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 components of ITD statistics from human head-related transfer functions (HRTFs), which can be assumed invariant across contexts. 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 are anticipated by the human brain, influencing spatial discriminability and novelty detection. Previously reported thresholds for discriminating ITD changes were 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 variability of the standard location. 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; Reliability; 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[1]. ITD is detected by auditory brainstem neurons within narrow frequency bands[2–5].

Classical psychophysical studies demonstrated that humans detect sound location better in the front than in the periphery[6–8]. The enhanced capability in frontal locations could be efficient for hunting and foraging, as analogously proposed for vision[9–11]. Physiological evidence indicates that coding of binaural spatial cues could support finer discriminability in the front[5,12–14]. 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[13,15,16].

For azimuth detection based on ITD, sound diffraction by structures surrounding the ear modulates the ITD vs azimuth relationship[17]. In addition, because coincidence detector neurons compute ITD in narrow frequency bands[3,4], the interaction of nearby frequencies may be an important source of ITD variability. Previous studies in barn owls have shown that variability of binaural cues is low in the front and high in the periphery, and varies across frequency[18,19]. 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[20].

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

Results

ITD statistics (i.e., the derivative of the mean ITD over azimuth and the standard deviation of ITD) were estimated from human HRTFs. We tested whether the ITD statistics predict human spatial discrimination thresholds[6]. Next, we tested if natural ITD statistics are built into spatial deviants detection using EEG and mismatch negativity signals (MMN).

ITD statistics estimated from human HRTFs

We estimated ITD statistics from human HRTF databases acquired in anechoic environments to test whether statistics invariant across contexts are built into brain processing (Listen HRTF database, http://recherche.ircam.fr/equipes/salles/listen; 51 subjects).

The rate of change of ITD (ITD slope) across location and frequency is shown in Fig. 1a (see Suppl. Fig. 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[21,22], it was already observable (Suppl. Fig. S2), indicating that it is consistent across approaches.

Second, we estimated the ITD standard deviation over time (ITD variability; Fig. 1b and Suppl. Fig. S1), calculated from the IPD variation along broadband signals convolved with the impulse responses. We confirmed that this variability results from the interaction between

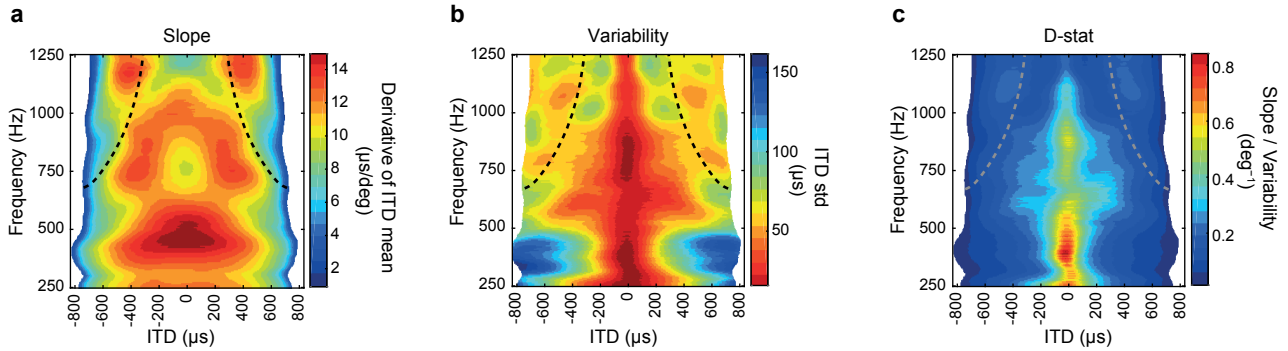

Fig. 1. Natural ITD statistics from human HRTFs. **a)** ITD slope across azimuth and frequency, plotted as a function of the ITD corresponding to each azimuth. As previously reported 16,23, the slope is higher mostly in the front, although the pattern is complex. **b)** ITD variability (standard deviation 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.

neighboring frequencies, within a single cochlear filter (Suppl. Fig. S3a-c); the physical basis of this variability is presumably differential gain modulation in neighboring frequencies affecting ongoing interaural phase difference. ITD standard deviation was consistent across broadband 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. Fig. S3d).

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

Prediction of spatial discrimination thresholds from ITD statistics

ITD just-noticeable difference (JND) as a function of azimuth and ITD was estimated from classic published data on human sound localization[6] (Fig. 2a,b).

We tested whether ITD slope, ITD variability or their combination (ITD D-stat) predicted the JND. Fig. 2c shows the JND as a function of ITD slope, ITD variability and ITD D-stat across the full frequency and ITD ranges (Fig. 2c, top), for the ITD range within the pi-limit (removing the phase-ambiguous range; Fig. 2c, middle) and for the low frequency range (Fig. 2c, bottom). JND was significantly correlated with ITD slope (Fig. 2c, left column) and ITD variability (Fig. 2c, middle column) but the highest correlation coefficients were found for ITD D-stat (Fig. 2c, right column). Correlation coefficients for each plot are shown in Fig. 2d. These results suggest that spatial discrimination thresholds are determined by expected ITD statistics, and therefore a brain representation of these statistics exists. ITD D-stat was also the best predictor of JND when ITD slope data from previous studies were used[21,22] (Suppl. Fig. S4a).

To consider proposed mechanisms supporting greater ITD sensitivity at the midline, we first verified whether a linear function of ITD (Suppl. Fig. S5a) predicted JND. This was expected by the monotonic function adopted by Mills[6] to fit the data. Additionally, we considered previously described mechanistic brainstem models that also predicted greater ITD sensitivity in the front. Specifically, Stern and Colburn (1978)[24] proposed an exponential (also monotonic) increase in density of ITD sensitive cells towards the midline (Suppl. Fig. S5b) that is consistent with increased sensitivity in the front. Similarly, Harper and McAlpine (2004)[25] showed that the brainstem ITD coding that maximizes Fisher information (Suppl. Fig. S5c) also predicts increased sensitivity in the front at low frequencies. These models showed varied levels of predictive power of Mills (1958)[6] JND data (Suppl. Fig. S5d). These results indicate that JND can be linked to properties of the network supporting sound localization, which is not inconsistent with the claim that ITD statistics are represented in the brain. Furthermore, previous work in the avian brain has postulated non-uniform cell density and ITD tuning as coding elements underlying Bayesian inference in sound localization[26,27].

ITD statistics built into deviance detection

To further test if statistics are built into the human brain processing, we recorded EEG mismatch negativity responses[29] for spatial deviants in tones presented through earphones. The inserted earphones removed the natural ITD variability induced by the filtering effect of the head and ears. 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 stimuli is unexpectedly interrupted by a low probability deviant stimulus[29–32] (Fig. 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[30,33–35].

We measured MMN responses in 33 subjects. A set of ITD and frequency conditions was selected in order to

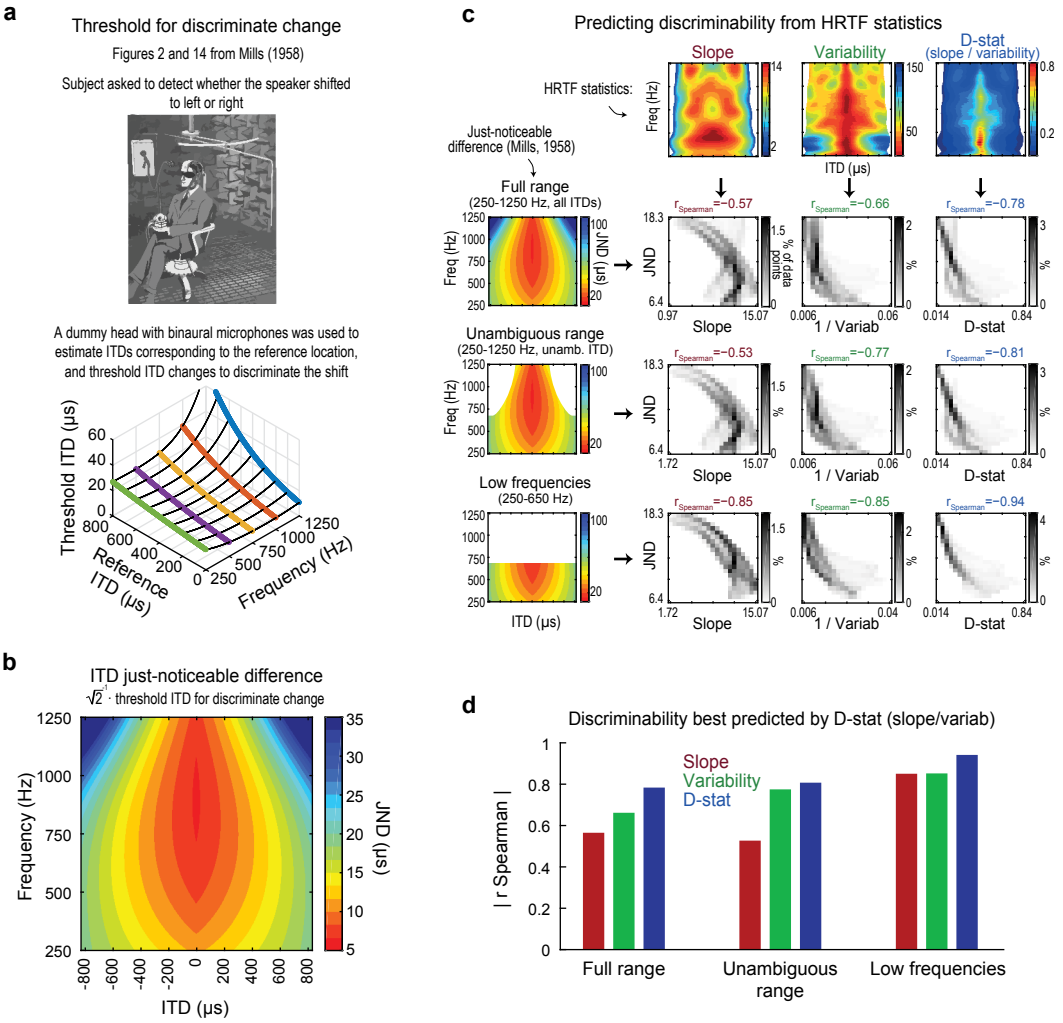

Fig. 2. Human detection thresholds of ITD changes match expected ITD statistics. **a)** Mills (1958) estimated the minimum ITD change (threshold ITD, proportional to JND) that could be detected across frequency and location (top). Azimuth was converted to ITD using a dummy head with binaural microphones (bottom). Reproduced from Mills[6]; 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 threshold estimations from Mills (1958)[6] data by the square root of 2 to take into account reported underestimation of JND by minimum audible angle measurements[28]. 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.

sample critical ranges drawn from the HRTF analysis (Fig. 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[36] (Fig. 3c). The peak amplitude of MMN was used to quantify the effect of stimulus manipulations on novelty detection.

The MMN peaks averaged across subjects showed weak negative correlation to the ITD difference (absolute difference between the ITDs of standard and deviant stimuli) (Fig. 3d). This result is consistent with studies demonstrating a correlation between MMN amplitude and ITD difference between standard and deviant[30]. However, we observed a strong correlation between ITD difference and MMN amplitude when the ITD of the standard was frontal (Fig. 3d, right) and no correlation with the standard in the most peripheral location. This can be explained by the absence of MMN when the standard is at the most peripheral location, which shows that this relationship depends not only on the ITD difference but on the location of the standard. Consistently, standard stimuli located in the front led to more negative MMNs than deviants in the front (Fig. 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 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

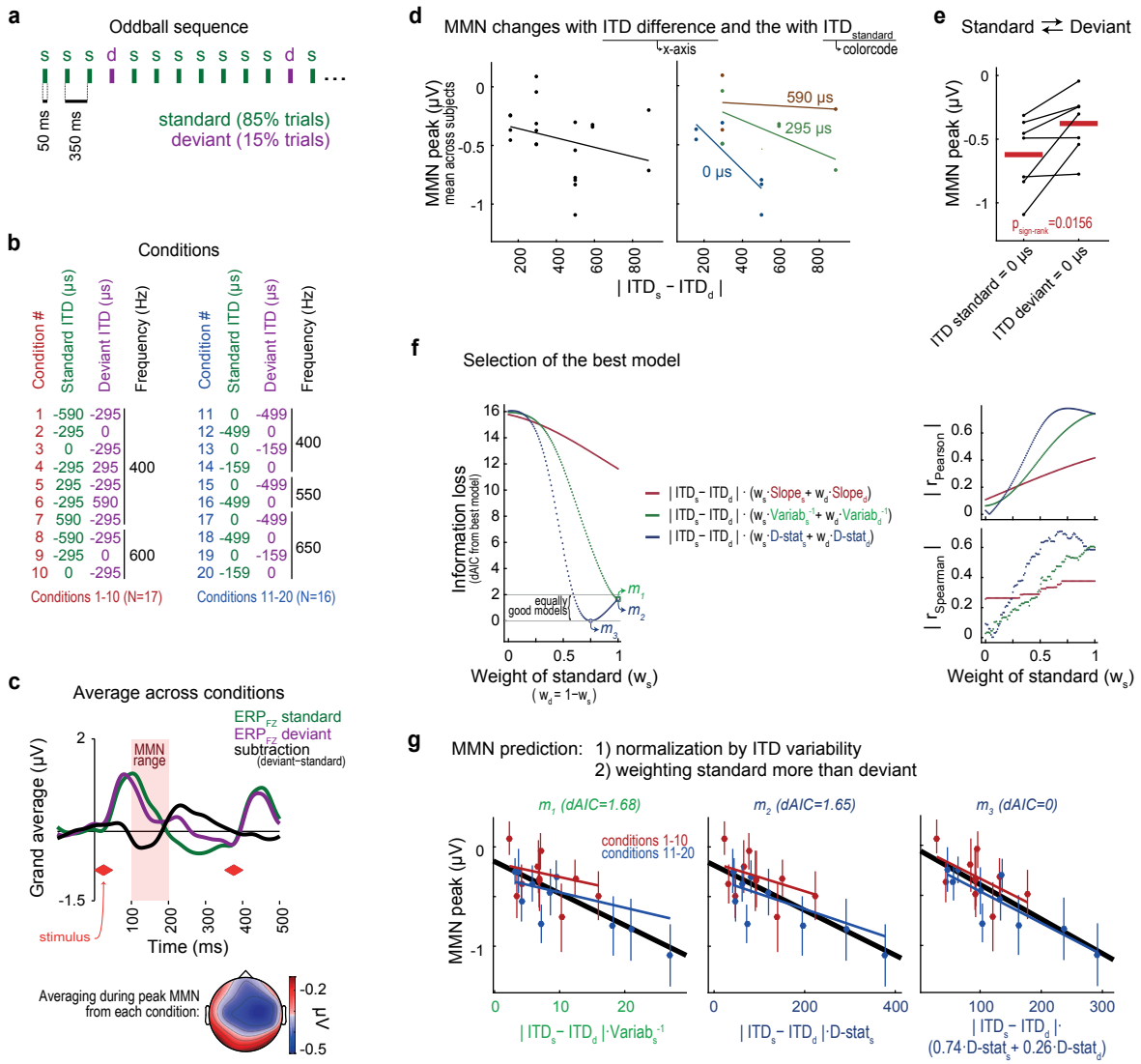

Fig. 3. Discriminability of spatial deviants indexed by MMN responses is weighted by expected ITD variability at the standard location. **a**) Passive oddball sequence protocol, in which subjects listened to frequent "standard" stimuli embedded with rare "deviants." **b**) In each condition, two tones 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), classifying the data by the ITD of the standard (0 µs, 295 µs and 590 µs), exposed different degrees of association (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.

between standard and deviant by the ITD slope (Fig. 3f, red), ITD variability (Fig. 3f, green) or ITD D-stat (Fig. 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"; Fig. 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"; Fig. 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 performed this analysis using ITD slope described in previous studies[21,22]. These models showed performance similar to the ITD slope from our study, with poorer prediction compared to models that accounted for ITD variability (Suppl. Fig. 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 and Colburn (1978)[24] (Suppl. Fig. S5b), and (2) the firing rate changes across ITD weighted by the distribution of cells reported in Harper and McAlpine (2004)[25] (Suppl. Fig. S5c). Normalizing the ITD difference between standard and deviant by either of these ITD sensitivity metrics lowered the predictive power relative to the models relying on ITD variability (Suppl. Fig S5e). However, these models are not incompatible, since cell density and tuning may represent ITD statistics[26,27].

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 (ITD variability), across frequency and ITD. ITD slope and ITD variability, and their ratio (ITD D-stat) were used to test whether 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[24,37] or greater variation in firing rate as a function of azimuth in the front compared to the periphery[5,25]. These mechanisms 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[16,23]. 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.

Previous reports have investigated the selectivity of midbrain neurons to the variability of spatial cues in the owl's auditory system[18,19]. These studies provided evidence of how sensory reliability could be represented[26,27,38] and integrated into an adaptive behavioral command[39]. Although the coding observed in owls cannot be assumed for all species, the present study shows that ITD statistics also affect human sound localization. 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 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[16,23]. This interpretation does not consider ITD variability. Here we show that natural 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[6] consisted of tones, the across-frequency interaction was not a source of ITD variability. On the other hand, Mills[6] 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'[6] 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'[6] 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 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[6], 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[6], 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)[40] 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[6], is influenced by attention and training; the novelty detection protocol allowed for control of these factors. 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 preserved. 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 idea that high-order statistics of sensory information relevant to sound localization are built into human brain processing.

Acknowledgements

This work was supported by NIH BRAIN Initiative (Grant NS104911) and by the NIDCD (Grant R01 DC004263). We thank Ruben Coen Cagli, Gervasio Batista and Michael Beckert 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).

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. Fig. 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 1 and that corresponds to range that have the thresholds estimated[6]), and bandwidth equivalent to the estimated bandwidth for human cochlear filters[41]. 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 (Fig. 1a).

We computed the median across subjects of the ITD variability for each azimuth and frequency sampled (Suppl. Fig. S1). These data were then interpolated across azimuth and represented as a function of ITDs (Fig. 1b). Although both ITD slope and ITD variability are likely affected by diffraction and interference between neighboring frequencies, their correlation was not strong ($r_{Pearson}$=-0.41; $r_{Spearman}$=-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 (Fig. 1c). Note that d-prime and D-stat are different quantities.

Estimation of spatial discriminability thresholds

Human spatial discriminability thresholds were estimated from the classic Mills[6] 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 (Fig. 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[6] 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[6]. 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[6] and allowed us to confirm that our value extraction was correct (Fig. 2a, bottom). Hartmann and Rakerd (1989)[28] demonstrated that the 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 (Fig. 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[6]; (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) (Fig. 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 (Fig. 3a). Each subject was presented with 10 conditions, which differed in frequency and ITD (Fig. 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-Hertz 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 (Fig. 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 (Fig. 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 (Fig. 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 peak was performed with linear mixed-effect models, for each condition and 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.

References

1. Middlebrooks, J. C. & Green, D. M. Sound Localization by Human Listeners. *Annu. Rev. Psychol.* **42**, 135–159 (1991).
2. 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).
3. Yin, T. C. & Chan, J. C. Interaural time sensitivity in medial superior olive of cat. *J. Neurophysiol.* **64**, 465–488 (1990).
4. 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).
5. McAlpine, D., Jiang, D. & Palmer, A. R. A neural code for low-frequency sound localization in mammals. *Nat. Neurosci.* **4**, 396–401 (2001).
6. Mills, A. W. On the Minimum Audible Angle. *J. Acoust. Soc. Am.* **30**, 237–246 (1958).
7. Yost, W. A. Discriminations of interaural phase differences. *J. Acoust. Soc. Am.* **55**, 1299–1303 (1974).
8. Makous, J. C. & Middlebrooks, J. C. Two-dimensional sound localization by human listeners. *J. Acoust. Soc. Am.* **87**, 2188–2200 (1990).
9. 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).
10. Cartmill, M. Pads and Claws in Arboreal Locomotion. in *Primate Locomotion* 45–83 (1974).
11. Changizi, M. A. & Shimojo, S. 'X-ray vision' and the evolution of forward-facing eyes. *J. Theor. Biol.* **254**, 756–767 (2008).
12. van Bergeijk, W. A. Variation on a Theme of Békésy: A Model of Binaural Interaction. *J. Acoust. Soc. Am.* **34**, 1431–1437 (1962).
13. 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).
14. Grothe, B., Pecka, M. & McAlpine, D. Mechanisms of Sound Localization in Mammals. *Physiol. Rev.* **90**, 983–1012 (2010).
15. Woodworth, R. *Experimental Psychology.* (New York: Holt, 1938)
16. Gelfand, S. A. *Hearing: An Introduction to Psychological and Physiological Acoustics.* (CRC Press, 2016).
17. 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).
18. Cazettes, F., Fischer, B. J. & Pena, J. L. Spatial cue reliability drives frequency tuning in the barn Owl's midbrain. *Elife* **3**, e04854 (2014).
19. Fischer, B. J. & Peña, J. L. Optimal nonlinear cue integration for sound localization. *J. Comput. Neurosci.* **42**, 37–52 (2017).
20. Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics.* (1966).
21. Kuhn, G. F. Model for the interaural time differences in the azimuthal plane. *J. Acoust. Soc. Am.* **62**, 157–167 (1977).
22. Benichoux, V., Rébillat, M. & Brette, R. On the variation of interaural time differences with frequency. *J. Acoust. Soc. Am.* **139**, 1810 (2016).
23. 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)
24. 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).
25. Harper, N. S. & McAlpine, D. Optimal neural population coding of an auditory spatial cue. *Nature* **430**, 682–686 (2004).
26. Fischer, B. J. & Peña, J. L. Owl's behavior and neural representation predicted by Bayesian inference. *Nat. Neurosci.* **14**, 1061–1066 (2011).
27. 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).
28. Hartmann, W. M. & Rakerd, B. On the minimum audible angle—A decision theory approach. *J. Acoust. Soc. Am.* **85**, 2031–2041 (1989).
29. Näätänen, R., Gaillard, A. W. & Mäntysalo, S. Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.* **42**, 313–329 (1978).
30. 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).
31. Sussman, E. S. A New View on the MMN and Attention Debate. *J. Psychophysiol.* **21**, 164–175 (2007).
32. 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).

33. 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).
34. Sams, M., Paavilainen, P., Alho, K. & Näätänen, R. Auditory frequency discrimination and event-related potentials. *Electroencephalogr. Clin. Neurophysiol.* **62**, 437–448 (1985).
35. 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).
36. 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).
37. 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).
38. 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).
39. 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).
40. Parras, G. G. et al. Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nat. Commun.* **8**, 2148 (2017).
41. Glasberg, B. R. & Moore, B. C. Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* **47**, 103–138 (1990).

Supplementary Figures

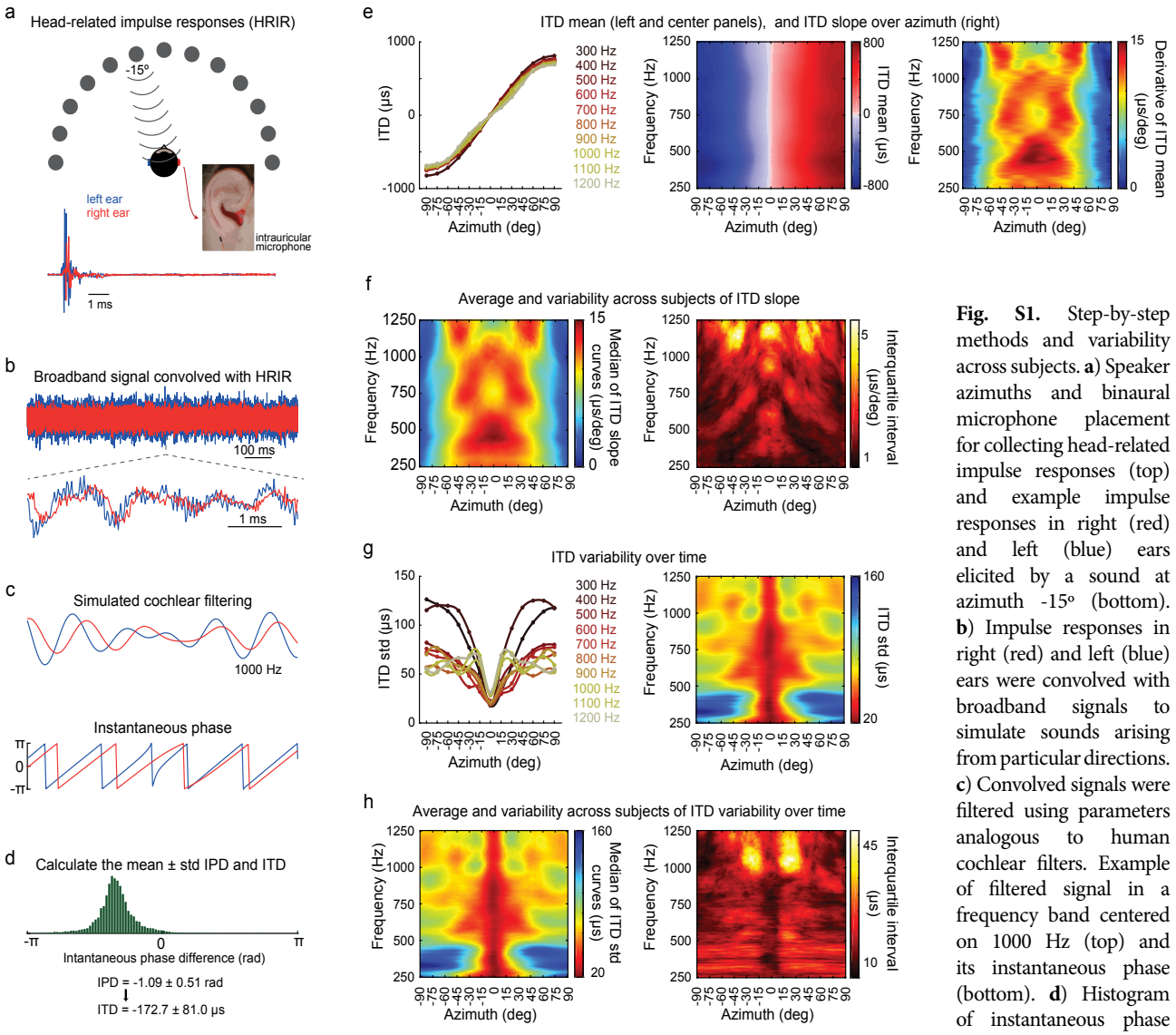

Fig. S1. Step-by-step methods and variability across subjects. **a**) Speaker azimuths and binaural microphone placement for collecting head-related impulse responses (top) and example impulse responses in right (red) and left (blue) ears elicited by a sound at azimuth -15º (bottom). **b**) Impulse responses 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 e to the data from each subject. The derivative of the median curve across subjects (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).

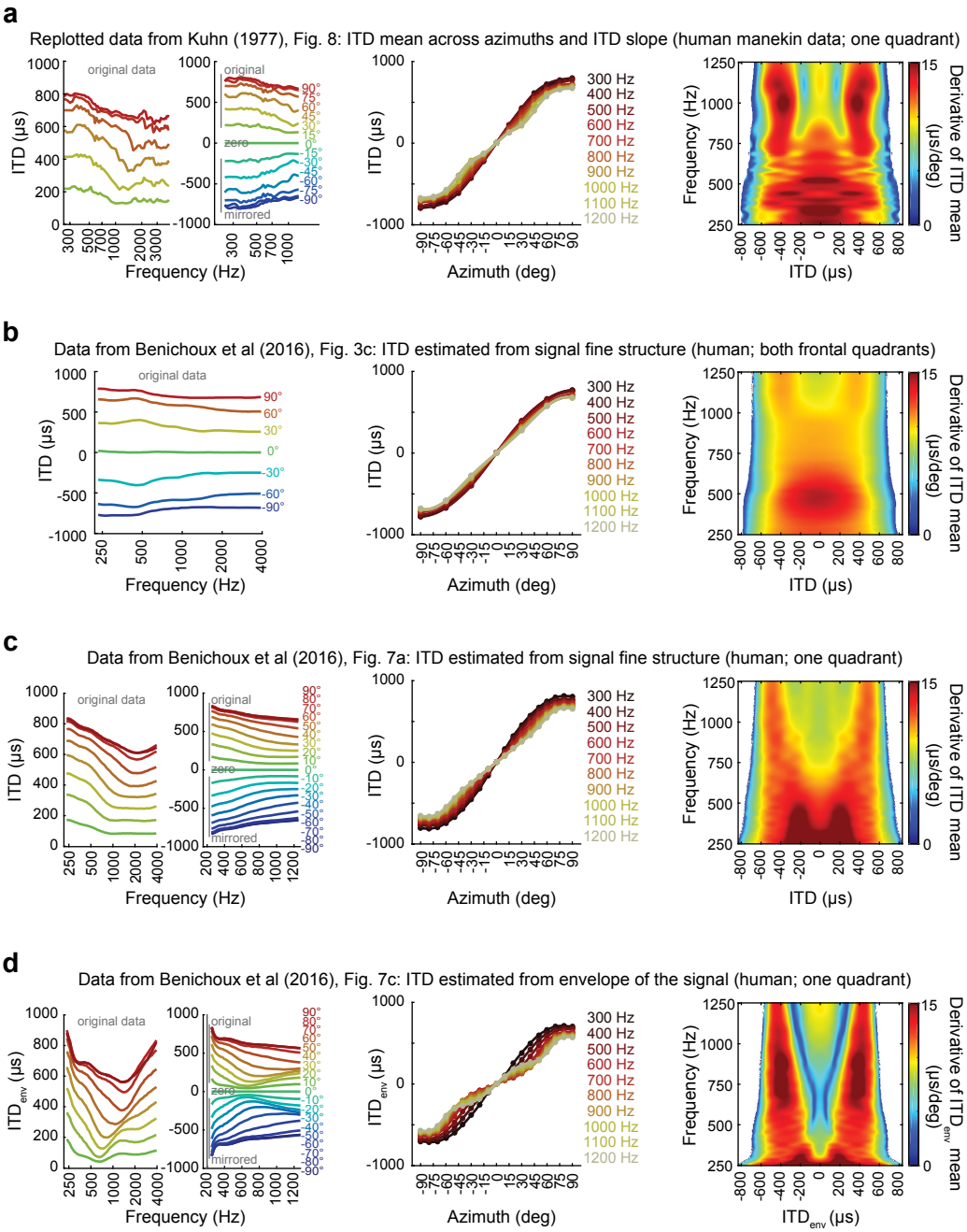

Fig. S2. ITD slope across reported datasets. **a**) Data of ITD across frequency and azimuth recorded in human manikins, reported in Kuhn (1977)[21]. Plots on the left: original figure and replot of these data adding an assumed straight line at ITD zero and mirrored data representing the contralateral 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)[22]. 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).

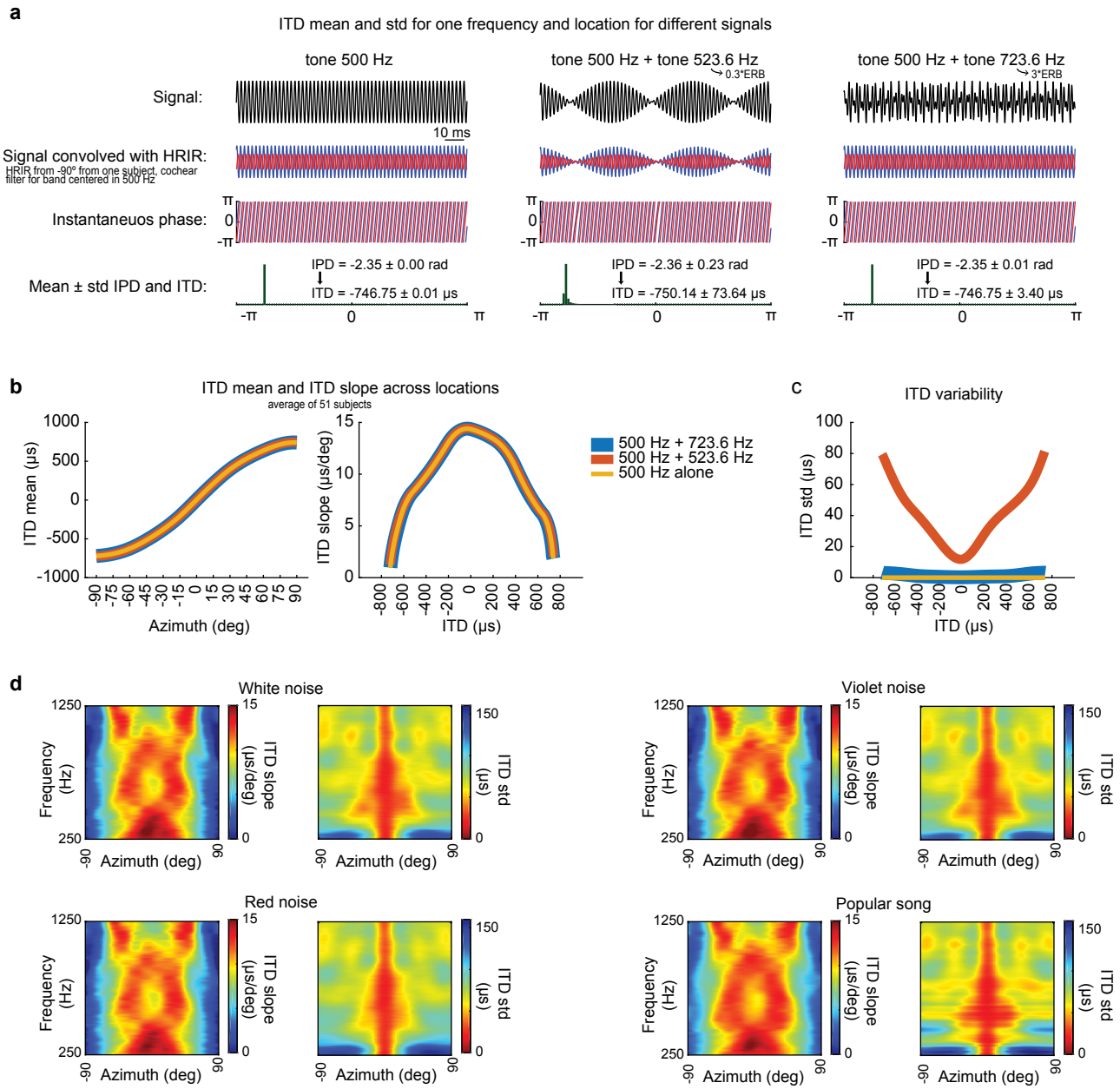

Fig. S3. Estimation of ITD statistics. **a**) ITD mean and standard deviation (std) for a sound at -90° in azimuth, calculated for three different signals: 500 Hz tone alone (left), 500 Hz combined with 523 Hz (middle), and 500 Hz combined with 723 Hz (right). 523 Hz and 723 Hz lie within and outside the cochlear filter range with center frequency 500 Hz, respectively. We applied the same procedure for calculating mean ITD and std described in Fig. 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 Fig. 1b. This shows that ITD variability is caused by interference between neighboring frequencies. **d**) ITD variability (left column) and ITD slope (right column) 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).

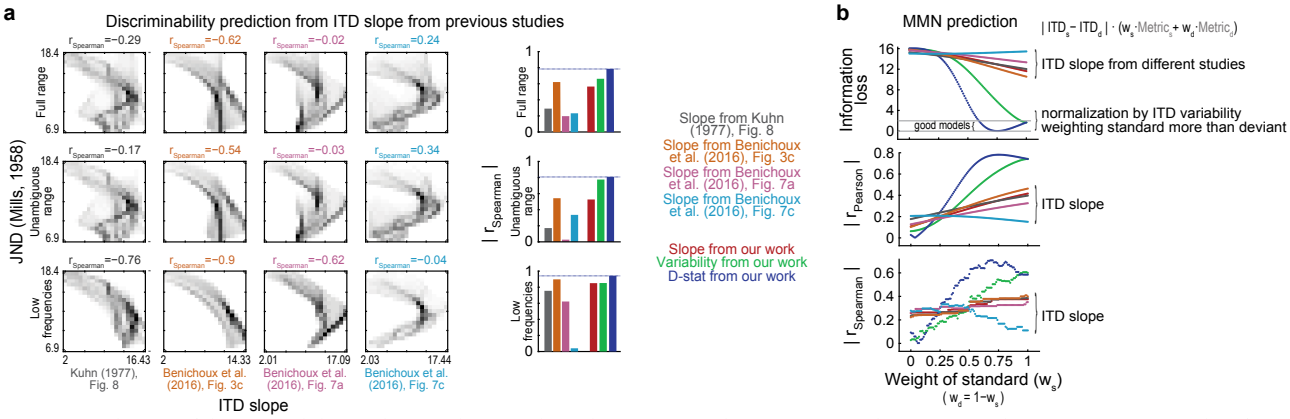

Fig. S4. Prediction of discriminability and MMN by ITD slope data across studies. **a**) Same analysis presented in Fig. 2 using JND data from Mills (1958)[6] and ITD slope data from our study, but here showing the correlation between JND and ITD slope data from Kuhn (1977)[21] and Benichoux et al. (2016)[22] shown in Supl. Fig. 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 Fig. 3 with MMN amplitude indexing discriminability of spatial deviants, but using ITD slope data from Kuhn (1977)[21] and Benichoux et al. (2016)[22]. 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).

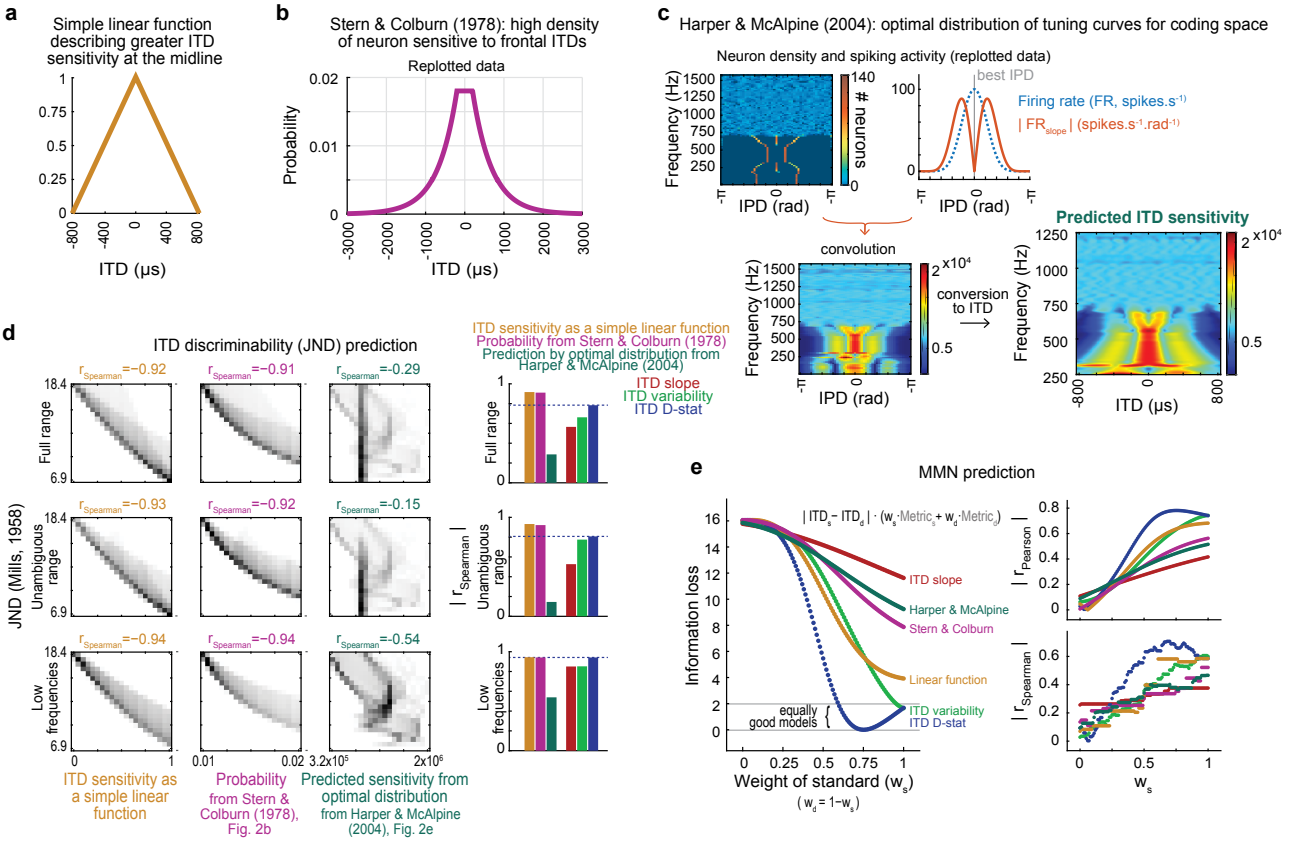

Fig. S5. JND and MMN prediction by a linear model of ITD discriminability and by ITD discriminability predicted by models of brainstem-neurons response properties. **a**) Simple linear function describing ITD sensitivity, varying from 0 in the periphery to 1 in the midline. **b**) Probability function replotted from Stern & Colburn (1978)[24], which proposes higher density of ITD-sensitive cells in the MSO near the midline. **c**) IPD tuning distribution maximizing information carried by firing rate (FR) change across the midline, proposed by Harper & McAlpine (2004)[25]. Top: distribution of best IPD tunings and firing rate slope as a function of IPD, replotted from the original article. The predicted IPD sensitivity was calculated by convolving the tuning distribution with the FR slope (bottom left), which was converted to ITD (bottom right). **d**) Same analysis presented in Fig. 2, but testing the predictive power of JND data[6] of the models show in panels a-c. Bar plots on the right show correlation coefficients compared to the correlation coefficients obtained for the ITD statistics of the present study. **e**) Same analysis shown in Fig. 3, but here testing the predictive power of MMN amplitude by the models presented in panels a-c, compared with the ITD statistics models of the present study.