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; rodrigo.pavao@ufabc.edu.br

The variability of natural scenes places perception in the realm of statistical inference. In this context, 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, where interaural time difference (ITD) is a key sensory cue for inferring the direction of acoustic signals in the horizontal plane. ITD statistics were estimated from human head-related transfer functions (HRTFs) and properties of cochlear filters. As previously shown, ITD changed with azimuth following a sigmoid relationship, whose slope was steepest at the center in most frequencies. However, 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 rate of change with azimuth (ITDrc) and ITD variability (ITDv), are anticipated by the human brain, influencing spatial discriminability and novelty detection. Our results showed that thresholds for discriminating ITD changes could be better predicted by a model relying on both ITDrc and ITDv than on ITDrc alone. In addition, we tested how novelty detection in a spatial oddball paradigm was weighted by anticipated ITD statistics, as indexed by the amplitude of the mismatch negativity (MMN) brain signal, a pre-attentive indicator of novelty detection. ITDrc and ITDv predicted MMN amplitude, showing a significantly stronger effect of ITD statistics corresponding to repeated (standard) than sporadic (deviants) stimuli. We further show that these statistics could be represented in parameters underlying ITD discriminability postulated by classic neural models. 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.

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

Introduction

Humans and other species localize sound sources in the horizontal plane using sub-millisecond interaural time difference (ITD) between signals arriving at the two ears (Middlebrooks and Green 1991). ITD is detected by auditory brainstem neurons within narrow frequency bands (Goldberg and Brown 1969; Yin and Chan 1990; Carr and 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 and 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 and 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 can affect the ITD vs azimuth relationship (Aaronson and Hartmann 2014; Roth et al. 1980). In addition, because coincidence detector neurons compute ITD in narrow frequency bands, the interaction of nearby frequencies within a given cochlear filter may also be a source of ITD variability over time (ITDv). 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 stimuli but also on the variability of the sensory evidence (Green and Swets 1966).

This study, focusing on spatial hearing, tested the hypothesis that there is a built-in representation of environmental statistics in the human brain. Using human HRTF databases, we estimated ITD rate of change across azimuth (ITDrc) and ITDv. We tested whether these statistics, separately and combined, predicted spatial discrimination thresholds and deviant detection. Using sound stimuli that removed ITDv (tones delivered by earphones inserted in the ear canal) allowed us to test whether anticipated, rather than actual, statistics predicted novelty detection. Our results indicate that expected ITDrc and ITDv 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 (rate of change) of the mean ITD over azimuth (ITDrc) and the standard deviation of ITD (ITDv) over time were estimated from human HRTFs. We first tested whether these ITD statistics predict human spatial discrimination thresholds using previously published datasets (Mills 1958) and data collected from specifically designed tests for measuring spatial discrimination. Next, we used EEG and mismatch negativity signals (MMN) to address the question of whether these natural ITD statistics are built into spatial deviant detection.

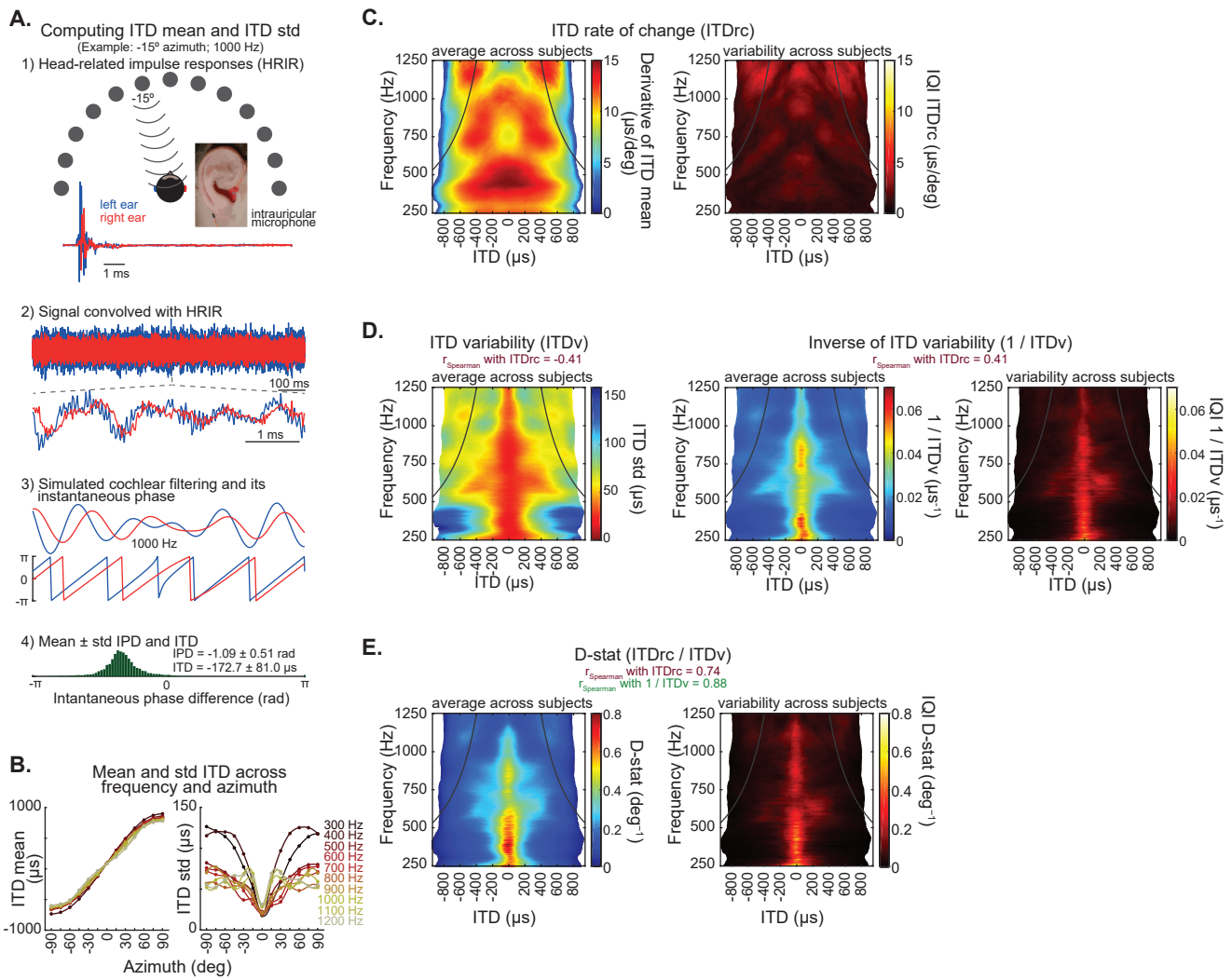

Figure 1. ITD statistics estimated from human HRTFs. (**A**) Estimation of ITD mean and ITD standard deviation (std): (1) Head-related impulse responses (HRIR) were collected by binaural microphones receiving signals from speakers located at different azimuth positions. Example impulse responses in right (red) and left (blue) ears elicited by a sound at azimuth −15º. (2) Overlaid broadband signals were convolved with impulse responses of the right (red) and left (blue) ears from each particular direction. Inset shows an amplified segment (indicated by dashed lines) of both traces. (3) 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). (4) Histogram of instantaneous phase differences between left and right signals (IPD), illustrating their variability over time for the example signal from −15º after cochlear filtering centered at 1000 Hz. IPD (in radians) was converted to ITD (in microseconds). (**B**) Mean and std ITD over time, across frequency and azimuth. Symbols represent the median values across subjects (N = 51), which were fit by a spline curve, color coded for each frequency. The derivative of the curves in the left was used to calculate ITD rate of change (ITDrc) across azimuth. The ITD variability (ITDv) was extracted directly from the distribution of ITD over time, computing its circular standard deviation. (**C**) ITDrc was computed from data of each subject. Azimuth was converted to ITD to obtain an estimate of ITDrc as a function of stimulus frequency and ITD. Left panel illustrates ITDrc (median across subjects) as a function of ITD for each azimuth and frequency. As previously reported (Mills et al. 1972; Gelfand 2016), ITDrc is higher mostly in the front, although the pattern is complex. Right panel depicts the interquartile intervals of ITDrc across subjects, which display low inter-individual variability. Since this figure represents variability of ITD statistics across subjects, a distinct color map was used. (**D**) Standard deviation of ITD over stimulus duration (ITDv), as shown in B (right). Right panels represent ITDv (median across subjects) as a function of ITD for each azimuth and frequency (color code highlights intervals of high discriminability in red). ITDv is lower in the front and higher in the periphery across frequency, with a more pronounced pattern in low frequencies. ITDv shows low correlation with ITDrc (coefficient shown in red), indicating that these statistics are independent. Middle panel shows the inverse ITDv (1/ITDv), since ITDv is inversely proportional to statistical discriminability. Right panel depicts the interquartile intervals of 1/ITDv across subjects; which also display low inter-individual variability (except for the frontal ITDs in frequency close to 250 Hz). (**E**) ITD statistical discriminability index (D-stat) was computed as the ITDrc/ITDv ratio. Left: D-stat (median across subjects) is shown as a function of ITD corresponding to each azimuth and frequency. As expected, D-stat shows strong correlation with ITDrc and ITDv (coefficients shown in red and green, respectively). Interquartile interval across subjects (right) shows low variability of D-stat across subjects (again, except for the frontal ITDs in lowest frequencies). Dark curved lines indicate the π-limit, beyond which ITD information becomes ambiguous within narrow frequency bands.

ITD statistics estimated from human HRTFs and properties of cochlear filters

To test whether ITD statistics are built into brain processing, we estimated ITDrc across locations and ITDv over time of sounds reaching the ears of humans subjects. **Figure 1A** shows the steps for estimating the ITD mean and standard deviation from a single location and frequency for one subject: (1) Impulse responses were obtained from publicly available human HRTF databases (Listen HRTF database, http://recherche.ircam.fr/equipes/salles/listen; 51 subjects). (2) Convolving acoustic signals by impulse responses, results in modulation of ongoing phase and gain that depends on both frequency and sound direction. (3) After filtering the sounds using parameters analogous to human cochlear filters (Glasberg and Moore 1990), we estimated the effect of the phase and gain modulations on the mean and standard deviation of instantaneous ITD (4). Two specific ITD statistics, ITDrc and ITDv, were derived from these measures (**Figure 1B**). ITDrc was higher (**Figure 1C**) while ITDv (**Figure 1D**) was lower in the midline for most frequencies. The inverse of ITDv (1/ITDv) was calculated in order to obtain a quantity directly proportional to statistical discrimination of azimuthal positions based on signal detection theory (**Figure 1D**). Finally, an index capturing both ITDrc and ITDv, referred to as D-stat, was computed by dividing ITDrc by ITDv at each location (**Figure 1E**). Although D-stat is inspired and conceptually similar to the signal detection theory's d-prime (difference in means divided by the square root of the mean variance), d-prime and D-stat are not identical quantities.

ITD statistics were computed for each of the 51 subjects and then the median was calculated. These statistics displayed low variability across individuals (**Figure 1C-E**, right panels) indicating they constitute a set of sensory statistics largely invariant across human subjects. This premise was the basis for considering the particular ITD statistics examined as 'predictable' across different groups of individuals, motivating the question whether they are anticipated in auditory spatial perception.

ITDrc is determined by the shape and filtering properties of the head, including diffractions (Woodworth, 1938; Aaronson and Hartmann, 2014; Roth et al. 1980). These features are supposed to affect ITDv as well; however, ITDv also depends on phase and gain modulations induced by interaction between neighboring

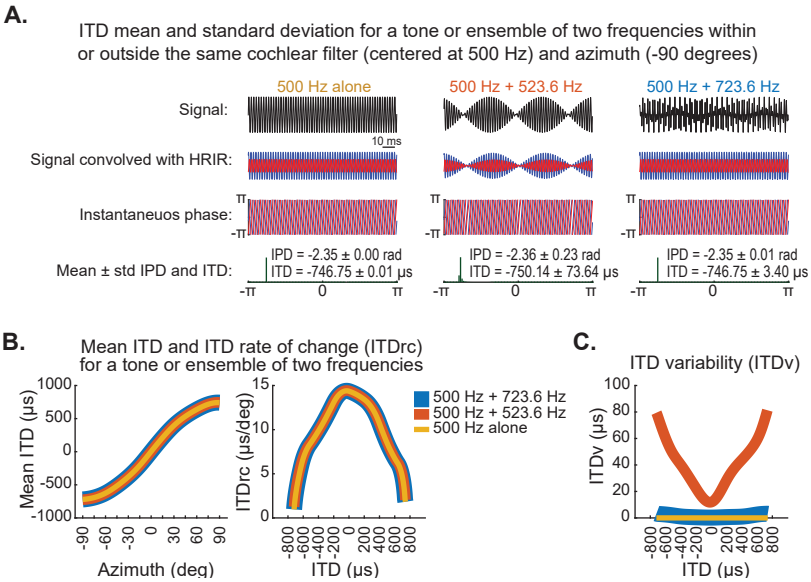

Figure 1—figure supplement 1. Bases of ITD statistics. (**A**) Mean and standard deviation (std) of IPD and ITD for three different signals coming from the same azimuth position (−90 deg): a 500 Hz tone (left) and two signals combining 500 Hz and 523 Hz (middle), and 500 Hz and 723 Hz (right). 523 Hz and 723 Hz were chosen to demonstrate the effect of combining signals lying, respectively, within and outside an ideal cochlear filter with a center frequency of 500 Hz. Mean ITD and std were estimated using the same method described in Figure 1, by convolving the signals (black) with impulse responses from the left and right ears, then filtering the signals by a model cochlear filter with 500 Hz centered frequency (blue and red). Note that the mean IPD and ITD (bottom histograms) are virtually identical across the three conditions, while std is considerably 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, leading to results shown in panel B. (**B**) Mean ITD (left) and ITDrc (right) across azimuths for locations from −90° to 90° in azimuth, averaged across subjects, for the same three signals shown in panel A. Note that mean ITD and ITDrc are similar across three signals, indicating no significant effect of frequency interference on these statistics. (**C**) ITDv (std) averaged across subjects. Note that ITDv is near zero for 500 Hz alone (orange) and 500 + 723 Hz (blue), but increases from front to periphery when neighboring frequencies are combined (500 + 523 Hz; red), consistently with the ITDv shown in Figure 1B. Hence, unlike mean ITD and ITDrc, ITDv depends on interference between neighboring frequencies within cochlear filter ranges.

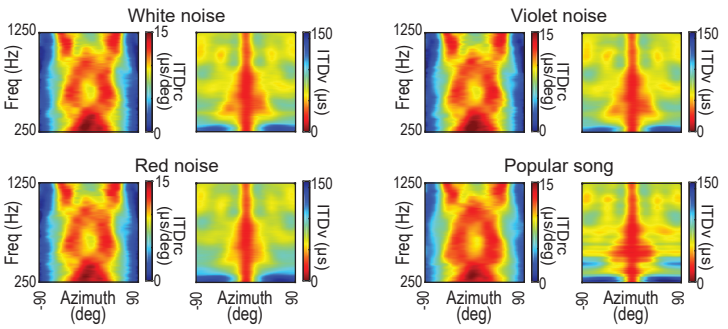

Figure 1—figure supplement 2. ITD statistics across types of broadband signals. ITDrc (left) and ITDv (right) computed for different types of broadband signals. From top to bottom: white noise, red noise, violet noise and a popular song (first second of the chorus of Alive, by Pearl Jam).

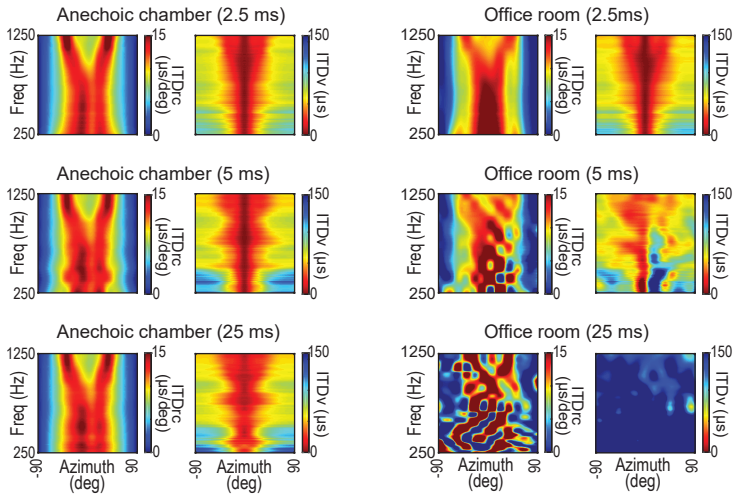

Figure 1—figure supplement 3. Dummy-head ITD statistics across acoustic environments and time-windows of analysis. ITDrc (left) and ITDv (right) computed from impulse responses recorded in human-head shaped manikin inside different types of environment and stimulus durations. Anechoic chambers and reverberant 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,

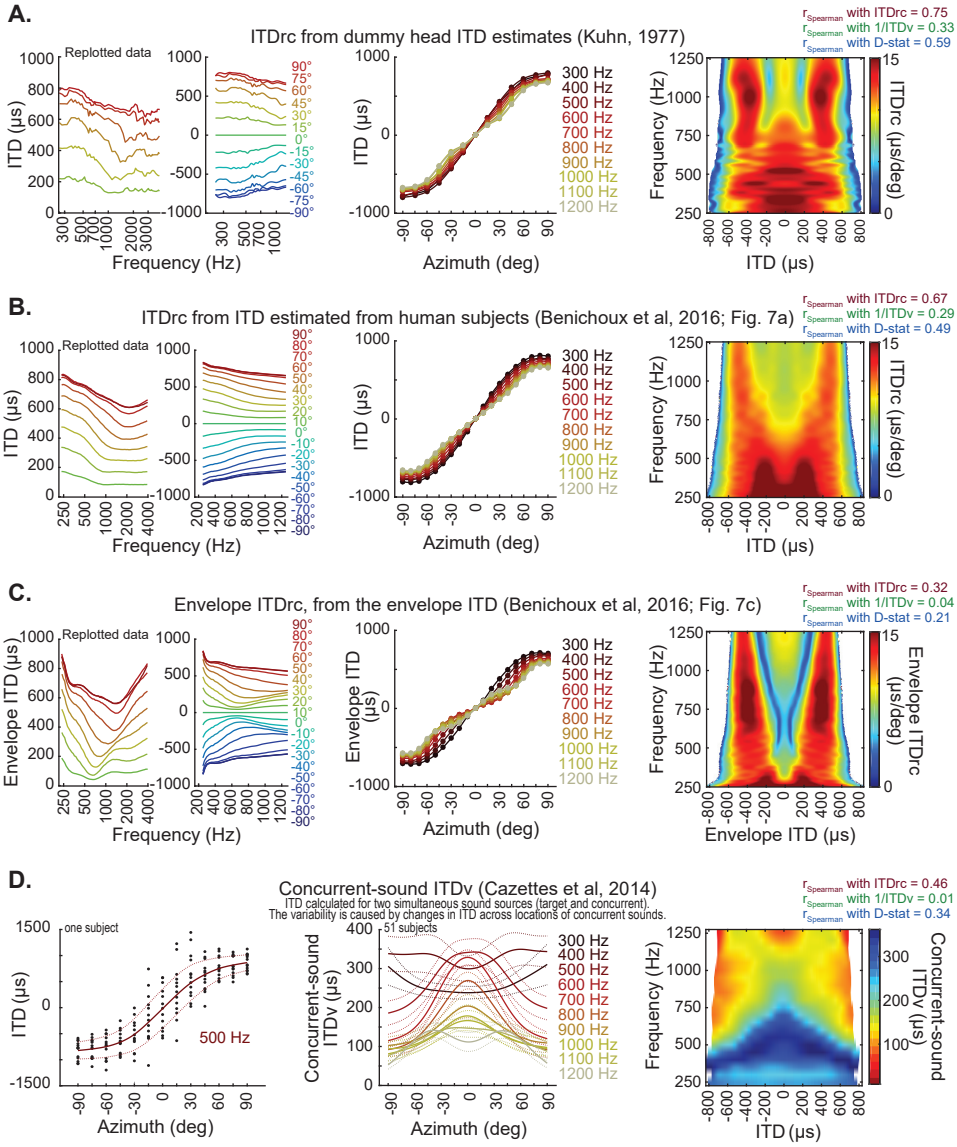

Figure 1—figure supplement 4. Previous reports of ITD statistics. (**A**) ITDrc data computed across frequency and azimuth from pioneer recordings made on human manikins (Kuhn, 1977). Two plots on the left: replot of the original figure (Kuhn, 1977, Fig. 8) showing ITD as a function of frequency for a range of angles of incidence (color code on the right) (left most plot) and the frequency range of interest for the present study (250-1250 Hz), adding their assumed ITD zero in the front and mirrored data representing the contralateral hemifield (second plot from left). Two plots on the right: Mean ITD across azimuth for each frequency, fit by a spline curve (third plot from the left), and ITDrc across azimuth (right most plot). The ITDrc estimated from manikin correlated strongly with our-study's ITDrc (red), weakly with ITDv (green) and moderately with D-stat (blue), indicating consistency of ITDrc estimations across datasets, independence between ITDrc and ITDv statistics and expected relationship between ITDrc and D-stat. (**B-C**) Human ITD across frequency and azimuth described in Benichoux et al. (2016). The panels depict ITD measurements estimated from interaural phase difference, i.e., the fine-structure (panel B) and envelope of the signals (panel C). (**D**) Estimation of ITDv induced by combining target and concurrent sounds, as described in Cazettes et al. (2014). For each target location (from an azimuth range of −90 to 90 degrees sampled in 15-deg intervals and zero elevation), target signals were combined with concurrent sounds coming from each of the 13 directions. The black dots in the left panel show the ITD distribution over concurrent-sound locations, for a 500 Hz target sound at azimuth locations from −90 to 90 degrees (x-axis) for a single example subject. Middle panel shows mean ITDv and standard deviation (dotted lines) over subjects, as a function of target azimuth and frequency. Right panel shows concurrent-sound ITDv estimates as a function of target sound ITD. Concurrent ITDv shows absent to moderate correlation with the ITD statistics estimated in our study (text in red, green and blue).

frequencies, within the critical band of single cochlear filters (**Figure 1—figure supplement 1**). Consistently, the correlation between ITDrc and ITDv was not strong ($r_{Spearman}$=−0.41), suggesting statistical independence. These ITD statistics were consistent across signals with different frequency spectra (**Figure 1—figure supplement 2**). Additionally, we tested consistency of ITD statistics across environments, comparing statistics estimated 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 (**Figure 1—figure supplement 3**). However, at latencies less than 5 ms, where the precedence effect is expected to segregate leading signals from their lagging echoes (Wallach et al. 1949; Brown et al. 2015), ITD statistics estimated in anechoic conditions and reverberant environments were similar, indicating that they are preserved in transient portions of sounds in reverberant environments. Finally, the estimated ITD variability over time was equivalent to the variability over trials ($r_{Spearman}$=0.99; see Methods), indicating that ITDv exhibits ergodicity.

Different auditory cue statistics were examined in previous studies, as well as the question whether consistent statistics across environments shape auditory spatial perception. High ITDrc in the midline supports better sound discrimination in frontal locations (Mills et al. 1972; Gelfand 2016). Consistent with previous reports (Woodworth 1938; Feddersen et al. 1957; Gelfand 2016), we found that ITDrc was higher in the midline for most frequencies (**Figure 1C**). However, ITDrc varied across frequency, including cases where the steepest change occurred at locations distant from the midline. Although this pattern was not directly address by previous studies (Kuhn 1977; Benichoux et al. 2016), it was already observable in their reported data (**Figure 1—figure supplement 4A-C**), indicating that variability of ITDrc

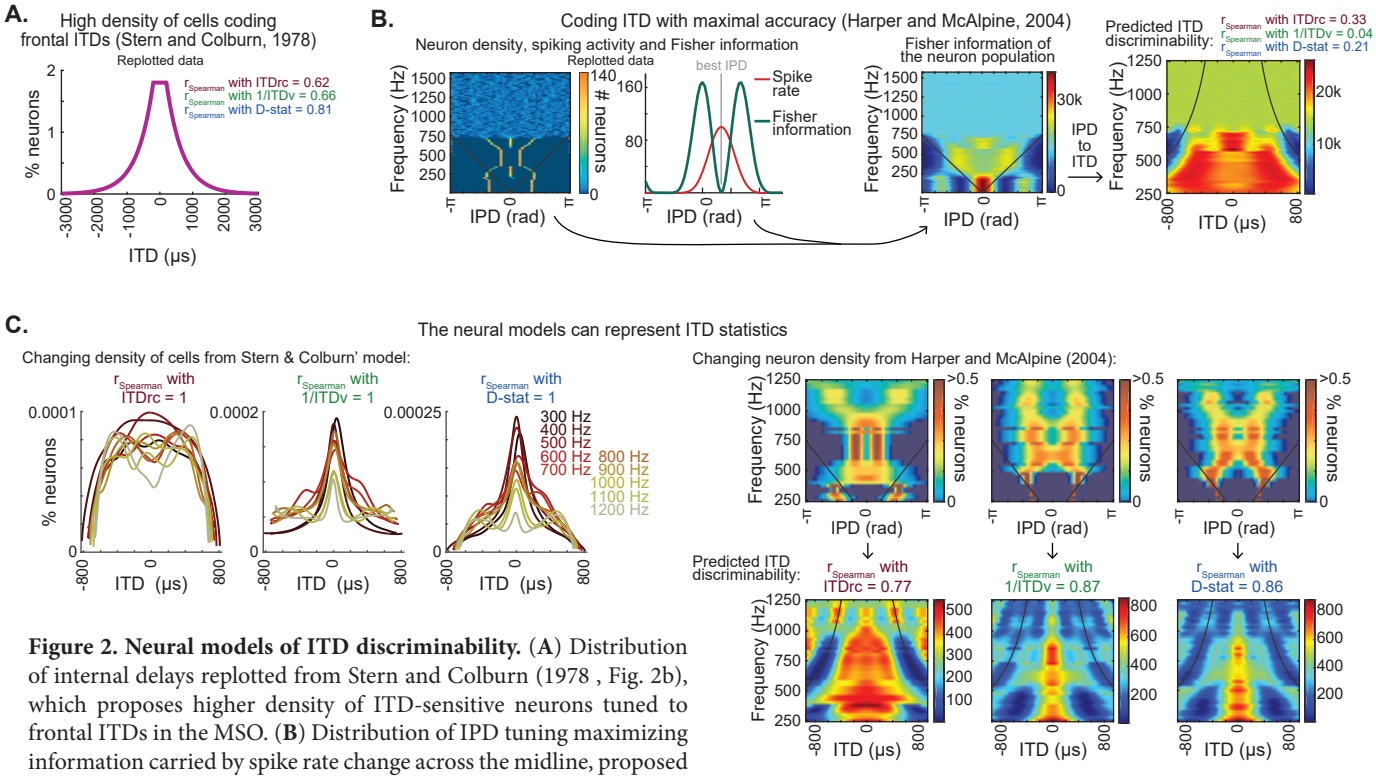

Figure 2. Neural models of ITD discriminability. (**A**) Distribution of internal delays replotted from Stern and Colburn (1978 , Fig. 2b), which proposes higher density of ITD-sensitive neurons tuned to frontal ITDs in the MSO. (**B**) Distribution of IPD tuning maximizing information carried by spike rate change across the midline, proposed by Harper and McAlpine (2004). Distribution of expected best IPDs of ITD-sensitive neurons in the mammal brain across frequency (left-most panel) and spike rate and Fisher information as a function of best IPD (second panel from the left), replotted from data reported in Harper and McAlpine (2004, Fig. 2). Dark straight lines indicate physiological range determined by the distance between ears. The predicted IPD discriminability was calculated by convolving the IPD tuning distribution with the Fisher information (third panel from the left), and converting IPD to ITD for each frequency (right-most panel). Dark curved lines indicate the π-limit, beyond which ITD information becomes ambiguous within narrow frequency bands. (**C**) The coding bases proposed in Stern and Colburn (1978) and Harper and McAlpine (2004) models could effectively achieve the representation of ITD statistics. Plots on the left: the density of cells proposed by Stern and Colburn (1978) as the mechanism underlying ITD discriminability was adjusted to match the pattern of ITDrc (left-most panel), the inverse of ITDv (second panel from the left) and D-stat (third panel from the left). Note that neuron distributions matching ITDv and D-stat largely preserve the shape of the distribution of the original model (shown in panel A), further supporting the explanatory power of this model across experimental datasets. Plots on the right: Top: the IPD tuning distribution proposed by Harper and McAlpine (2004) as a mechanism underlying ITD discriminability was adjusted to match ITDrc, ITDv and D-stat. Note that neuron distributions matching ITD statistics display best IPDs away from midline across frequency, consistent with a coding strategy based on two clustered subpopulations tuned to IPDs away from the front (McAlpine et. al 2001; Harper and McAlpine 2004; Hancock and Delgutte 2004; Pecka et al. 2008). Bottom, ITD discriminability estimated by convolving each of the distributions shown on top with the Fisher information. These estimates of ITD discriminability are highly correlated with the corresponding ITD statistics ($r_{Spearman}$≥0.77, shown on top of each plot).

over sound frequencies is consistent across approaches. ITDv induced by concurrent sounds was proposed by Cazettes et al. (2014) as relevant to the owl's auditory spatial perception. This method applied to human HRTFs led to ITDv values weakly correlated to those obtained with the method used in the present study (**Figure 1—figure supplement 4D**), suggesting that the presence of concurrent sounds represents an independent source of ITDv. Młynarski and Jost (2014) also estimated auditory cue statistics across environments. However they did so without reporting the location of sound sources, which restricts the use of their estimated statistics for testing the prediction of sound discrimination across locations. In the current study, we hypothesize that ITDv induced by frequency interference within cochlear filters is a significant high-order statistic of the crucial binaural cue ITD, and test ITDv's predictive power (alone and combined with ITDrc) of sound discrimination performance and novelty detection.

Classic models of ITD coding postulate brainstem mechanisms underlying ITD discriminability. The model by Stern and Colburn (1978) postulates increased density of delay lines leading to more cells tuned to ITDs near the midline as the basis for increased ITD discriminability in the front (**Figure 2A**). This distribution showed high correlation with all ITD statistics examined, especially the D-stat (**Figure 2A**). On the other hand, a more recent model relies on the maximization of Fisher information within the physiological range of ITD to propose the optimal IPD-tuning distribution in a population of brainstem neurons (Harper and McAlpine 2004). From their proposed distribution and the Fisher information of individual tuning curves, the population Fisher information across frequency and ITD can be estimated. The population Fisher information may predict differential discriminability over locations and frequencies (**Figure 2B**).

We conducted an analysis to test whether the parameters of these previous models could implement a representation of ITD statistics (**Figure 2C**). Interestingly, representing ITDv and D-stat required only a subtle change in the distribution proposed by Stern and Colburn (1978) (**Figure 2C**, plots on the left), supporting the idea that this seminal model, which explains multiple experimental findings, is also consistent with a representation of ITD statistics.

We also found that ITD statistics could be represented by a model relying on the Fisher information framework (**Figure 2C**, plots on the right), proposed by Harper and McAlpine (2004). Although the IPD tuning distribution that correlates best with ITD statistics (**Figure 2C**, plots on the top-right) differs from the distribution that maximizes Fisher information across ITDs (as proposed by Harper and McAlpine, 2004; **Figure 2B**, plot on the left), representing ITD statistics under this framework predicted clustered peripheral IPD tuning across the frequency range tested, consistent with previous reports of brainstem recordings in mammals (McAlpine et al. 2001; Hancock and Delgutte 2004; Pecka et al. 2008). These results suggest that mechanisms believed to underlie IPD discriminability can also implement a representation of ITD statistics, thus providing a plausible biological connection between the coding of sensory statistics and perceptual functions.

Below, we examine the correlation between ITD statistics (estimated in the present study and previously reported) and human auditory spatial discrimination performance (spatial discrimination thresholds and novelty detection responses) or ITD discriminability predicted by neural model. This analysis tests the hypothesis that ITD statistics drive auditory spatial perception.

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 3**). In addition, an ITD-change discrimination test was

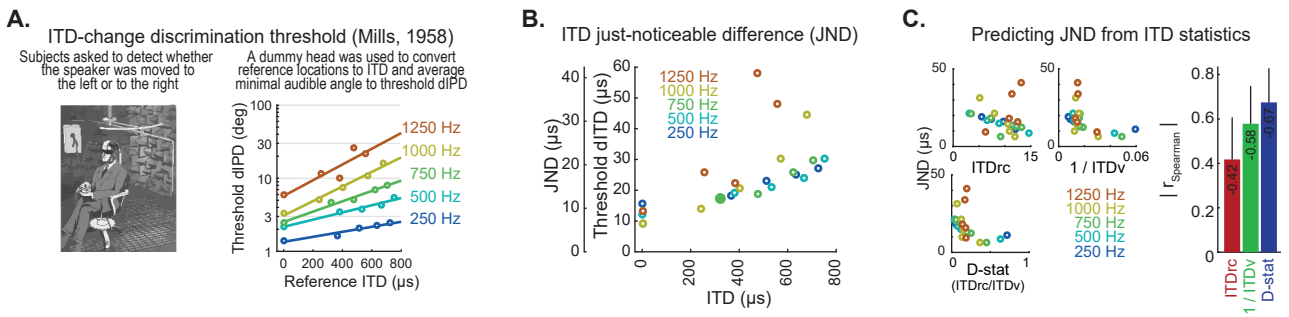

Figure 3. ITD statistics predict human ITD-change detection thresholds previously reported. (**A**) Mills (1958) estimated the minimum ITD change (threshold dITD) inducing just-noticeable differences (JND) in perceived azimuth across frequency and locations. The average minimum audible angle across 3 subjects and speaker azimuth were converted to threshold dIPD and reference ITD using a dummy head with binaural microphones. Data 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 and Rakerd 1989). JND is proportional to the inverse of discriminability. (**C**) Scatter plots (left) show JND as a function of ITDrc, 1/ITDv and D-stat, i.e., ITDrc/ITDv. The color of the symbols depicts the frequency. Bar plot (right) displays the Spearman correlation coefficients calculated for plots shown on the left. JND as a function of D-stat displays the highest correlation coefficient. Error bars indicate standard error of the mean across conditions (25 combinations of frequency and reference locations averaged across subjects).

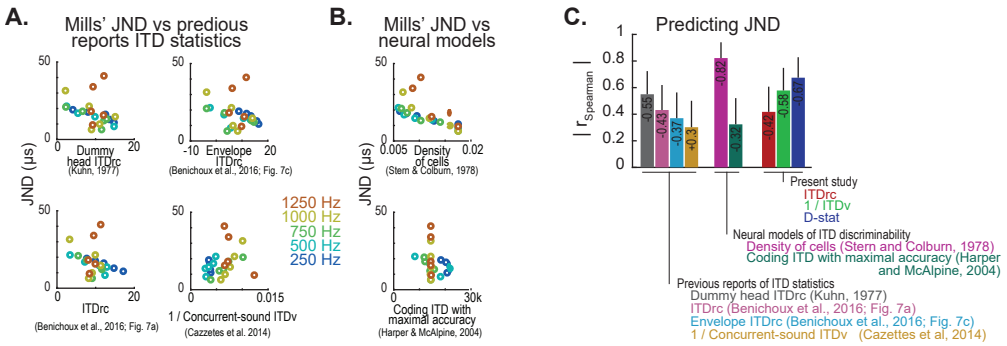

Figure 3—figure supplement 1. Predicting JND from previous reports of ITD statistics and neural models of ITD discriminability. (**A**) Scatter plots of JND reported in Mills (1958) as a function of ITD statistics from three previous reports (Kuhn 1977; Benichoux et al. 2016; Cazettes et al. 2014). (**B**) Scatter plots of JND reported in Mills (1958) as a function of ITD discriminability predicted by two classic neural models (Stern and Colburn 1978; Harper and McAlpine 2004). **(C)** Bar plot displays the Spearman correlation coefficients calculated for plots shown in previous panels and also for ITD statistics from the present study (Figure 3C). ITD discriminability by Stern and Colburn (1978) and present study's D-stat show the highest correlation coefficients. Error bars indicate standard error of the mean across conditions (25 combinations of frequency and reference locations averaged across subjects).

specifically designed and conducted for the purpose of this study (**Figure 4**).

We first tested whether ITDrc, ITDv or their ratio (D-stat) predicted the JND estimated from the previously reported dataset (Mills 1958). **Figure 3A** replots original figures from Mills (1958) illustrating the study's method and detection thresholds of changes in interaural phase difference (dIPD thresholds) across frequency in human subjects. These dIPD thresholds were converted to dITD, then scaled into an estimate of JND as proposed by Hartmann and Rakerd (1989) (**Figure 3B**). When plotted as a function of ITDrc, 1/ITDv and D-stat (**Figure 3C, left**), the highest correlation was observed between JND and D-stat (**Figure 3C, right**). Thus, while discriminability reported in Mills (1958) is significantly correlated with ITDv and ITDrc, combining both ITDrc and ITDv in D-stat yielded the highest correlation with JND estimates based on Mills (1958) data. These results are consistent with the hypothesis that ITD statistics are built into the brain processing underlying discrimination thresholds.

Additionally, in **Figure 3—figure supplement 1** we compared the correlation between JND values reported by Mills (1958) and: (1) ITD statistics reported by other studies (Kuhn 1977; Benichoux et al. 2016; Cazettes et al 2014); and (2) ITD discriminability reported by classic neural models (Stern and Colburn 1978; Harper and McAlpine 2004). D-stat displayed higher correlation with JND than ITD statistics from other studies. The neural model of ITD discriminability by Stern and Colburn (1978) was also highly correlated with JND. Notably, the cell density parameter proposed by Stern and Colburn (1978) to underlie ITD discriminability shows strong similarity to the D-stat pattern across ITD (**Figure 2A**).

Because the previously reported dataset used free field stimulation (Mills 1958), JND data reported in this study may have been influenced by the subjects' use of additional sound localization cues and not only ITD. Furthermore, because free-field stimulation inherently incorporates the effect of filtering properties of stimuli reaching the eardrums, this dataset may not unequivocally distinguish anticipated or actual ITD statistics determining the observed JND. In order to specifically measure ITD-change discrimination using earphones, and be able to evaluate the models' predictive power, a test of ITD-change discriminability was specifically defined. Briefly, subjects underwent a discrimination test consisting of detecting a change in ITD within a pair of tonal sounds delivered through earphones (**Figure 4A**). This test was conducted on 24 normal-hearing adult subjects (see sample description in Methods), providing JND of ITD changes across frequency, for the range of interest (**Figure 4B**). The dITD thresholds averaged across subjects were strongly correlated with thresholds reported by Mills (1958) (**Figure 4C**), although considerably higher. Higher dITD thresholds of the present study compared to Mills (1958) can be explained by the following approach differences: (1) recruitment of naive subjects in the present study, as opposed to highly trained individuals in Mills (1958); and (2) delivery of sounds through earphones, as opposed to free field in Mills (1958). Average dITD thresholds across subjects plotted as a function of ITD statistics (**Figure 4D**) showed the highest correlation with ITDrc and D-stat (**Figure 4E, left**). The AIC test, which accounts for individual variability, further revealed that D-stat was the best predictor of dITD thresholds (**Figure 4E, right**). Particularly, dITD thresholds for 250 Hz were a consistent outlier across subjects. Higher thresholds for this frequency have previously been reported (Brughera et al. 2013). Moreover, exclusion of this frequency further improved D-stat's predictive power (**Figure 4—figure supplement 1**), suggesting that dITD thresholds at 250 Hz may be determined by additional parameters not addressed by our model.

D-stat predicts dITD thresholds better than the ITD statistics from other studies (Kuhn 1977; Benichoux et al. 2016; Cazettes et al 2014) and better than classical models of ITD discriminability (Stern and Colburn 1978; Harper and McAlpine 2004) (**Figure 4—figure supplement 2**). These results further support the notion that ITD statistics are built into ITD discrimination thresholds. To further test the hypothesis that ITDv is relevant to 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 mismatch negativity (MMN) EEG responses (Näätänen et al. 1978) for

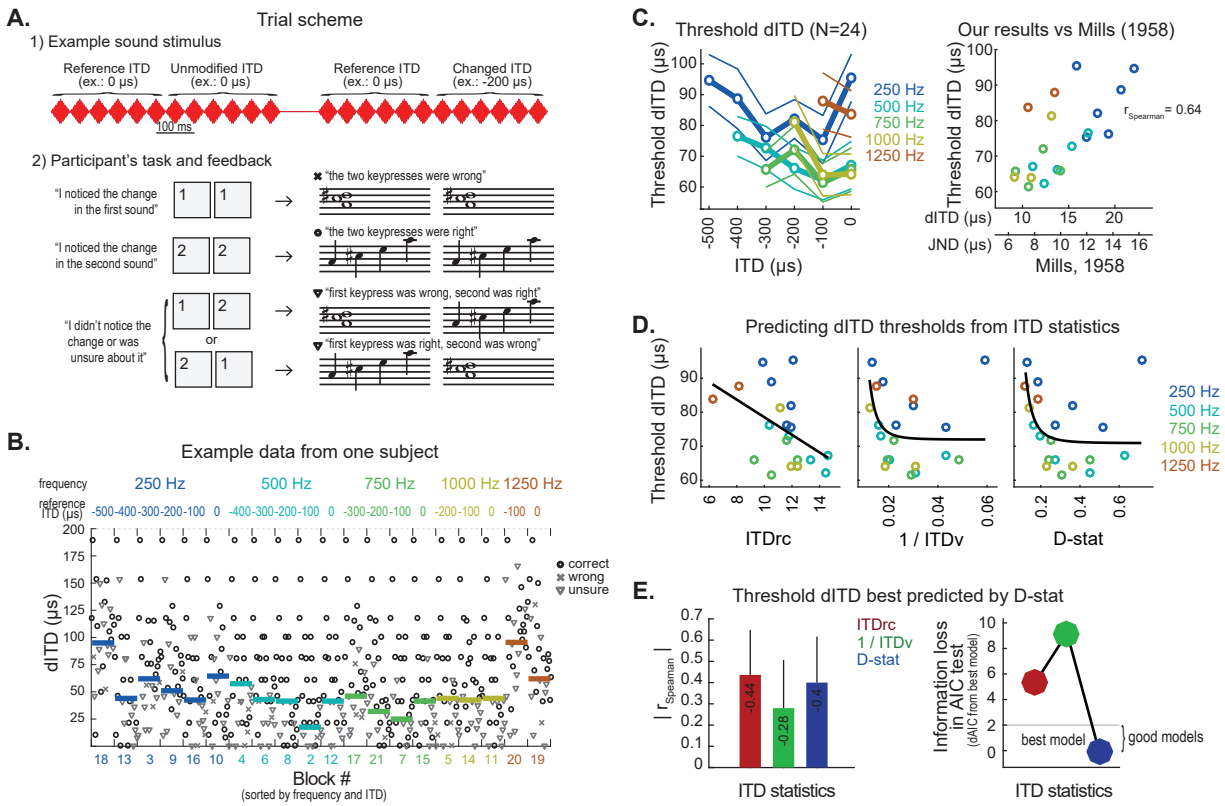

Figure 4. ITD statistics predict human ITD-change detection thresholds measured with headphones. (**A**) Scheme of one trial of the test conducted by the present study to specifically assess ITD-discrimination (dITD) thresholds for tonal sounds delivered through headphones. Subjects press the spacebar to start the trial. Two binaural tones with an identical 'reference' ITD were presented in sequence through headphones. ITD was changed in the last half of one of these sounds. Subjects (N=24) were asked to respond which of these sounds contained a change in ITD by pressing the '1' or '2' keys twice when sure, or alternate both keys when unable to detect a change or unsure. Feedback sounds were delivered indicating whether each response was correct or wrong. (**B**) Example dITD thresholds estimated across frequency and reference ITDs for one subject. The dITD thresholds were estimated from responses in 25 trials within a range of dITDs between the reference and changed ITDs: dITD threshold was estimated as the dITD that optimally separated correct from wrong and unsure categories of responses (black and gray symbols, respectively). (**C**) Mean (bold lines) and standard error of the mean (thinner lines) of dITDs over subjects, across frequencies and ITDs (left). Threshold dITDs tended to be lower in the front, but this trend varied across frequencies. dITD thresholds across frequency estimated from this test were larger, yet correlated with estimates from Mills (1958) (right; $r_{Spearman}$=0.64). (**D**) Threshold dITD (averaged across subjects) as a function of ITDrc (left), 1/ITDv (middle) and D-stat (right). Black lines indicate the function which best fit the data (linear or power function). (**E**) Correlation coefficients (left) and AIC test (right) for selecting the model with the best predictive power of threshold dITD. The group average discriminability displays moderate correlation with ITDrc, 1/ITDv and D-stat (left). Error bars indicate standard error of the mean across conditions (25 combinations of frequency and reference locations averaged across subjects). The AIC test, which considers the data of each subject individually, shows that D-stat carries the best predictive power.

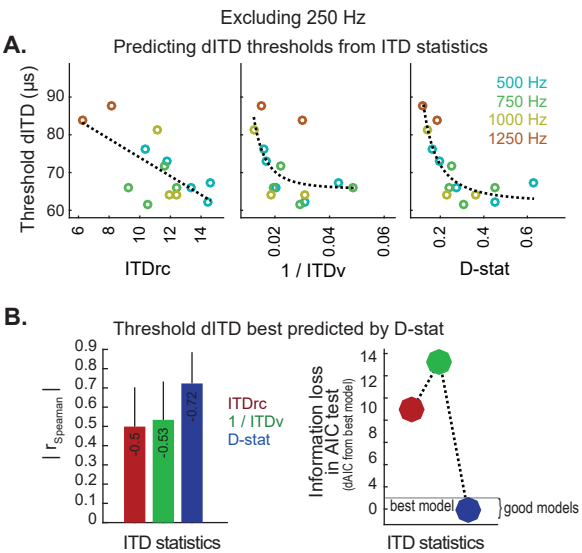

Figure 4—figure supplement 1. Prediction of ITD-change detection thresholds, excluding the condition of 250 Hz. (**A**) Threshold dITD as a function of ITDrc (left), 1/ITDv (middle) and D-stat (right) for 500 Hz, 750 Hz, 1000 Hz and 1250 Hz sounds (250 Hz are excluded). Dotted lines indicate the function which best fit the data (linear or power function). (**B**) Correlation coefficients (left) and AIC test (right) for selecting the model with the best predictive power of threshold dITD; D-stat displays the best predictive power in both metrics. Error bars indicate standard error of the mean across conditions (25 combinations of frequency and reference locations averaged across subjects).

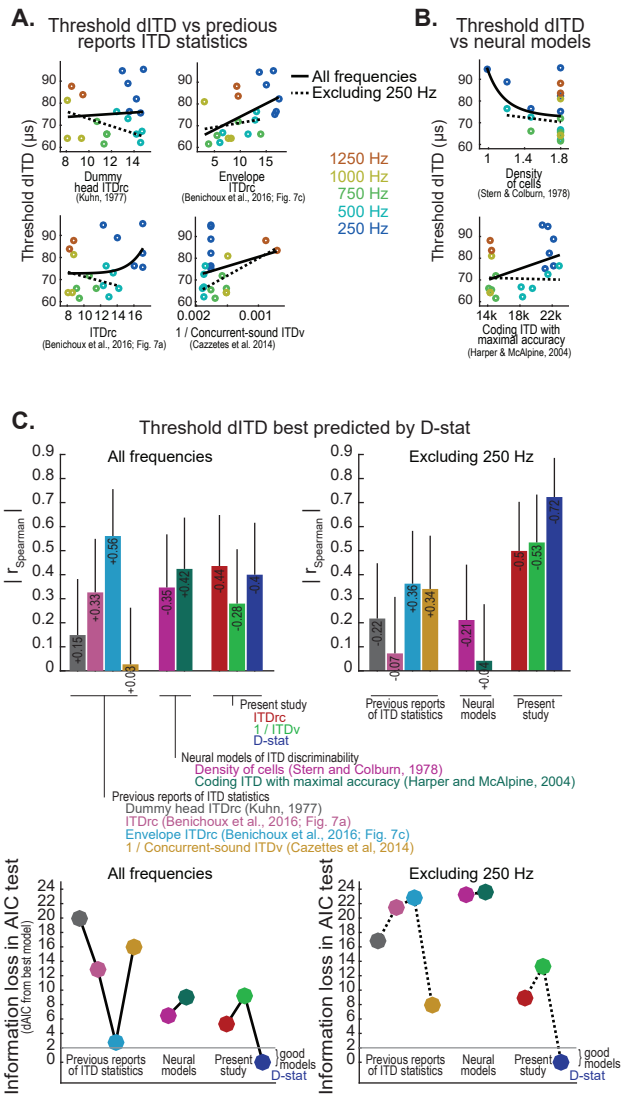

Figure 4—figure supplement 2. Predicting ITD-change detection thresholds from previous reports of ITD statistics and neural models of ITD discriminability. (**A-B**) Scatter plots of threshold dITD as a function of ITD statistics from three previous reports (Kuhn 1977; Benichoux et al. 2016; Cazzetes et al. 2014; panel A) and as a function of ITD discriminability predicted by two classic neural models (Stern and Colburn 1978; Harper and McAlpine 2004; panel B). Solid and dotted lines represent regressions including all frequencies tested and excluding 250 Hz, respectively. **(C)** Top plots: Bar plots display the Spearman correlation coefficients calculated for plots shown in previous panels and also for ITD statistics from the present study. ITD discriminability by Stern and Colburn (1978) and present study's D-stat show the highest correlation coefficients. When all the frequencies are analyzed (left plot), the Envelope ITD (light blue bar) shows the highest correlation coefficient; however, this correlation is positive, indicating a relationship between envelope ITDrc and threshold dITD opposite to what is expected from the relationship between ITDrc and threshold. When the 250 Hz frequency is excluded from the analysis (right plot), D-stat from the present study yields the highest correlation with threshold dITD. Error bars indicate standard error of the mean across conditions of frequency and reference locations averaged across subjects (the number of conditions is 20 or 14 – all frequencies and excluding 250 Hz, respectively). Bottom plots: AIC test for selecting the best model, showing that D-stat carries the best predictive power of dITD thresholds compared to any previous report of ITD statistics or neural models of ITD discriminability, both when all the frequencies are analyzed and when the 250 Hz frequency is excluded from the analysis.

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, MMN could be used to address the question of whether anticipated ITD statistics would modulate its amplitude.

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 5A**). The MMN signal is obtained by subtracting event-related potentials (ERPs) evoked by standard stimuli from ERPs evoked by deviant stimuli. The larger the tone feature 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 presented tones through insert earphones and recorded MMN responses to spatial deviants. By presenting tones directly to the ear canal through insert earphones, ITDv induced by frequency interference within the critical band of cochlear filters and gain modulation by the filtering effect of outer ear and surrounding structures was avoided. Thus, differences in novelty detection could be assigned to expected, rather than to current stimulus statistics. Importantly, the pre-attentive nature of MMN signals provides a direct brain measure of discriminability that did not require training subjects to perform behavioral tasks, which might introduce confounding effects.

We measured MMN responses in 33 normal-hearing adult participants. The data showed the characteristic frontal topography of MMN responses (Giard et al. 1990) (**Figure 5B**). A set of ITD and frequency conditions was selected in order to sample critical ranges drawn from the HRTF analysis (**Figure 5C**). Frequencies of 400, 550, 600 and 650 Hz were chosen, where ITDrc and ITDv changed as a function of azimuth in a manner that could maximize the difference in predictive power across models. Frequencies lower than 400 Hz were not tested because of observed distortion in the sound stimulation system, while frequencies above 650 Hz were excluded to avoid phase ambiguity confounds. MMN signals were measured across participants and conditions. The peak amplitude of MMN was used to quantify the effect of stimulus manipulations on novelty detection.

Consistent with previous reports (Pakarinen et al. 2007), we found that average MMN peak amplitude across

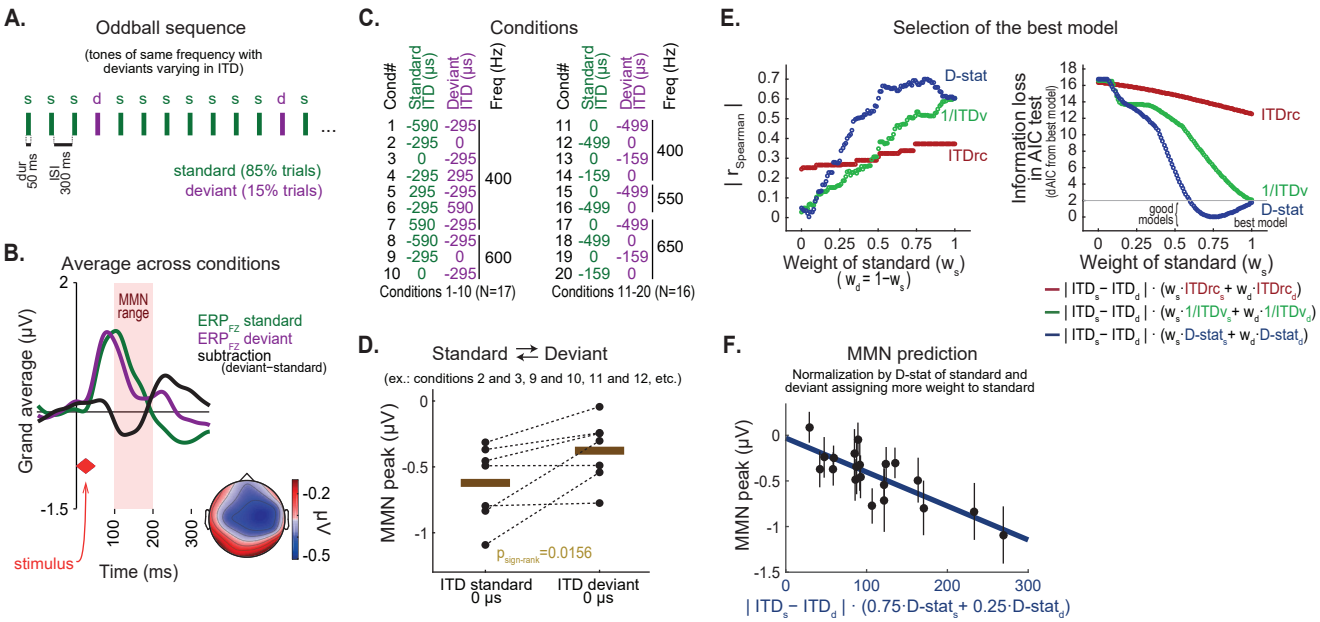

Figure 5. Discriminability of spatial deviants indexed by MMN responses is weighted by expected ITDv at the standard location. (**A**) Passive oddball sequence protocol, in which subjects listened to frequent "standard" stimuli embedded with rare "deviants." (**B**) In each condition, two tones of the same frequency and distinct ITDs were presented. 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 brain areas. (**C**) Two sets of 10 conditions were tested in two different groups of subjects. (**D**) 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). (**E**) Correlation coefficients ($r_{Spearman}$, left) and AIC test (right) for selecting the model with the best predictive power of MMN amplitude. The models normalized ITD difference by anticipated ITDrc and ITDv, weighting standard (w_s) and deviant (w_d) differently. (**F**) Prediction of MMN amplitude by D-stat, weighting standard more than deviant, which is the best of the models according to the AIC test (dAIC=0).

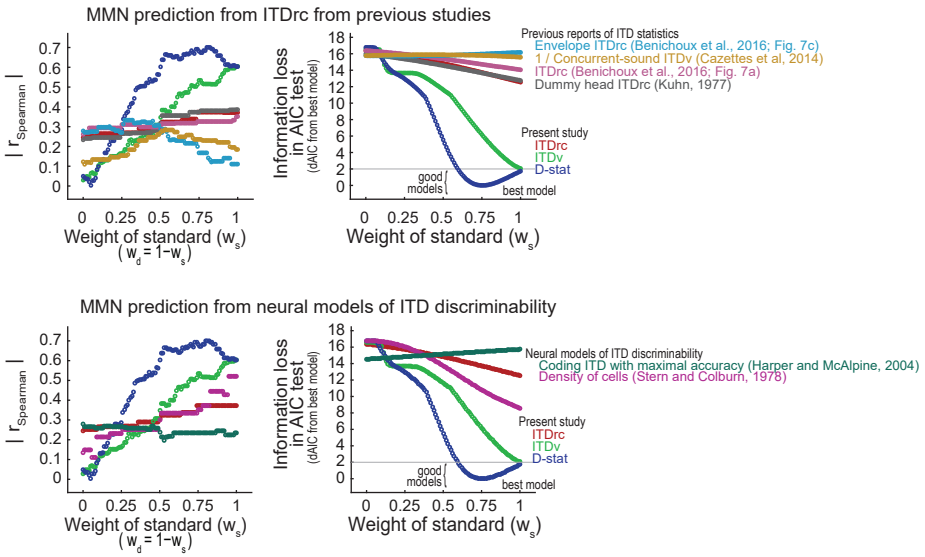

Figure 5—figure supplement 1. Prediction of MMN peak amplitude from previous reports of ITD statistics and neural models of ITD discriminability. Correlation coefficients (left) and AIC test (right) for selecting the model with the best predictive power of MMN amplitude. D-stat (dark blue) shows the best predictive power of MMN amplitude compared to any previous report of ITD statistics (top) or neural models of ITD discriminability (bottom). Color codes indicated on the right.

subjects showed strong correlation with ITD difference between standard and deviant stimuli when the standard stimulus was positioned in the front. Surprisingly, however, the correlation with ITD difference was weaker when the standard was in peripheral locations. Accordingly, standard stimuli located in the front led to larger MMNs than deviants in the front (**Figure 5D**). 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 show 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.

We then examined the predictive power of MMN amplitude by models relying on ITD statistics. Three groups of models were tested, which normalized the ITD difference between standard and deviant by the ITDrc, ITDv or D-stat corresponding to the frequency and location for the standard and deviant stimuli (**Figure 5E**). For each model, the effect of changing the relative weights of standard and deviant was examined. We found that multiplication by 1/ITDv was required for a good

predictive power. Interestingly, models that assigned stronger weight to the ITD statistics of the standard performed better. This was valid to both methods applied: correlation to the average amplitude of MMN peaks or AIC test statistics, which considers individual variability. Specifically, the best prediction of MMN peak was obtained by normalizing the ITD difference between standard and deviant by D-stat, and assigning 75% weight to the standard and 25% to the deviant (**Figure 5F**). Furthermore, D-stat is also a better predictor of MMN peak amplitude (**Figure 5—figure supplement 1**) than ITDrc or ITDv alone, reported in this and other studies (Kuhn 1977; Benichoux et al. 2016; Cazettes et al 2014) or classic models explaining ITD discriminability (Stern and Colburn 1978; Harper and McAlpine 2004). The predictive power of MMN by D-stat further supports that combined ITDrc and ITDv information is critical for auditory spatial processing involving both discriminability and novelty detection.

Discussion

Human HRTFs were used to estimate ITD statistics across subjects. Namely, the rate of change of ITD over azimuth (ITDrc) and standard deviation of ITD over time or trials (ITDv) across frequency and ITD. ITDrc and ITDv and their ratio (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 as variation in density of brainstem ITD-sensitive neurons (Colburn 1977; Stern and 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 and McAlpine 2004). These are valid explanations of biological mechanisms underlying discriminability. Other mechanisms based on the amount of spatial information carried by auditory stimuli have also been invoked, such as the rate of change of ITD as a function of azimuth (Mills et al. 2018; Gelfand 2016). Our study specifically tests whether anticipated ITDrc and ITDv are additional relevant parameters determining ITD discriminability. In fact, our results, rather than contradicting previous findings, 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 and Peña 2017). These studies provided evidence of how sensory reliability could be represented (Fischer and Peña 2011; Rich et al. 2015; Cazettes et al. 2016) and integrated into an adaptive behavioral command (Cazettes et al. 2018). 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), results of the present study support the hypothesis that specific ITD statistics, ITDrc and ITDv, determine human sound localization discriminability and novelty detection of acoustic spatial deviants. We show that classic neural models are not incompatible with the representation of these statistics, although the present work does not provide evidence to support which mechanism underlies ITD and ITD statistics coding in the human brain. The code underlying the causal link between inherent statistics 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 ITDv. Here we show that ITDv improves the prediction of these thresholds. The correlation between discriminability and each of the ITD statistics shows that the ratio between ITDrc and ITDv (D-stat) is the best predictor of discriminability thresholds. Yet some specific patterns of the JND data are not captured by ours or other models, such as the unexpected higher ITDrc or 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 considering ITDrc alone, supporting the idea that the brain anticipates, and therefore represents auditory space discriminability as a function of frequency, location and statistics of spatial cues.

Because sound stimulation in Mills (1958) consisted of tones, the across-frequency interaction was not a source of ITDv. 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 ITDv) from anticipated ITDv on discriminability thresholds. However, the better predictive power of Mills' (1958) data by models including anticipated ITDv suggests that these statistics play a role.

We estimated detection thresholds of ITD changes using headphones, finding a better correlation between these thresholds and the ratio between ITDrc and ITDv (D-stat) than with either ITDrc or ITDv statistics alone, reported in this and previous reports, and predictions by neural models of ITD discriminability (Stern and Colburn 1978; Harper and McAlpine 2004). However, thresholds for tonal stimuli in 250 Hz constituted and outlier that revealed limited prediction by ITD statistics from both present and previous studies. Increased thresholds within the 250 Hz range were reported in previous studies (Brughera et al. 2013). These results thus suggest that although additional factors may determine discriminability in the lowest frequencies, combined ITD statistics are significant predictors of discrimination thresholds in human sound localization.

Novelty detection weighted by anticipated ITD statistics

The novelty detection studies conducted with tonal sounds delivered by insert earphones were designed to further test the question of whether anticipated ITDv is built into spatial perception. Consistent with observations on the discriminability thresholds, the ratio between ITDrc and ITDv (D-stat) was a better predictor of deviant detection responses.

Although ITD-change detection thresholds and the novelty detection paradigm required discriminating a change in ITD, the latter also involved 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 ITDv 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 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 statistics are 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 higher statistical discriminability, 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, in the sense that larger ITD changes are required for detecting shifts in location where anticipated statistical discriminability are lower. 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. The present work benefits from the simplicity of testing ITD in isolation; future work should address whether statistics of other spatial cues are anticipated by the human brain, such as interaural level difference.

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 BandK 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 (**Figure 1**). 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 in 5 Hz steps. These center frequencies are within the range where ITD is a primary spatial binaural cue (Lord Rayleigh, 1907) and also correspond with the frequency range of thresholds estimated by Mills (1958). The bandwidth of these filters matched the estimated bandwidth of human cochlear filters (Glasberg and Moore 1990). We used the gamma-tone 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 ITDrc being corrupted by artificial negative values caused by phase ambiguity. We also selected IPD values within the $-\pi$ to π 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: $ITD=10^6 IPD(2\pi Freq)^{-1}$. These steps are shown in **Figure 1A**.

For each subject, we interpolated the mean ITD across azimuths using the cubic spline method and calculated the rate of change of ITD across azimuth

(ITDrc) as the derivative of this curve. Then, for each combination of azimuth and frequency, we computed the median and interquartile interval ITDrc across subjects. Finally, we converted azimuth to ITD (using the original relationship between azimuth vs ITD), obtaining an estimate of ITDrc across frequency and ITD (**Figure 1C**). We performed a similar procedure for estimation of ITD variability (ITDv). For each subject, we interpolated the standard deviation of ITD over time across azimuths using the cubic spline method. Then, for each combination of azimuth and frequency, we computed the median and interquartile interval ITDv across subjects (**Figure 1D**). To confirm that this metric of ITDv is ergodic, i.e., variability across time and trials is equivalent, we also computed the ITDv over trials, by selecting ITDs at the midpoint of 1-second signals, across 200 different broadband sounds. The combined effect of ITDrc and ITDv was assessed by their ratio. This ratio was referred to as D-stat (**Figure 1E**).

Neural models

Two seminal models (Stern and Colburn 1978; Harper and McAlpine 2004) addressing discriminability of azimuth positions in acoustic space based on ITD were used to examine the potential link between the brain representation of sensory statistics and perceptual functions.

The relative number of fiber pairs encoding interaural delays, $p(\tau)$, used as predictors of ITD discriminability (**Figure 2A**), was extracted from Stern and Colburn (1978).

From Harper and McAlpine (2004), we extracted the optimal population distribution of IPD tuning estimated for humans, and the Fisher information from single-cell IPD-tuning curves. The M-shaped Fisher information curve was positioned at the best IPD of each neuron within the distribution, to obtain an estimate of the population Fisher information across IPD and frequency. The population Fisher information across IPD was scaled to ITD, obtaining the prediction of ITD discriminability induced by the neuron distribution (**Figure 2B**).

To test whether ITD statistics could be represented by parameters of these models, these parameters were adjusted to match statistics (**Figure 2C**). For the Stern and Colburn (1978) model, the analysis consisted in matching the relative number of fiber pairs to each ITD statistic normalized from 0 to 1, then scaling these data to obtain a probability distribution where the sum of all probabilities was equal to 1 (**Figure 2C**, left). For the Harper and McAlpine (2004) model, we first generated two midline-mirrored gaussian neural population distributions with random mean IPD tunings and standard deviations from 0 to π, then selected the distribution that displayed the highest correlation between Fisher information and ITD statistics for each frequency. In order to select the population distribution predicting ITD discriminability that best correlated to ITD statistics across frequency, the distributions for each frequency were scaled (multiplying the original distribution from 0 to 1.2 with steps of 0.1) from other frequencies, substituting the original distribution only when the overall correlation increases.

Estimation of spatial discriminability thresholds from previously published datasets (Mills, 1958)

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 3A**, left). 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 judgments 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 (**Figure 3B**, left).

For the current study, we extracted the data points of threshold IPD across 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 precise (not shown). Hartmann and 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 (an inverse measure of discriminability measure) led to the same correlation coefficients, since this analysis is unaffected by scaling constants (**Figure 3C**).

Estimation of spatial ITD discriminability thresholds

Healthy adult subjects were included in the sample (N=24; 12 females and 12 males; mean age 28.0 ± 8.6; 5 left-handed and 19 right-handed; 19 from São Paulo and 5 from New York). After the procedure was described to the subjects, they provided written informed consent. The protocol was approved by the Ethics Committee of Universidade Federal do ABC and 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 or hearing impairments.

A test was designed to estimate detection thresholds of changes in stimulus ITD (dITD) across specific frequencies of interest for this study. In this computer-

based test, subjects were presented with 65 dB (A scale) tones within a range of frequencies through calibrated headphones using an Instrutherm DEC-460 decibel meter or B&K 4947 microphone with an artificial ear.

Trials started by pressing the spacebar key. In each trial, two binaural tones were presented in sequence. The ITD of both sounds started at the same value (reference ITD) and changed by different amounts (dITD) in the second half of either the first or the second sound in the sequence. Subjects were instructed to press the keys '1' or '2' depending on which sound in the pair they perceived a shift in location, and press a given key twice if confident, or alternate both (1 and 2) keys if unable to perceive a shift or unsure about it. Trials could be repeated as many times as needed by pressing the spacebar key. Feedback sounds indicated whether each of the pressed keys was correct or wrong (**Figure 4A**). The range of reference ITDs spanned from frontal (0 μs) to peripheral locations within the unambiguous range of ITD for each frequency. ITD change (dITD) varied from 1 μs to 200 μs towards the periphery to cover a range from unequivocally detectable and undetectable changes for each frequency. The direction of dITDs relative to the reference ITD was always away from the front, to avoid direction dependent biases affecting threshold measurements. Each condition (a given combination of frequency and reference ITD) was presented 29 times: 4 training trials with dITD 200 μs which were not computed and 25 testing trials which were used for estimating the dITD threshold (**Figure 4B**). The first (training) block of the session used 750 Hz tones and −50 μs reference ITD for all subjects and was not included in the analysis. Blocks 2 to 21 presented tones of frequencies 250, 500, 750, 1000 or 1250 Hz, and reference ITDs from −500 to 0 varied in steps of 100 μs. Only reference IPDs of absolute value smaller than π/2 radians were included, to avoid phase ambiguity. The sequence of conditions was randomly varied across subjects. Sessions lasted approximately 60 minutes.

Ongoing estimation of dITD thresholds was conducted since trial 1 of each block, to optimize the test estimate. A psychometric sigmoid curve was progressively fit to a plot of correct responses (assigned 1) and incorrect or unsure responses (assigned 0) as a function of dITD; the dITD corresponding to 0.5 accuracy was selected as the estimated dITD threshold. Preliminary dITD thresholds were computed from subsets of trials within varying dITD ranges. The first 6 trials ranged from dITDs 10 to 190 μs, spaced by 36 μs. From these trials, a first preliminary dITD threshold was estimated. In the following 6 trials, dITD was varied from −50 to 50 μs in steps of 20 μs, centered on the first preliminary dITD threshold; by the end of trial 12 a second preliminary dITD threshold was estimated using trials 1 to 12 using the same sigmoid fitting procedure. A third preliminary dITD threshold was then estimated using trials 1-18. The final 7 trials ranged from −21 to 21 μs spaced by 7 μs centered on the third preliminary dITD threshold. When a set of dITDs centered on a preliminary dITD threshold extended beyond the 1 to 200 μs range, dITDs were adjusted to fall within this range. While this procedure permitted an efficient estimate of dITD thresholds, it did not yield plausible dITD threshold estimates in cases where subjects provided largely random responses across dITDs. To address this limitation, a nonparametric receiver operating characteristic (ROC) classifier was conducted offline to independently verify the validity of estimated dITD thresholds.

The ROC classifier used for estimating threshold dITDs was computed over the 25 trials of each condition for each subject. The threshold was estimated by averaging the subset of possible thresholds within the 1 to 200 μs range that jointly maximized the number of correct (hit) responses and minimized false positive (type II error) ones. This optimization was obtained by selecting candidate dITD thresholds within a minimum euclidean distance from the perfect discrimination (i.e., 100% hit rate and 0% type II error rate) yielded by the ROC analysis. The ROC classifier was robust enough to estimate consistent dITD thresholds to all conditions from all subjects (**Figure 4C**). The threshold within 1-200 μs range estimated by the sigmoid fitting method were significantly correlated with those estimated by the ROC classifier ($r_{Pearson}$=0.85).

Prediction of spatial discriminability thresholds by ITD statistics

An initial suboptimal analysis of the ranked (Spearman) correlation coefficients was performed for the relationship between the threshold dITD averaged across subjects and the ITD statistics of the reference ITD (**Figure 4D**). Additional analyses with selection of best predictor of MMN peaks were performed with linear mixed-effect models (**Figure 4E**). 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.

Collection and analysis of the Mismatch Negativity (MMN) component

Healthy adult participants 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 Institutional 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 or better at 500, 1000, 2000, and 4000 Hz).

Participants 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 to 53 dB (A

scale) 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.

Participants listened to oddball sequences consisting of repetitive ("standard") tones embedded with sporadic ("deviant", 15% of the trials) tones, with duration of 50 ms and inter-stimulus interval of 300 ms (**Figure 5A**). Grand-averages of ERPs recorded at FZ electrodes were computed for standard and deviant trials across all subjects and conditions; 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 5B**). Each subject was presented with 10 conditions, which differed in frequency and ITD (**Figure 5C**). 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.

Prediction of MMN by ITD statistics

An initial suboptimal analysis of the ranked (Spearman) correlation coefficients was performed for the relationship between the MMN peak averaged across subjects and absolute ITD difference normalized by ITD statistics, from both standard and deviant ITD (**Figure 5E**, left). Additional analyses with selection of best predictor of MMN peaks were performed with linear mixed-effect models using AIC to compare the performance of the models (**Figure 5E**, right).

All data processing was performed in Matlab (Mathworks) using built-in or custom made routines. The datasets generated and analyzed in 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 André Cravo, Boris Marin, Fanny Cazettes, Gervasio Batista, Michael Beckert, Peter Claessens and Ruben Coen Cagli for thoughtful discussions or 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; JP, Acquisition of data, Conception and design, Drafting or revising the article. BF, ES, Conception and design, Drafting or revising the article.

Ethics

This study was performed in accordance with the NIH Human Subjects Policies and Guidance and with the Brazilian National Health Council, and it was approved by the Internal Review Board of the Albert Einstein College of Medicine (#1999-023) and Ethics Committee of Universidade Federal do ABC (#2968291).

References

Aaronson, N. L. and 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. and Brette, R. On the variation of interaural time differences with frequency. *J. Acoust. Soc. Am.* **139,** 1810–1821 (2016).

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

Brughera, A., Dunai, L. and Hartmann, W. M. Human interaural time difference thresholds for sine tones: the high-frequency limit. *J. Acoust. Soc. Am.* **133,** 2839-55 (2013).

Carr, C. E. and 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., and 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. and 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. and Pena, J. L. Spatial cue reliability drives frequency tuning in the barn Owl's midbrain. *eLife* **3,** e04854 (2014).

Changizi, M. A. and 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., and 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. and 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. and Jeffress, L. A. Localization of high-frequency tones. *J. Acoust. Soc. Am.* **29,** 988–991 (1957).

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

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

Goldberg, J. M. and 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. and Swets, J. A. *Signal Detection Theory and Psychophysics*. (1966).

Grothe, B., Pecka, M. and McAlpine, D. Mechanisms of sound localization in mammals. *Physiol. Rev.* **90,** 983–1012 (2010).

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

Hancock and Delgutte, 2004; Hancock, K. E., and Delgutte, B. A physiologically based model of interaural time difference discrimination. *J. Neurosci.* **24,** 7110–7117 (2004).

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

Kayser, H.; Ewert, S. D.; Anemüller, J.; Rohdenburg, T.; Hohmann, V. and Kollmeier B. Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. *EURASIP J. Adv. Signal Process.* **2009,** id298605 (2009).

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

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

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

Middlebrooks, J. C. and 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).

Młynarski and Jost Młynarski, W. and Jost, J., Statistics of natural binaural sounds. *PloS one* **9,** e108968 (2014) .

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

Pakarinen, S., Takegata, R., Rinne, T., Huotilainen, M. and 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).

Pecka, M., Brand, A., Behrend, O., and Grothe, B. Interaural time difference processing in the mammalian medial superior olive: the role of glycinergic inhibition. *J. Neurosci.* **28**, 6914–6925. (2008).

Rayleigh L. XII. On our perception of sound direction. *Philos. Mag.* **13,** 214–232 (1907).

Rich, D., Cazettes, F., Wang, Y., Peña, J. L. and Fischer, B. J. Neural representation of probabilities for Bayesian inference. *J. Comput. Neurosci.* **38,** 315–323 (2015).

Roth. G.L., Kochhar, R.K., Hind, J.E. Interaural time differences: implications regarding the neurophysiology of sound localization. *J. Acoust. Soc. Am.* **68,** 1643-1651 (1980).

Sams, M., Paavilainen, P., Alho, K. and Näätänen, R. Auditory frequency discrimination and event-related potentials. *Electroencephalogr. Clin. Neurophysiol.* **62,** 437–448 (1985).

Schnupp, J. W., and Carr, C. E. On hearing with more than one ear: lessons from evolution. *Nature Neurosci*, ***12***, 692-697. (2009)

Stern, R. M. and 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. and 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. and 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).

Wallach, H., Newman, E.B. and Rosenzweig, M.R. The precedence effect in sound localization. *Am. J. Psychol.* **62,** 315-336 (1949).

Woodworth, R. *Experimental Psychology*. (New York: Holt, 1938)

Yin, T. C. and 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).