Natural 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

Anticipating invariant statistical structure of sensory cues may optimize perception. We investigated this hypothesis in human sound localization, focusing on context-invariant interaural time difference (ITD) statistics. Natural ITD rate of change across azimuth (ITDrc) and a novel metric of ITD variability over time (ITDv) were estimated using head-related transfer functions (HRTFs) and modeled cochlear filters. Using protocols which disabled ITDrc estimation and delivered zero-ITDv stimuli, we found that spatial discriminability and novelty detection are consistent with the hypothesis that the brain represents and predicts natural ITD statistics, exploiting them for optimizing discrimination. We further show that natural ITD statistics could be represented under coding frameworks postulated by classic neural models, highlighting the validity of these models and providing potential explanation for the observed distribution of best ITDs across frequency found in mammalian brainstem. These findings support the hypothesis that natural ITD statistics are built into the neural substrate underlying perception.

Keywords: Binaural cues; HRTF; Cochlear filtering; Sensory reliability; Discriminability; MMN; Novelty detection; Predictive coding.

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 higher ITD rate of change (ITDrc) 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-azimuth relationship (Aaronson and Hartmann 2014; Roth et al. 1980). In addition, because the brain computes 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 ITD statistics in the human brain. Using human HRTF databases and models of cochlear filters, we estimated ITDrc and ITDv. We tested whether these statistics, separately and combined, predicted spatial discrimination thresholds and deviance detection. Using sound stimulation that removes ITDv and disables ITDrc estimation (tones delivered by earphones) allowed us to test whether anticipated, rather than recent or ongoing, statistics predicted spatial auditory perception. Our results indicate that expected ITDrc and ITDv determine 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 and models of cochlear filters. We first tested whether these ITD statistics predict human spatial discrimination thresholds using previously published datasets (Mills 1958) and data collected with tests specifically designed for measuring ITD discrimination. Next, we used EEG and mismatch negativity signals (MMN) to address the question of whether these natural ITD statistics influence ITD deviance detection.

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 in sounds reaching the ears of human subjects. The method for estimating the ITD mean and standard deviation (**Figure 1A**), which was applied across locations and frequencies, included: (1) Impulse responses obtained from publicly available human HRTF databases (Listen HRTF database, http://recherche.ircam.fr/equipes/salles/listen; 51 subjects) where used to convolve acoustic signals by head related impulse responses, which results in modulation of ongoing phase and gain that depends on

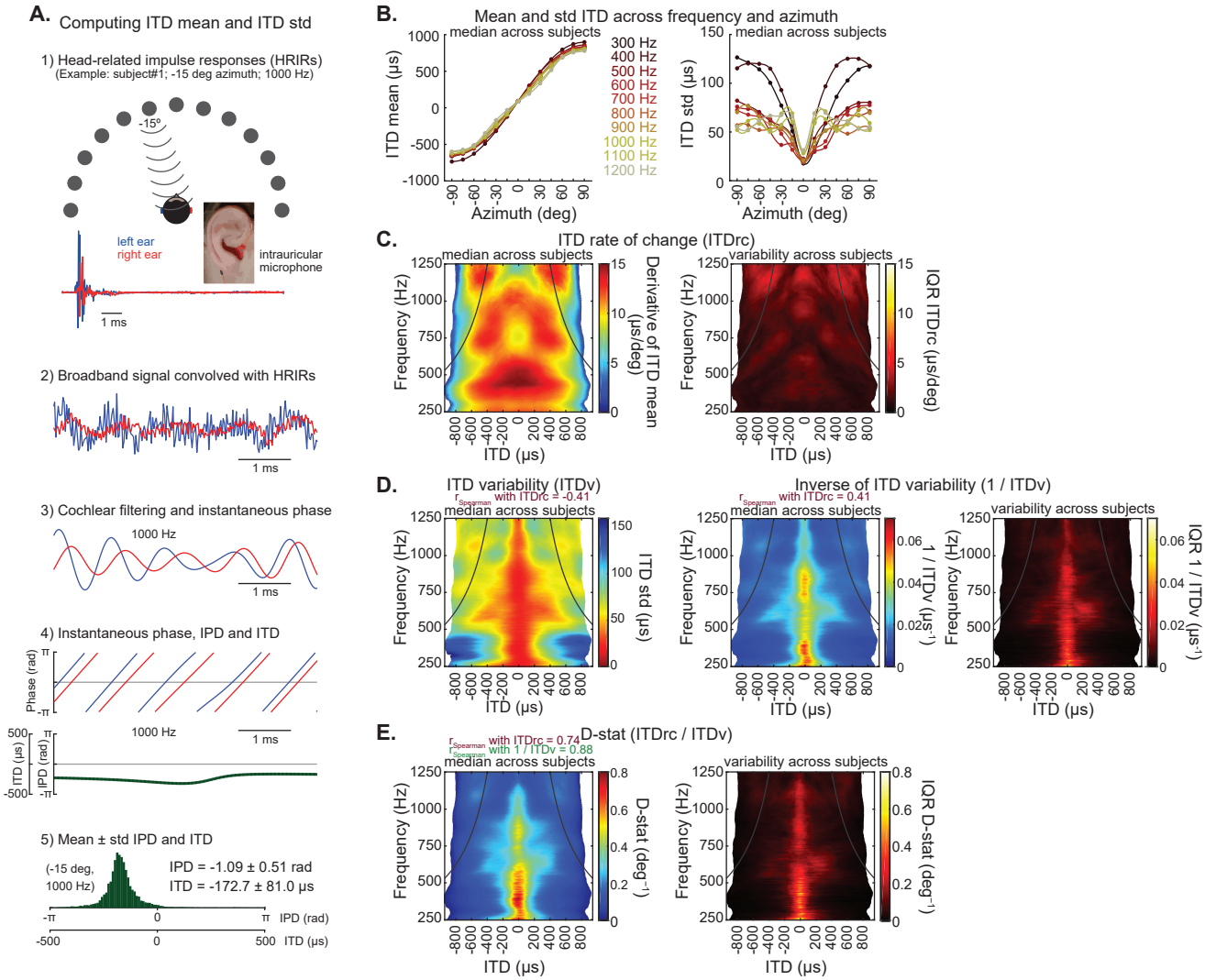

Figure 1. ITD statistics estimated from human HRTFs and modeled cochlear filters. (**A**) Estimation of ITD mean and ITD standard deviation (std) over time in broadband signals filtered by HRTFs and cochlear filters. (1) Head-related impulse responses (HRIRs), computed using sounds from speakers located at different azimuth positions and collected by microphones positioned in each ear, were obtained from a publicly available dataset (LISTEN). Traces show example impulse responses in right (red) and left (blue) ears elicited by a sound at azimuth −15°. (2) A broadband signal was convolved with HRIRs from right (red) and left (blue) ears for each particular direction. (3) Convolved signals were then filtered using parameters analogous to human cochlear filters. Example of signal passed through a cochlear filter with a frequency band centered on 1000 Hz. (4) The instantaneous phase of the resulting signals on the left and right ears was computed. Top, instantaneous phase for the left and right signals shown in 3. Bottom, instantaneous phase differences (IPD, in radians) and instantaneous time differences (ITD, in microseconds) between left and right signals. (5) Histogram of instantaneous IPD and ITD, illustrating their variability over time for the example signal shown in 3. (**B**) ITD mean (left) and std (right) over time, as a function of frequency and azimuth. Plots represent median values across subjects (N = 51), fit by spline curves, and color coded for each frequency. The derivative of the curves on the left was used to calculate ITD rate of change (ITDrc) across azimuth. The ITD variability (ITDv) was computed as the standard deviation of the ITD distribution over time. (**C**) Left, median ITDrc computed from data across subjects as a function of stimulus frequency and azimuth (azimuth was converted to ITD to obtain an estimate of ITDrc as a function of frequency and ITD, matching the stimulus metrics and model parameters used in our study). As previously reported (Mills et al. 1972; Gelfand 2016), ITDrc is higher mostly in the front, although the pattern is complex. Right, inter-individual variability represented by interquartile range of ITDrc across subjects, which displays low levels. (**D**) Left, median ITDv as a function of stimulus frequency and azimuth. ITDv is lower in the front and higher in the periphery across frequency, with a more pronounced pattern in low frequencies. ITDv showed low correlation with ITDrc (coefficient shown in red above the plot). Middle, inverse ITDv (1/ITDv), to obtain a metric of ITD reliability proportional to statistical discriminability. Right, inter-individual variability represented by interquartile intervals of 1/ITDv across subjects; which also display low levels (except for the frontal ITDs in the frequency close to 250 Hz). (**E**) Left, median ITD statistical discriminability index (D-stat) as a function of frequency and ITD corresponding to each azimuth, computed as the ITDrc/ITDv ratio across subjects. D-stat shows strong correlation with ITDrc and ITDv (coefficients shown above the plot in red and green, respectively). Right, interquartile range of D-stat across subjects (right) shows low inter-individual variability (except for frontal ITDs in lower frequencies). Black curves on each panel indicated the π-limit across frequency, beyond which ITD information becomes ambiguous for narrowband sounds.

both frequency and sound direction. (2) Filtering sounds using parameters analogous to models of human cochlear filters (Glasberg and Moore 1990). (3) Extracting instantaneous phase of the resulting signals and instantaneous interaural phase difference (IPD). (4) Computing mean and standard deviation of instantaneous IPD, and conversion to ITD. As result, we estimated ITD mean and ITD standard deviation over time, at each azimuth and frequency, for each of the subjects (**Figure 1B**). (5) ITDrc was calculated as the derivative of mean ITD over azimuth, while ITDv is the standard deviation of ITD. ITDrc was higher (**Figure 1C**) while ITDv was lower (**Figure 1D**) 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**). An index capturing both ITDrc and ITDv, referred to as D-stat, was computed by dividing ITDrc by ITDv at each location and frequency (**Figure 1E**). Although D-stat is inspired by and is conceptually similar to the d-prime calculation, d-prime and D-stat are not identical quantities (while d-prime is the subtraction of two means divided by the standard deviation, D-stat is the derivative of the mean divided by the standard deviation).

ITD statistics were computed for each of the 51 subjects and then the median and interquartile range were calculated. These statistics displayed low variability across individuals (**Figure 1C-E**, right panels) indicating they constitute a set of statistics largely invariant across human subjects. This provided justification for using ITDrc and ITDv estimated on these 51 subjects for predicting ITD discriminability collected from other groups of subjects.

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 should affect ITDv as well; however, ITDv also depends on phase and gain modulations induced by interaction between neighboring frequencies, within the critical band of single cochlear filters (**Figure 1—figure supplement 1A**). Consistently, the correlation between ITDrc and ITDv was not strong ($r_{Spearman} = -0.41$). These ITD statistics were consistent across broadband signals with different frequency spectra (**Figure 1—figure supplement 1B**). Additionally, we tested the 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). Echoes significantly disrupted ITD statistics; however, the precedence effect is expected to segregate leading signals from their lagging echoes (Wallach et al. 1949; Brown et al. 2015). Accordingly, ITD statistics estimated in anechoic and reverberant environments were similar when signal transients were considered (**Figure 1—figure supplement 1C**). 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. Invariance across contexts is a premise motivating the question of whether ITD statistics are anticipated in auditory spatial perception.

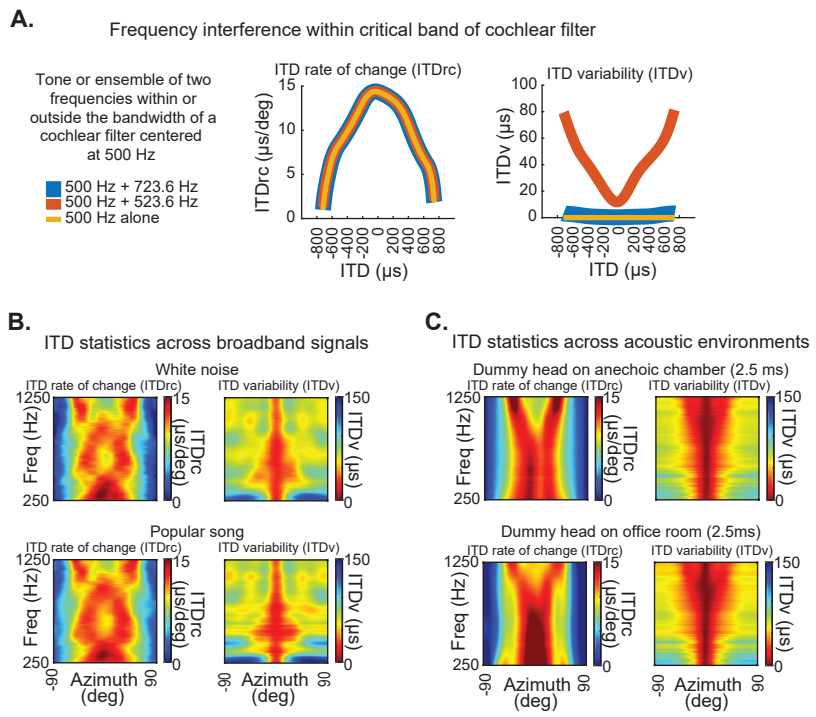

Figure 1—figure supplement 1. Basis of ITD statistics and consistency across conditions. (**A**) ITDrc (left) and ITDv (right) as a function of ITDs corresponding to azimuth locations from -90° to 90°, averaged across subjects, for three example signals: a 500 Hz tone (yellow) and two signals combining 500 Hz and 523 Hz (red), and 500 Hz and 723 Hz (blue). 523 Hz and 723 Hz were chosen to demonstrate the effect of combining signals lying, respectively, within and outside the bandwidth of a cochlear filter with a center frequency of 500 Hz. The ITD statistics were estimated using the same method described in Figure 1, by convolving each of these signals with impulse responses from the left and right ears, then filtering using parameters analogous to human cochlear at 500 Hz center frequency. This procedure was applied across all other locations and subjects in the reported dataset. Note that ITDrc is similar for the three signals, indicating no significant effect of frequency interference on this statistic. However, while ITDv is near zero across azimuth locations for 500 Hz alone (yellow) and 500 + 723 Hz (blue), it increases from front to periphery when neighboring frequencies within a single cochlear filter are combined (500 + 523 Hz; red). Hence, unlike ITDrc, ITDv depends on interference between neighboring frequencies within cochlear filter ranges. (**B**) ITD statistics are consistent across broadband signals. ITDrc (left) and ITDv (right) computed using white noise (the broadband noise used to estimate ITD statistics in Figure 1C and 1D; top) and a distinct broadband signal (first second of the chorus section of the popular song Alive, by Pearl Jam; bottom) displayed similar patterns. (**C**) ITD statistics are consistent across acoustic environments. ITDrc (left) and ITDv (right) computed from impulse responses recorded in human-head shaped manikins inside two different types of environment - anechoic chambers (top) and reverberant office spaces (bottom). We used the first 2.5 ms of impulse responses (corresponding mostly to the direct click reaching the ears, avoiding echo components in the reverberant office).

Different auditory cue statistics have been examined in previous studies, as well as whether consistent statistics across environments shape auditory spatial perception. High ITDrc in the midline supports the better sound discrimination observed 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, we also found that ITDrc varied across frequency. Interestingly, we found cases where the steepest ITDrc occurred at locations distant from the midline. Although this was not directly addressed by previous studies (Kuhn 1977; Benichoux et al. 2016), it is observable in the data of these studies (**Figure 1—figure supplement 2A-C)**, providing convergent evidence of ITDrc patterns across studies. In particular, rate-of-change of envelope ITD, also reported by Benichoux et al. (2016) showed maximum values away from the midline across most frequencies (**Figure 1 - figure supplement 2C**). This contrasts the premise of a direct relationship between ITD rate of change across azimuth and spatial discriminability (Mills et al. 1972; Gelfand 2016). Thus, we excluded envelope ITDrc from our analysis.

Across-trial ITDv induced by concurrent sounds was reported by Cazettes et al. (2014) as a metric relevant to the owl's auditory spatial perception. When this method was applied to human HRTFs and human cochlear filters, for the range of frequencies of interest in human ITD detection, we found ITDv values weakly correlated to those obtained with the method used in the present study (**Figure 1—figure supplement 2D**). However, a stronger correlation between both metrics was observed for the range of frequencies most relevant to owls' sound localization (above 2000 Hz), suggesting that the effect of concurrent sounds on ITDv may not represent a significant source of ITDv in humans. Młynarski and Jost (2014) also estimated auditory cue statistics across environments. However they did so without reporting the location of sound sources, which restricted the use of their estimated statistics for testing the prediction of sound discrimination across locations.

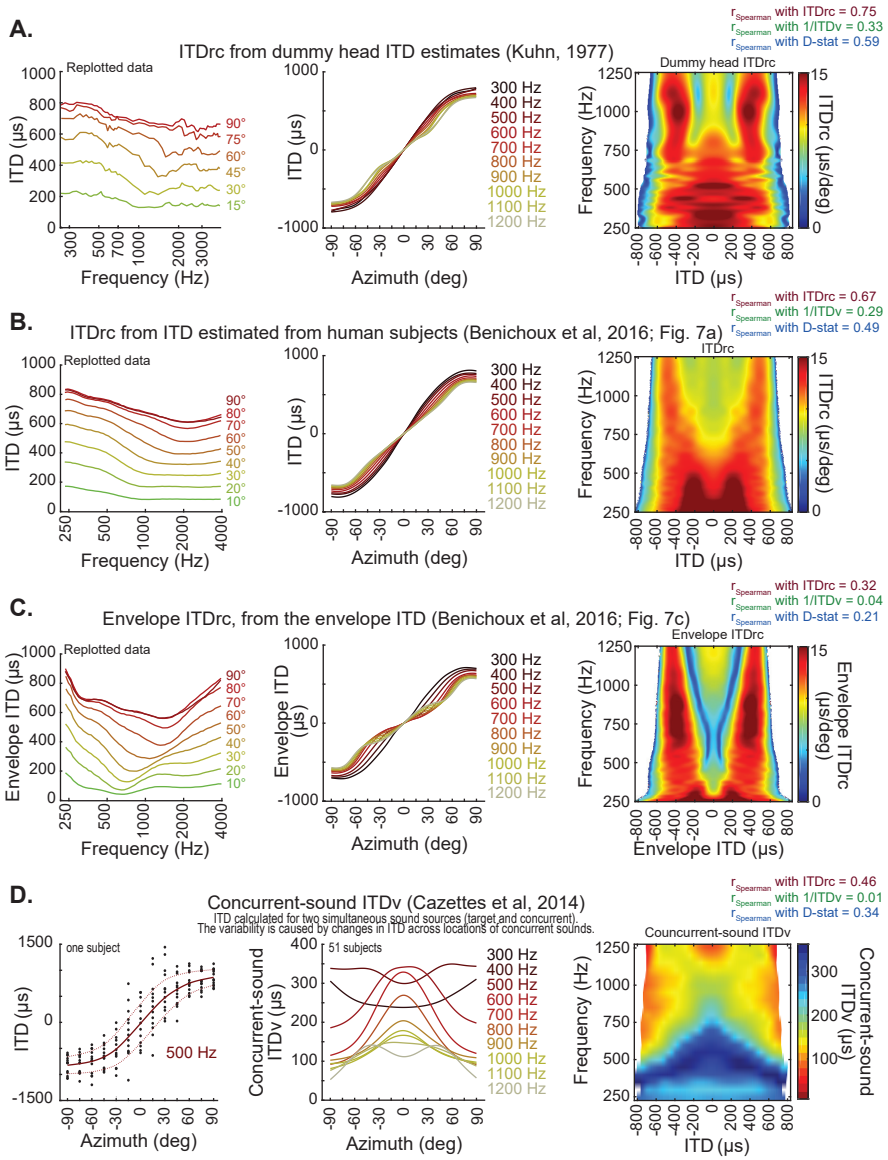

Figure 1—figure supplement 2. Previous reports of ITD statistics. (**A**) ITDrc computed across frequency and azimuth from pioneer recordings made on human manikins (Kuhn, 1977). Left, original figure of Kuhn (1977) showing ITD as a function of frequency for a range of angles of incidence (color code on the right). Middle, ITD across azimuth for each frequency, fit by a spline curve. Right, ITDrc across ITD corresponding to the azimuth. The ITDrc estimated from Kuhn (1977) data correlated strongly (Spearman correlation coefficients above right panel) with our-study's ITDrc (red), weakly with ITDv (green) and moderately with D-stat (blue), indicating consistency of ITDrc estimations across datasets and independence of ITDrc and ITDv statistics. (**B-C**) Human ITD across frequency and azimuth described in Benichoux et al. (2016). The panels depict ITD measurements estimated from 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 location within an azimuth range of −90 to 90 degrees sampled in 15-degree intervals and zero elevation (13 different locations), target broadband signals were combined with concurrent broadband sounds coming from each of the 13 directions. Left, example distribution of ITDs computed for the 500-Hz cochlear filter; black dots represent ITDs resulting from the combined target and concurrent sounds, across −90 to 90 degrees of azimuth, for a single subject. Middle, concurrent-sound ITDv estimates (mean across subjects) as a function of target signal's azimuth for each frequency of interest. Right, concurrent-sound ITDv estimates as a function of target sound ITD. Concurrent-sound ITDv shows weak to moderate correlation with the ITD statistics estimated in our study for the range of frequencies of interest (text in red, green and blue).

In the current study, we hypothesized that ITDv induced by frequency interference within cochlear filters is a significant high-order statistic of the crucial binaural cue ITD, and tested ITDv's prediction accuracy (alone and combined with ITDrc) of sound discrimination performance and novelty detection.

ITD discriminability predicted by classical neural models of ITD coding

Classic models of ITD coding propose brainstem mechanisms underlying ITD discriminability. The model by Stern and Colburn (1978) postulates an increased density of pairs of fibers underlying tuning to ITDs near the midline as the basis for increased ITD discriminability in the front (**Figure 2A**). This parameter of the model showed high correlation with all ITD statistics examined in the current study, especially the D-stat (**Figure 2A**). Thus, this was tested as a predictor of ITD discriminability in the following sections. Additionally, we conducted an analysis to test whether this model parameter could implement a representation of ITD statistics (**Figure 2B**). Consistently, representing D-stat required only a subtle change in the distribution proposed by Stern and Colburn (1978), supporting the idea that this seminal model, which explains multiple experimental findings, is also consistent with a representation of ITD statistics.

On the other hand, the model postulated by Harper and McAlpine (2004) relies on the maximization of Fisher information within the physiological ITD range to explain the optimal IPD-tuning distribution in a population of brainstem neurons. This model provides an explanation for the tuning to peripheral ITDs reported in brainstem recordings

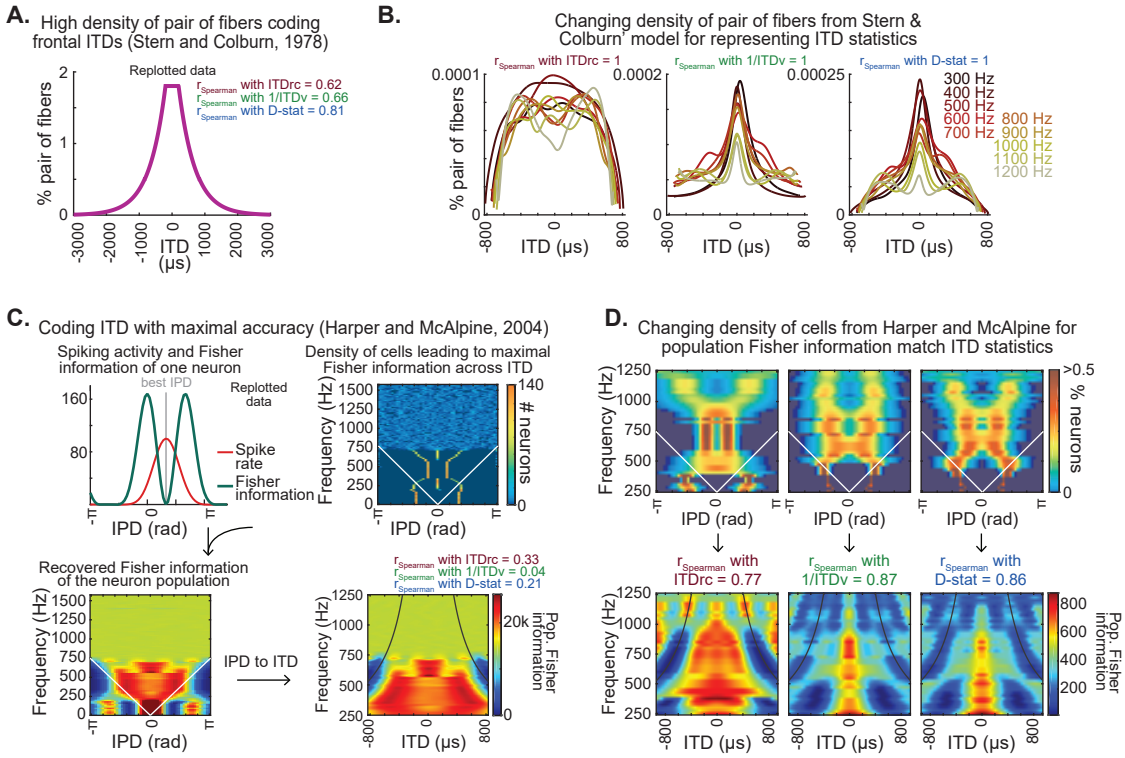

Figure 2. Classic models of neural properties underlying ITD discriminability and their potential for explaining representation of ITD statistics. (**A**) Distribution of internal delays replotted from Stern and Colburn (1978, Fig. 2b), which proposes a higher density of pairs of fibers encoding frontal ITDs. (**B**) The density of pairs of fibers proposed by Stern and Colburn (1978) as the mechanism underlying ITD discriminability could effectively achieve the representation of ITD statistics. The density was adjusted to match the pattern of ITDrc (left), the inverse of ITDv (middle) and D-stat (left). Note that distributions matching 1/ITDv and D-stat largely preserve the shape of the distribution of the original model (shown in A), further supporting the explanatory power of this model across experimental datasets. (**C**) Distribution of IPD tuning maximizing information carried by firing rate across the midline, proposed by Harper and McAlpine (2004). Spike rate and Fisher information as a function of best IPD (top-left) and distribution of expected best IPDs of ITD-sensitive neurons in the mammalian brain across frequency (top-right), replotted from data reported in Harper and McAlpine (2004). Bottom-left, Fisher information summed across the complete neuron population, calculated by convolving the IPD tuning distribution with the Fisher information of a single neuron; white straight lines indicate physiological range determined by the distance between ears. Bottom-right, by converting population Fisher information from IPD to ITD for each frequency, we obtain the predicted ITD discriminability (black curved lines indicate the π-limit, beyond which ITD information becomes ambiguous within narrow frequency bands). (**D**) The IPD tuning distribution proposed by Harper and McAlpine (2004) as a mechanism underlying ITD discriminability was adjusted making the population Fisher information match ITDrc, ITDv and D-stat. Note that neuron distributions matching ITD statistics shown in the top plots depict 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, population Fisher information obtained by convolving each of the distributions shown on top with the single-neuron Fisher information; population Fisher information highly correlated with the corresponding ITD statistics ($r_{Spearman} \geq 0.77$, shown on top of each plot).

of mammals (McAlpine et al. 2001; Hancock and Delgutte 2004; Pecka et al. 2008). We followed the model's method of calculating Fisher information of individual neurons' IPD tuning curves (**Figure 2C**, top left) and the proposed distribution of best IPDs of the neuronal population (**Figure 2C**, top right) for computing the population's Fisher information (**Figure 2C**, bottom left). The population's Fisher information across ITD and frequency corresponds to the estimated ITD discriminability from this neural model (**Figure 2C**, bottom right) and will be used as a predictor of experimental ITD discriminability in the following sections.

Furthermore, we found that ITD statistics could also be represented by a model relying on the Fisher information framework (**Figure 2D**). The IPD tuning distributions that correlate best with ITD statistics from the current study (**Figure 2C**, plots on the top-right) differ from the distribution originally proposed by Harper and McAlpine (2004). However, representing ITD statistics under this coding framework predicted neurons' IPD tuning clustered within peripheral IPDs across the frequency range tested, consistent with the reports of brainstem recordings in mammals (McAlpine et al. 2001; Hancock and Delgutte 2004; Pecka et al. 2008).

These results suggest the mechanisms underlying ITD discriminability proposed by Stern and Colburn (1978) and Harper and McAlpine (2004) models can implement a representation of ITD statistics, thus providing a plausible biological connection between the coding of sensory statistics and perceptual functions.

To test the hypothesis that ITD statistics drive auditory spatial perception, we explored the correlation between discrimination performance and three groups of predictors (current-study ITD statistics, previously-reported ITD statistics, and ITD discriminability predicted by classical neural models, presented in **Figures 1-2**).

Prediction of spatial discrimination thresholds from ITD statistics

Free-field dITD thresholds as a function of ITD and frequency were obtained from a classic study of human sound localization (Mills 1958) (**Figure 3A-B**). In addition, a test measuring dITD thresholds through dichotic (earphone) sound delivery was conducted (**Figure 3C-D**). Neither of these protocols delivered stimuli carrying natural ITD statistics: they both used tones, which abolishes ITDv, and disables ITDrc estimation by either fixing the head of subjects (in free field) or by decoupling head movement and ITD input (in dichotic). Thus the effect of predicted natural ITD statistics could be assessed in both approaches.

We first tested whether ITDrc, ITDv or their ratio (D-stat) predicted the free-field dITD thresholds 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, converted to dITD) across reference ITD and frequency in human subjects. **Figure 3B** shows the comparison of correlation between free-field dITD thresholds reported by Mills (1958) and: (1) ITD statistics reported by other studies (Kuhn 1977; Benichoux et al. 2016; Cazettes et al 2014); (2) ITD discriminability predicted by classic neural models (Stern and Colburn 1978; Harper and McAlpine 2004); and (3) ITDrc, 1/ITDv and D-stat estimated in our study. D-stat displayed higher correlation with dITD thresholds than other ITD statistics. The neural model of ITD discriminability by Stern and Colburn (1978) was also highly correlated with dITD thresholds. 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**). These results are consistent with the hypothesis that ITD statistics are built into the brain processing underlying discrimination thresholds.

Additionally, to test ITD change discrimination thresholds and evaluate prediction accuracy of ITD statistics we presented sounds through earphones instead of free-field stimulation in the current study, thus avoiding the potential effect of ongoing ITD statistics and the influence of other sound localization cues. The 24 normal-hearing adults that participated in the testing were instructed to detect a change in ITD within a pair of tonal sounds (**Figure 3C-left**), providing dichotic dITD thresholds across reference ITD and frequency, for the range of interest (**Figure 3C-right**). The dichotic dITD thresholds averaged across subjects correlated with free-field thresholds reported by Mills (1958) ($r_{Spearman} = 0.64$). However, dITD thresholds were higher in the present study compared to Mills (1958); other studies also using dichotic stimuli reported higher dITD thresholds (e.g., Brughera et al., 2013) than Mills (1958). This may be due to approach differences, such as presentation of sounds through earphones vs. free field stimulation and testing of untrained subjects in the present study rather than highly trained individuals in Mills (1958). Using free-field stimulation leaves open the possibility that listeners rely on cues other than ITD to detect sound location in azimuth, which may have lowered the thresholds as was done in Mills (1958). Additionally, training in ITD detection may have an effect on threshold levels compared to normal untrained individuals.

We computed the average dichotic dITD thresholds across participants and quantified the correlation between them and: (1) ITD statistics reported by other studies, (2) ITD discriminability predicted by classic neural models, and (3) ITD statistics estimated in our study. When all frequency conditions were analyzed, average dichotic dITD thresholds showed moderate correlation with several of the predictors, including ITDrc and D-stat of our study, ITDrc as estimated by Benichoux et al 2016, and also the ITD discriminability predicted by both Stern and Colburn's and Harper and McAlpine's models (**Figure 3D**, top-left). Particularly, dITD thresholds for 250 Hz were a consistent outlier across predictors (**Figure 3D**, right plots). Higher thresholds for this frequency have previously been reported (Brughera et al. 2013). Moreover, exclusion of this frequency further improved D-stat's prediction accuracy (**Figure 3D**, bottom-

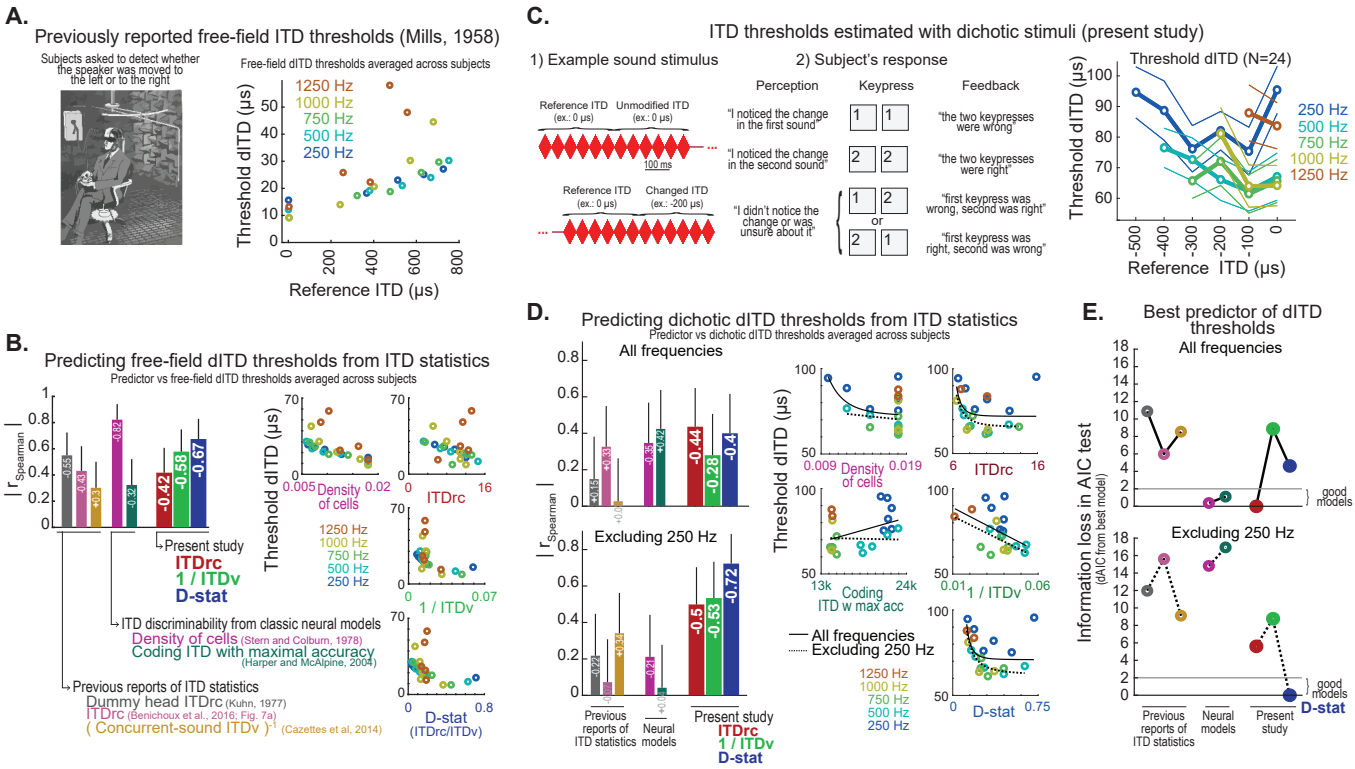

Figure 3. ITD statistics predict human ITD-change detection thresholds to free-field and dichotic tones. (**A**) Minimum ITD change detection (threshold dITD) across frequency and locations for sounds in free-field, reproduced from Mills (1958); doi: 10.1121/1.1909553, with the permission of the Acoustical Society of America). Left, free-field setup where minimum azimuth change detection as a function of location and frequency were estimated and converted to dIPD a function of reference ITD, averaged across subjects. Right, dIPD as a function of reference ITD reported by Mills (1958) replotted as dITD as a function of reference ITD for the frequencies of interest. (**B**) Left, coefficients of the correlation between free-field dITD thresholds across subjects reported by Mills (1958) and ITD statistics from three previous reports (Kuhn 1977; Benichoux et al. 2016; Cazzetes et al. 2014), ITD discriminability predicted by two classic neural models (Stern and Colburn 1978; Harper and McAlpine 2004) and ITD statistics estimated in the present study. Bars and whiskers indicate Spearman correlation coefficients and standard error across conditions (25 combinations of frequency and reference locations averaged across subjects). Scatter plots on the right show free-field dITD thresholds as a function of each of the ITD statistics estimated in the present study (ITDrc, 1/ITDv, D-stat) and also ITD discriminability by Stern and Colburn (1978); this classic model and present study's D-stat show the highest correlation coefficients. (**C**) Left, schematic of the test conducted in the present study to specifically assess ITD-discrimination (dITD) thresholds for tonal sounds delivered through headphones. 1) Two binaural tones with an identical 'reference' ITD were presented in sequence through headphones. ITD was changed towards the periphery in the last half of one of these sounds. 2) 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. Right, dichotic threshold dITDs as a function of reference ITD across frequency. Bold lines represent the mean across subjects and thin lines standard error of the mean. (**D**) Left, coefficients of the correlations between dichotic dITD thresholds across subjects and ITD statistics of previous reports, ITD discriminability predicted by classic neural models and ITD statistics estimated from the present study. Bars and whiskers indicate Spearman correlation coefficients and standard error across conditions. Scatter plots on the right show dichotic dITD thresholds as a function of the ITD statistics (black line: all the frequencies are analyzed; dotted line: 250 Hz frequency is excluded from the analysis). When all the frequencies are analyzed (top bar plot), several predictors show moderate correlation with dITD threshold. When the 250 Hz frequency is excluded from the analysis (bottom bar plot), D-stat from the present study yields the highest negative correlation with threshold dITD. (**E**) AIC test for selecting the best predictor of dITD thresholds across all frequencies (top), showing that ITDrc and ITD discriminability predicted by both Stern and Colburn's model and Harper and McAlpine's model (although with a positive correlation in this last one, indicating a relationship between population Fisher information and threshold dITD opposite to the expected). When 250-Hz conditions were not included, AIC test shows that D-stat carries the best prediction accuracy of dITD thresholds compared to any previous report of ITD statistics or neural models of ITD discriminability (bottom).

left), suggesting that dITD thresholds at 250 Hz may be determined by additional parameters not addressed by the ITD statistics investigated in our study. The AIC test, which accounts for individual variability, replicated results of the Spearman correlation analysis showing that several models (ITDrc and ITD discriminability predicted by both Stern and Colburn's and Harper and McAlpine's models) have equivalent prediction accuracy of dITD across all frequencies. Further analysis excluding 250-Hz conditions revealed that D-stat was the best predictor of dITD thresholds (**Figure 4E**, bottom-left).\

ITD statistics built into deviance detection

To further test if statistics are built into the human brain processing, we measured the mismatch negativity (MMN) component of event-related brain potentials (ERPs) (Näätänen et al. 1978) for sounds coming from standard (repeated) and deviant (sporadic) locations. MMNs are observable when a sequence of standard stimuli is unexpectedly interrupted by a low probability stimulus, without the listener making an overt response (Näätänen et al. 1978; Pakarinen et al. 2007; Sussman 2007; Sussman et al. 2013) (**Figure 4A**). Thus, the MMN signals provide a direct brain measure of discriminability that does not require training subjects to perform behavioral tasks. The MMN signal is displayed by subtracting the mean ERP response elicited by the standard stimuli from the mean ERP elicited by deviant stimuli. The amplitude of the MMN indexes discriminability between standard and deviant sounds. The larger the tone features separation between standard and deviant stimuli, the more negative the MMN amplitude (Deouell et al. 2006; Sams et al. 1985; Pakarinen et al. 2007; Tiitinen et al. 1994). Thus, MMN was used to address the question of whether anticipated ITD statistics influences the robustness of ITD discrimination.

A set of ITD and frequency conditions was selected and presented to 33 normal hearing adults in order to sample critical ranges drawn from the HRTF analysis (**Figure 4B**). 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 prediction accuracy across ITD statistics. 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 separately across participants and conditions. The peak amplitude of MMN was used to quantify the subject's capacity to discriminate between ITDs of the standard and deviant. The characteristic fronto-central scalp topography of the MMN responses were observed (Giard et al. 1990) (**Figure 4C**).

We then examined the prediction accuracy of MMN amplitude from model equations relying on ITD statistics from previous work, ITD discriminability from classic neural models, and ITD statistics proposed by the current study. In these model equations, the absolute difference between the ITDs of standard and deviant stimuli was adjusted by the weighted sum of ITD statistics (or ITD discriminability predicted by each neural model) of standard and deviant stimuli. This adjustment was implemented by multiplying or dividing the absolute ITD difference by metrics that magnify (e.g. ITDrc) and blur (e.g. ITDv) ITD discriminability, respectively (**Figure 4D**). As an example, the equation we used to test the prediction of MMN amplitude by D-stat was:

$$MMNpeak \sim |ITD_d - ITD_s|(w_s Dstat_s + (1 - w_s)Dstat_d)\,,$$

where ITD_s and ITD_d are the ITD of standard and deviant, w_s and $1\text{-}w_s$ are the relative weights of the standard and deviant, and $Dstat_s$ and $Dstat_d$ are the estimated D-stat values corresponding to the frequency and ITD of the standard and deviant stimuli.

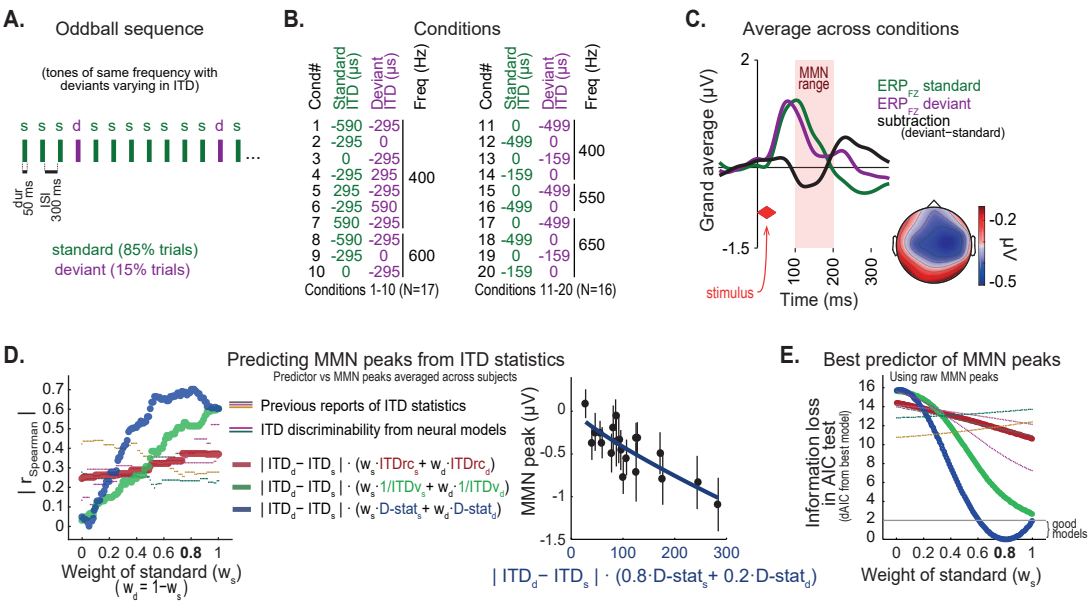

Figure 4. ITD statistics predict discriminability of spatial deviants indexed by MMN responses. (**A**) Passive oddball sequence protocol, in which subjects listened to frequent "standard" stimuli embedded with rare "deviants". In each condition, two tones of the same frequency and distinct ITDs were presented. (**B**) 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 brain areas. (**D**) Left, coefficients of correlations between MMN amplitude and different predictor equations normalizing ITD difference between standard and deviant by ITD statistics and ITD discriminability (same color code as in Figure 3B), as a function of the relative weight of the standard stimulus (w_s), relative to the weight of the deviant (w_d). Right plot: best prediction of MMN amplitude by D-stat, weighting standard more than deviant (80%:20%). (**E**) AIC test for selecting the best model, with the same outcome of correlation analysis in D.

Figure 4D shows the correlation between each of the predictors' output and the amplitude of MMN peaks (averaged across subjects) as a function of the weight of the standard. The highest correlation is found when multiplying the ITD difference between standard and deviant by D-stat, assigning 80% weight to the standard and 20% to the deviant. The same outcome was found using the AIC test (**Figure 4E)**, which considers individual variability. The good prediction of MMN by the model relying on D-stat further supports the idea that combined ITDrc and ITDv information is critical in auditory spatial perception.

Discussion

Human HRTFs and models of cochlear frequency filtering were used to estimate two specific 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. Models relying on ITDrc, ITDv and their ratio (D-stat) were good predictors of ITD discriminability and novelty detection, demonstrating 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, including an uneven density of brainstem ITD-sensitive neurons (Colburn 1977; Stern and Colburn 1978) and greater change in the firing rate of these neurons 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. 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 ITDrc and ITDv, inherent in natural acoustic scenes, are relevant parameters determining ITD discriminability even for tonal stimuli, which do not contain these ITD statistics. In fact, our results, rather than contradicting the classic neural models for coding ITD, further support the notion that network properties supporting discriminability are evolutionarily adapted to capture sensory statistics. Both classic neural models are compatible with the representation of these statistics, but the present work provides evidence supporting the idea that spatial cue statistics are relevant parameters for explaining auditory spatial discriminability. Specifically, this study demonstrates that ITD discriminability could be best predicted by D-stat, and thus provides additional justification for the uneven density of brainstem ITD-sensitive neurons postulated by Stern and Colburn (1978) and the clustered subpopulations tuned to IPDs away from the midline reported in the mammalian brain (McAlpine et. al 2001; Harper and McAlpine 2004; Hancock and Delgutte 2004; Pecka et al. 2008).

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 properties of the neural mechanisms underlying human and owl sound localization differ in frequency range and putative ITD coding schemes (Schnupp and Carr 2009), studies in both species support the concept that natural ITD statistics guide ITD processing. Specifically, our results support the hypothesis that ITDrc and ITDv determine human sound localization discriminability and novelty detection of acoustic spatial deviants.

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 demonstrate 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. We found that the D-stat statistic is a better predictor of ITD discriminability than ITDrc alone. This supports the hypothesis that the brain anticipates and therefore represents both sets of natural statistics included in the D-stat ratio.

The current study found 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, the high dITD thresholds for 250 Hz tonal stimuli (see also Brughera et al. 2013) constituted an outlier that revealed limited prediction by ITD statistics. 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

An MMN paradigm was used to find converging evidence to address the question of whether anticipated ITDv is built into spatial perception. 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. Consistent with results of discriminability thresholds, the ratio between ITDrc and ITDv (D-stat) was a better predictor of deviant detection responses.

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 have modulated novelty

detection at both levels. However, the model that best described the MMN signals weighted the difference between the ITDs of standard and deviant primarily by the anticipated ITD statistics of the standard stimulus. 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. Results demonstrate that when the standard is in a location with anticipated higher statistical discriminability, a 'stronger' standard is built.

In sum, we found evidence that anticipated ITD statistics correlated with auditory spatial perception. 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, where larger ITD changes are required for detecting shifts in location in regions of space and frequency ranges for which ITD discriminability is anticipated to be weaker. This study provides evidence supporting the hypothesis that natural ITD 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, at a distance of 1.95 m from the center of the head. 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 on 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-made amplifier with 40 dB gain, and recorded using an RME sound card interface with Max/MSP real time library which deconvolved the microphone signal.

HRTF analysis

HRTF data from 51 subjects were included in the HRTF analysis (**Figure 1**). Head-related impulse responses (h) for the left (L) and right (R) ears corresponding to speaker locations at 0-degree in elevation and −90 to 90 degrees in azimuth (θ) were denoted as a function of time, $h_{L,\theta}(t)$ and $h_{R,\theta}(t)$. The azimuth in the database was sampled in 15-degree steps. Impulse responses were convolved with a white noise signal of 1-second duration $s(t)$ to model the signals (x) received at the left and right ears:

$$x_L(t) = h_{L,\Theta}(t) * s(t)$$
$$x_R(t) = h_{R,\Theta}(t) * s(t),$$

where $*$ denotes convolution.

This procedure transfers temporal and level effects of the filtering properties of the head to the convolved signal. These convolved signals were filtered by narrow-band filters modeling cochlear processing, using the gamma-tone filter bank from Malcolm Slaney's Auditory Toolbox, available in https://engineering.purdue.edu/~malcolm/interval/1998-010). Gamma-tone filters are described in the following cochlear impulse response equation,

$$g(t; f_k) = t^3 e^{\frac{-t}{T_k}} cos(2\Pi f_k) U(t),$$

where $U(t)$ is the unit step function and the center frequencies of the filters (f_k) ranged 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 time constants (t_k) were chosen such that the bandwidth of these filters matched the estimated bandwidth of human cochlear filters (Glasberg and Moore 1990).

The outputs of the gamma-tone filter bank on the left ($y_L(t;f_k)$) and right ($y_R(t;f_k)$) sides were computed by convolving left- and right-ear input signals with gamma-tone filters,

$$y_L(t; f_k) = g(t; f_k) * x_L(t)$$
$$y_R(t; f_k) = g(t; f_k) * x_R(t).$$

Instantaneous phase was then computed for these output signals using the Signal Processing Toolbox (Mathworks). The instantaneous phase was computed as the phase (argument; *arg*) of the analytic representation of the signal,

$\Theta_L(t; f_k) = \arg\{y_L(t; f_k) + i\hat{y}_L(t; f_k)\}$

$\Theta_R(t; f_k) = \arg\{y_R L(t; f_k) + i\hat{y}_R(t; f_k)\}$,

where y is the signal and $\hat{y}$ is its Hilbert transform.

For each azimuth (θ) and frequency range (f_k), we then calculated the instantaneous interaural phase difference (IPD) over time,

$IPD(t; f_k) = \Theta_R(t; f_k) - \Theta_L(t; f_k)$,

where $IPD(t;f_k)$ is in radians.

The circular mean and standard deviation of the instantaneous IPD over time was then computed. To avoid the ITD rate of change being corrupted by artificial values caused by phase ambiguity, we unwrapped (Matlab function) the mean IPD over azimuth, and subtracted the value 2π repeatedly (from all IPD values jointly) until the IPDs corresponding to midline locations returned to the value before the shift. Finally, the circular mean and standard deviation of IPD was converted to ITD (in μs) using the following equation:

$$ITD = \frac{10^6 IPD}{2\pi f}.$$

These steps are shown in **Figure 1A**.

For each subject, the mean ITD across azimuth was interpolated using a cubic spline, and the rate of change of ITD across azimuth (ITDrc) was calculated as the derivative of this curve. Then, the median and interquartile range of ITDrc were computed for each combination of azimuth and frequency across subjects. Finally, azimuth was converted to ITD (using the relationship between azimuth vs ITD determined from the HRTFs), obtaining an estimate of ITDrc across frequency and ITD (**Figure 1C**). For estimating ITD variability (ITDv), the standard deviation of ITD over time and across azimuths was interpolated for each subject using a cubic spline. Then, for each combination of azimuth and frequency, the median and interquartile range of ITDv was computed across subjects (**Figure 1D**). To confirm that this metric of ITDv is ergodic, i.e., variability across time and trials is equivalent, ITDv was also computed 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, referred to as D-stat (**Figure 1E**)

$$Dstat = \frac{ITDrc}{ITDv}.$$

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(τ), used as predictors of ITD discriminability (**Figure 2A**), was extracted from Stern and Colburn (1978). To test whether ITD statistics could be represented by this model, the p(τ) parameter was adjusted to match statistics, by normalizing ITD statistics from 0 to 1 and scaling the resulting data to obtain a probability distribution where the sum of all probabilities was equal to 1 (**Figure 2B**).

Fisher information from single-cell IPD-tuning curves and the optimal population distribution of IPD tuning estimated for humans were extracted from data reported in Harper and McAlpine (2004). 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 converted to ITD, obtaining the prediction of ITD discriminability induced by the neuron distribution (**Figure 2C**). To test whether ITD statistics could be represented by parameters of this model, we 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 from signal traces using 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 data points of threshold dIPD across reference ITD and frequency from Mills (1958). Threshold dIPD was converted to threshold dITD (using the same equation described in HRTF analysis section) (**Figure 3A**, right). Spearman's correlation coefficients between dITD thresholds and ITD statistics (or ITD discriminability predicted by neural models) were computed across reference ITD and frequency.

Estimation of spatial ITD discriminability thresholds

A test was designed to estimate detection thresholds of changes in stimulus ITD (dITD) across specific frequencies of interest for this study. 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. There were no distinct groups in the experiment. All subjects had no reported history of neurological disorders or hearing impairments.

We used the statistical model comparison Akaike Information Criterion method (AIC; see below), which relies on both the number of model parameters and sample size (i.e., number of subjects) as a metric of goodness of fit. During the design of the experiments, sample sizes were made several times higher than the number of parameters of our models to ensure samples were sufficiently large. Pilot measurements of dITD thresholds were initially conducted in 10 subjects, using the same combination of frequencies and ITDs across subjects. Results from these measurements already showed that D-stat predicted dITD thresholds better than ITDrc and ITDv alone, for frequencies above 250 Hz. This pilot experiment, which lasted approximately 150 minutes, was performed in up to 5 sessions per subject.

Based on feedback from subjects undergoing pilot measurements, a new measurement of dITD threshold was conducted using a shorter protocol that lasted about 60 minutes, leading to the reported dITD results. The analysis reported in the manuscript was performed in the data across all 24 subjects.

In this computer-based test, subjects were presented with 65 dB (A scale) tones within a range of frequencies through headphones calibrated with an Instrutherm DEC-460 decibel meter or a 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 3C**, left).

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. An initial training block used 750 Hz tones and −50 μs reference ITD for all subjects and was not included in the analysis. The following 20 blocks presented tones of frequencies 250, 500, 750, 1000 or 1250 Hz, and reference ITDs from −500 to 0 varied in steps of 100 μs. Reference IPDs of absolute values smaller than $\pi/2$ radians were included, to avoid phase ambiguity. The sequence of conditions was randomly varied across subjects.

Ongoing estimation of dITD thresholds was conducted from 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.

For estimating threshold dITDs, an ROC classifier 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 3C**, right). 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 straightforward analysis of the ranked (Spearman's) correlation coefficients was performed for the relationship between the threshold dITD averaged across subjects and the ITD statistics of the reference ITD (**Figure 3D**, left). Spearman's correlations, which were computed from averages of dITD thresholds over subjects across multiple conditions (*N* number of combinations of reference ITD and frequency), were used to assess the monotonicity of this relationship. Since the *N* in this analysis reflects the number of conditions, not the number of subjects, the standard statistical power analysis does not apply. Accordingly, p-values were not computed for this correlation analysis.

Additional analysis for selection of best predictor of threshold dITD was performed using linear mixed-effect models (LMM; Magezi 2015), classifying ITD statistics (eg. ITDrc, D-stat) across conditions as 'fixed factors' and participant as 'random factor' (Figure 3E).

LMM analysis assumes linearity between measures and predictors, then we inspected whether the relationship between dITD thresholds and each of the predictors was linear. Although relationships were mostly linear, some were best fitted by a power function. In these cases, we applied the standard method for achieving linearity by log-transforming both predictors and dITD thresholds.

Linear regressions of the relationship between the multiple log-transformed dITD thresholds collected from each subject and log-transformed ITD statistics were performed, and AIC statistics computed. The AIC analysis 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).

A statistical power analysis for the MMN component using a stringent minimum MMN amplitude of -0.5 μV (SD 0.7 μV) revealed substantial power (1-β = 0.87) with an alpha level of 0.05 in 30 adult subjects. We performed the first set of conditions (1-10) in 17 subjects and found that D-stat predicted MMN peaks better than ITDrc and ITDv alone; 16 additional subjects were recruited for a second set of conditions (11-20), replicating the initial findings. No significant difference was found between groups and therefore the analysis reported in the manuscript was performed on the pooled data of all 33 participants. Participants were seated inside a sound-attenuated 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-weighted) using a B&K 4947 microphone with an artificial ear. Sound signals were 50-ms duration tones (25-ms rise-fall time) of different frequencies and ITDs synthesized with Matlab software.

Participants listened to oddball sequences presenting repetitive ("standard") tones embedded with sporadic ("deviant", 15% of the trials) tones with a 300 ms inter-stimulus interval (**Figure 4A**). Each subject was presented with 10 conditions, which differed in frequency and ITD. Subjects 1 to 17 performed the conditions 1 to 10, while subjects 18 to 33 performed the conditions 11 to 20 (**Figure 4B**). 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 of blocks was randomized for each subject. The first trials of each block (18 standards + 1 deviant) were used for training and not included in the analysis. Results reported included the following trials (385 standards and 70 deviants) for each subject. MMN values for each condition were estimated by subtracting mean ERP signals of 210 deviant trials by the mean of 1155 standard trials; after removal of trials with artifacts (see below). Sessions lasted approximately 2.5 hours, including placement of electrode caps and breaks during the EEG recording.

EEG was recorded with Neuroscan SynAmps and a 32-channel electrode cap following the modified international 10-20 System, including electrodes on the nose (reference), P09 (ground) 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 using a band-pass from 0.05 to 100 Hz. Impedances were maintained below 5 kOhms.

To measure the MMN, EEG signals from each subject were processed as follows: (1) 20-Hz low-pass filtering; (2) pooling (by concatenating) all EEG 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 signals from the FZ channel (frontal EEG electrode, in the middle of the forehead) averaged across subjects, for each condition; (10) measurement of MMN peak for each subject and condition within this time bin.

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 4C**).

Prediction of MMN by ITD statistics

An initial straightforward analysis of the ranked (Spearman) correlation coefficients was performed for the relationship between the MMN peak amplitude averaged across subjects and the absolute ITD difference between standard and deviant stimuli, normalized by the weighted sum of ITD statistics estimated for both standard and deviant stimuli (**Figure 4D**, left). Additional analyses with selection of best predictor of MMN peaks were performed with LMM and AIC (see "Prediction of spatial discriminability thresholds by ITD statistics" session) to compare performance across predictors (**Figure 4E**). The relationship between the MMN data and each of the predictors (ITD statistics or ITD discriminability as predicted by classic neural models) that followed a power function were linearized using log transformation. The conditions were treated as nested in the model selection analysis, since we fit a linear mixed-effects model for MMN peaks, with ITD statistics and participants as 'fixed' and 'random factors', respectively. The analysis of MMN values included all data collected. No outliers were detected or excluded. The AIC method was blind to the weight of standard vs weight of deviant in MMN data; different combinations of standard and deviant weights were treated as different models, rather than model parameters.

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 in https://doi.org/10.5061/dryad.h70rxwdf9.

Acknowledgements

This work was supported by the 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, Conception and design, Acquisition of data, 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 Peña, 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, K. E., and Delgutte, B. A physiologically based model of interaural time difference discrimination. *J. Neurosci*. **24,** 7110–7117 (2004).

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).

Magezi, D. A. Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (LMMgui). *Front. Psychol.* **22** (2015).

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, 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).