concurrent sound with amplitude zero (orange bar), which is equivalent to ITD variability of the present study (see Figure 2d), JND predictions by ITD variability across concurrent sounds are poor. **c)** AIC analysis (as shown in Figure 3), testing the predictive power of MMN amplitude by ITD variability across concurrent sounds, compared with predictions by the D-stat model of the present study. Except for the condition with concurrent sounds with amplitude zero (orange symbols), MMN prediction by the models using ITD variability across concurrent sounds is poor.