Anticipated ITD statistics built-in 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

Highlights

- Human HRTFs were used to estimate ITD statistics across frequency and azimuth.
- Spatial discrimination thresholds could be predicted by rate of change and ITD variability.
- Spatial deviants detection was weighted by anticipated ITD statistics.
- These results indicate that high-order sensory statistics are built-in human perceptual functions.

Abstract

The variability of natural scenes places perceptual processes in the realm of statistical inference, where sensory evidence must be weighted by its reliability. Absent prior information, estimating environmental variability requires real-time sampling and computations. However, a portion of environmental variability can be assumed invariant across conditions. Perceptual tasks relying on time-dependent information may be vastly enhanced if the invariant statistical structure of sensory cues is built into the underlying neural processing. We investigated this question in human sound localization, where the statistics of spatial cues can be estimated.

Localizing low frequency sounds in the horizontal plane relies on interaural time differences (ITD). We estimated the ITD statistics across frequency and azimuth from human head-related transfer functions (HRTFs). The mean ITD varied with azimuth following a sigmoid relationship, whose slope is steepest at the center. In addition, ITD was more variable over time for sounds located in the periphery compared to the center, in a frequency-dependent manner. We investigated the role of these statistics – ITD slope and ITD variability – in low-frequency lateralization of human subjects, to test the hypothesis that high-order sensory statistics are represented in the human brain influencing spatial discriminability and novelty detection.

Thresholds for discriminating ITD changes reported by classical studies (Mills, 1958) were predicted by a model that considered both ITD slope and ITD variability. To further test our hypothesis, EEG novelty responses were recorded in human subjects undergoing an oddball stimulation sequence, where repetitive ("standard") tones of a given ITD were combined with sporadic ("deviant") tones of a different ITD. By using insert earphones, ITD was shifted with zero variability across time and location. Mismatch negativity (MMN) brain signals were used as an index of discriminability between standard and deviant stimuli. We found that MMNs were weaker for standards in the periphery, where the ITD slope is lower and the ITD variability is higher. Overall, the amplitude of novelty EEG signals was predicted by the difference in ITD between the standard and deviant normalized by the anticipated discriminability of the standard location, indicating that change detection is weighted by expected statistics of the sensory input.

In sum, our results show that spatial discriminability thresholds and deviant detection are consistent with a representation of anticipated ITD statistics in the human brain, supporting the hypothesis that high-order statistics are built into human perceptual processes biasing behavior.

Keywords

Human sound localization; Sensory reliability; ITD; HRTF; JND; MMN; Novelty detection.

Introduction

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

Classical psychophysical studies demonstrated that humans detect sound location better in the front than in the periphery (Mills, 1958; Yost, 1974; Makous & Middlebrooks, 1990). The enhanced capability in frontal locations could be efficient for hunting and foraging, as analogously proposed for vision (Collins, 1922; Cartmill, 1974; Changizi & Shimojo, 2008). Physiological evidence indicates that coding of binaural spatial cues could support finer discriminability in the front (van Bergeijk, 1962; 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, 2017).

For azimuth detection based on ITD, sound diffraction by structures surrounding the ear modulates the ITD vs. azimuth relationship (Aaronson & Hartmann,

2014). In particular, because coincidence detector neurons compute ITD in narrow frequency bands (Yin and Chan 1990; Carr and Konisih 1990), the interaction of nearby frequencies may be an important source of ITD variability. Previous studies in barn owls have shown that variability of binaural cues is low in the front and high in the periphery, and varies across frequency (Cazettes et al., 2014; Fischer & Pena, 2016). Sensory variability is known to be fundamental in change detection. It has been shown that stimulus discrimination depends not only on the mean difference between standard and deviant stimuli but also on the variability of the sensory evidence (Green & Swets, 1966). There is evidence that variability of binaural cues is represented in the avian auditory system, determining sound-orienting behavior (Fischer & Pena 2011, 2016; Cazettes et al., 2016; 2017).

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

Results

ITD statistics (i.e., the derivative of the mean ITD over azimuth and the standard deviation of ITD) were estimated from human HRTFs (Listen HRTF database, 2003). We tested whether the ITD statistics predict human spatial discrimination thresholds (Mills, 1958). Next, we tested if natural ITD statistics are built-in spatial deviants detection using EEG and mismatch negativity signals (MMN).

ITD statistics from human HRTFs

Using the Listen HRTF database (http://recherche.ircam.fr/equipes/salles/listen; 51 subjects), we estimated the rate of change of ITD (ITD slope) (Figure 1A) and the ITD standard deviation over time (Figure 1B) at each azimuth location and frequency (see Methods). Azimuth discriminability carried by ITD, as proposed by Signal Detection Theory, was estimated by computing an index inspired by the d-prime, dividing ITD slope by ITD variability at each azimuth location (Figure 1C). From these data, frequencies and azimuth ranges of interest were selected for subsequent analyses and experimental design.

Prediction of spatial discrimination thresholds from ITD statistics

ITD just-noticeable difference (JND) as a function of azimuth and ITD was estimated from classic published data on human sound localization (Mills, 1958; Fig 2A,B).

We tested whether ITD slope, ITD variability or their combination (ITD d-prime) predicted the JND. Figure 2C shows the JND as a function of ITD slope, ITD variability and ITD d-prime across the full frequency and ITD ranges (Fig 2C, top), for the ITD range within the pi-limit (removing the phase-ambiguous range; Fig 2C, middle) and for the low frequency range (Fig 2C, bottom). JND was significantly correlated with ITD slope (Fig 2C left column) and ITD variability (Fig 2C middle column) but the highest correlation coefficients were found for ITD d-prime (Figure 2C right column). Correlation coefficients for each plot are shown in Fig. 2D. These results suggest that spatial discrimination thresholds are determined by expected ITD statistics, and therefore a brain representation of these statistics exists.

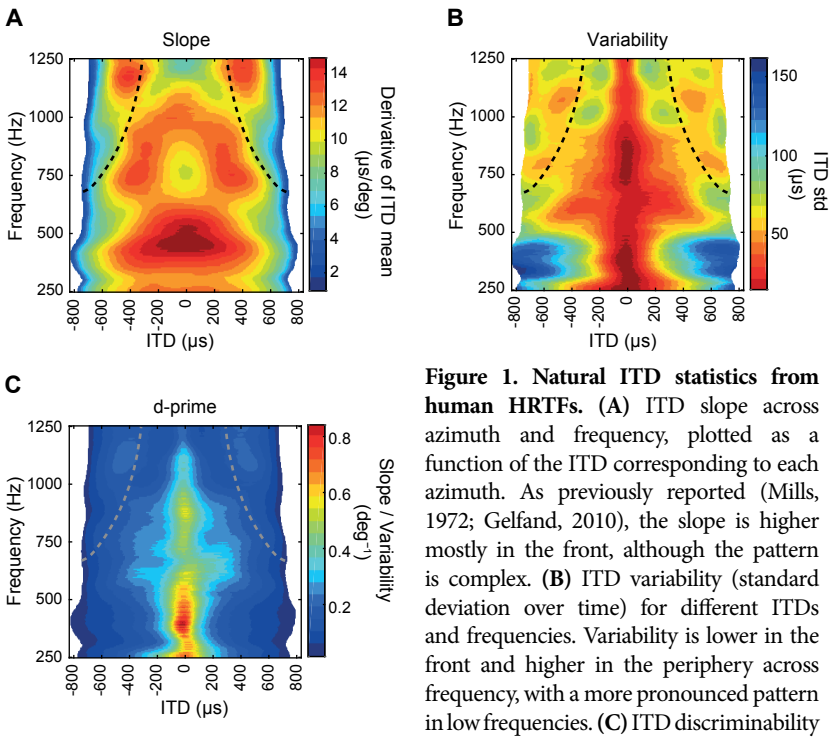

Figure 1. Natural ITD statistics from human HRTFs. (A) ITD slope across azimuth and frequency, plotted as a function of the ITD corresponding to each azimuth. As previously reported (Mills, 1972; Gelfand, 2010), the slope is higher mostly in the front, although the pattern is complex. **(B)** ITD variability (standard deviation over time) for different ITDs and frequencies. Variability is lower in the front and higher in the periphery across frequency, with a more pronounced pattern in low frequencies. **(C)** ITD discriminability (d-prime). Dashed lines indicate the pi-limit, beyond which ITD information becomes ambiguous within narrow frequency bands.

Built-in ITD statistics in deviant detection

To further test if statistics are built-in the human brain processing, we recorded EEG mismatch negativity responses (Näätänen et al., 1978) for spatial deviants in tones presented through earphones. The inserted earphones removed the natural ITD variability induced by the filtering effect of the head and ears. Thus, differences in deviant detection could be assigned to expected rather than current stimulus statistics. MMNs allowed for a quantitative indexing of the effect of reliability on either

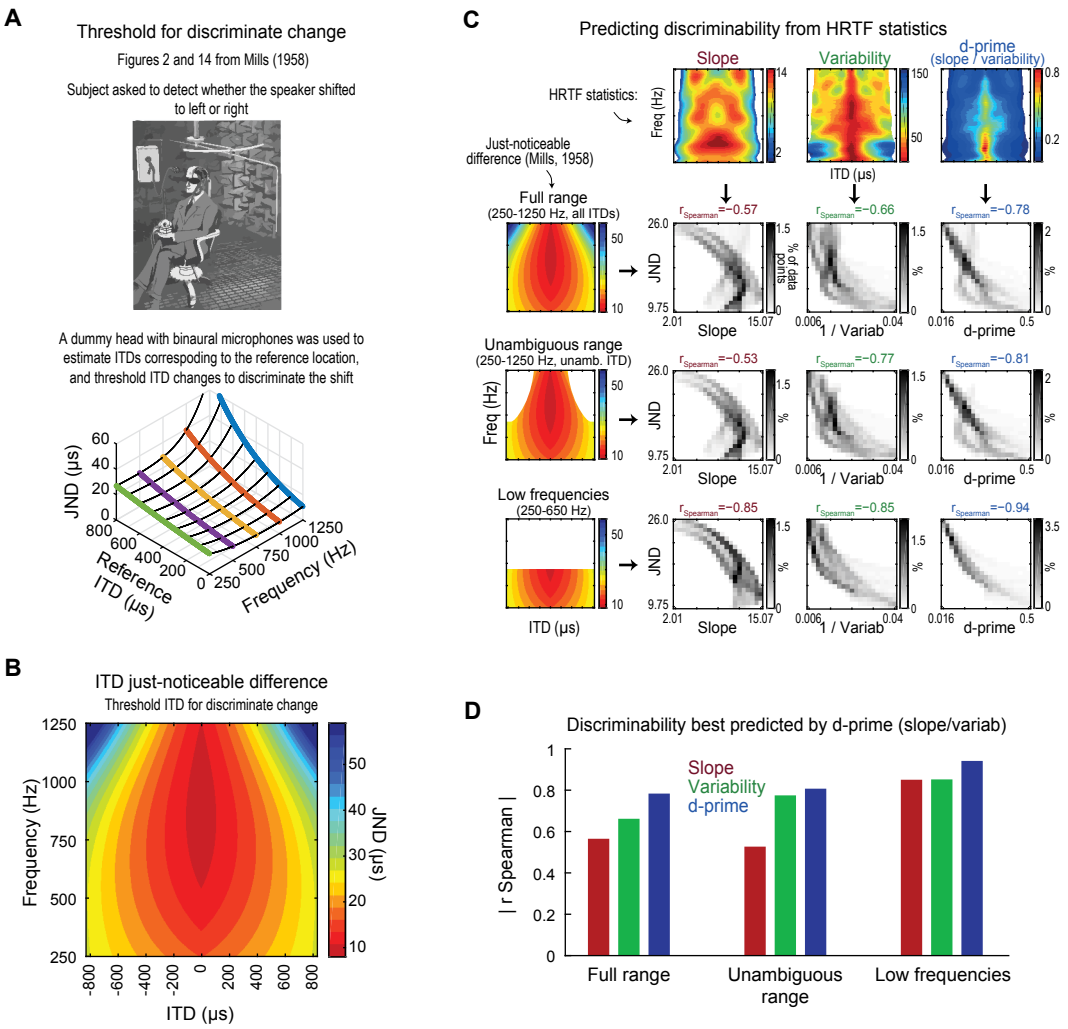

Figure 2. Human detection thresholds of ITD changes match expected ITD statistics. (A) Mills (1958) estimated the minimum ITD change (JND) that could be detected across frequency and location *(top)*. Azimuth was converted to ITD using a dummy head with binaural microphones *(bottom)*. Reproduced from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

standard and deviant stimuli. In addition, the pre-attentive nature of MMN signals did not require training subjects to perform behavioral tasks, which might introduce confounding effects.

MMNs are observable when a sequence of standard stimuli is unexpectedly interrupted by a low probability deviant stimulus (Näätänen et al., 1978; Pakarinen et al., 2007; Sussman, 2007; Sussman et al., 2014; Figure 3A). The MMN signal is obtained by subtracting event-related potentials (ERPs) evoked by deviant stimuli from ERPs evoked by the standard. The larger the separation between standard and deviant stimuli, the more negative the MMNs, which has led to consider the MMN amplitude an index of discriminability between standard and deviant (Sams et al., 1985; Tiitinen et al., 1994; Pakarinen et al., 2007).

We measured MMN responses in 33 subjects. A set of ITD and frequency conditions was selected in order to sample critical ranges drawn from the HRTF analysis (Figure 3B). MMN signals were measured across subjects and conditions. The data showed the characteristic frontal topography of MMN responses (Figure 3C; Giard et al., 1990). The peak amplitude of MMN was used to quantify the effect of stimulus manipulations on novelty detection.

The MMN peaks averaged across subjects showed weak correlation to the ITD difference (absolute difference between the ITDs of standard and deviant stimuli) (Figure 3D). However, a stronger correlation emerged between the ITD difference and MMN negativity when the ITD of the standards was frontal (Figure 3D, right). Consistently, standard stimuli located in the front led to more negative MMNs than deviants in the front (Figure 3E). This comparison was performed between pairs of standard and deviants of exactly the same frequency and ITDs, where only the probability of either stimulus was switched. These results indicate that MMNs were mainly elicited when the

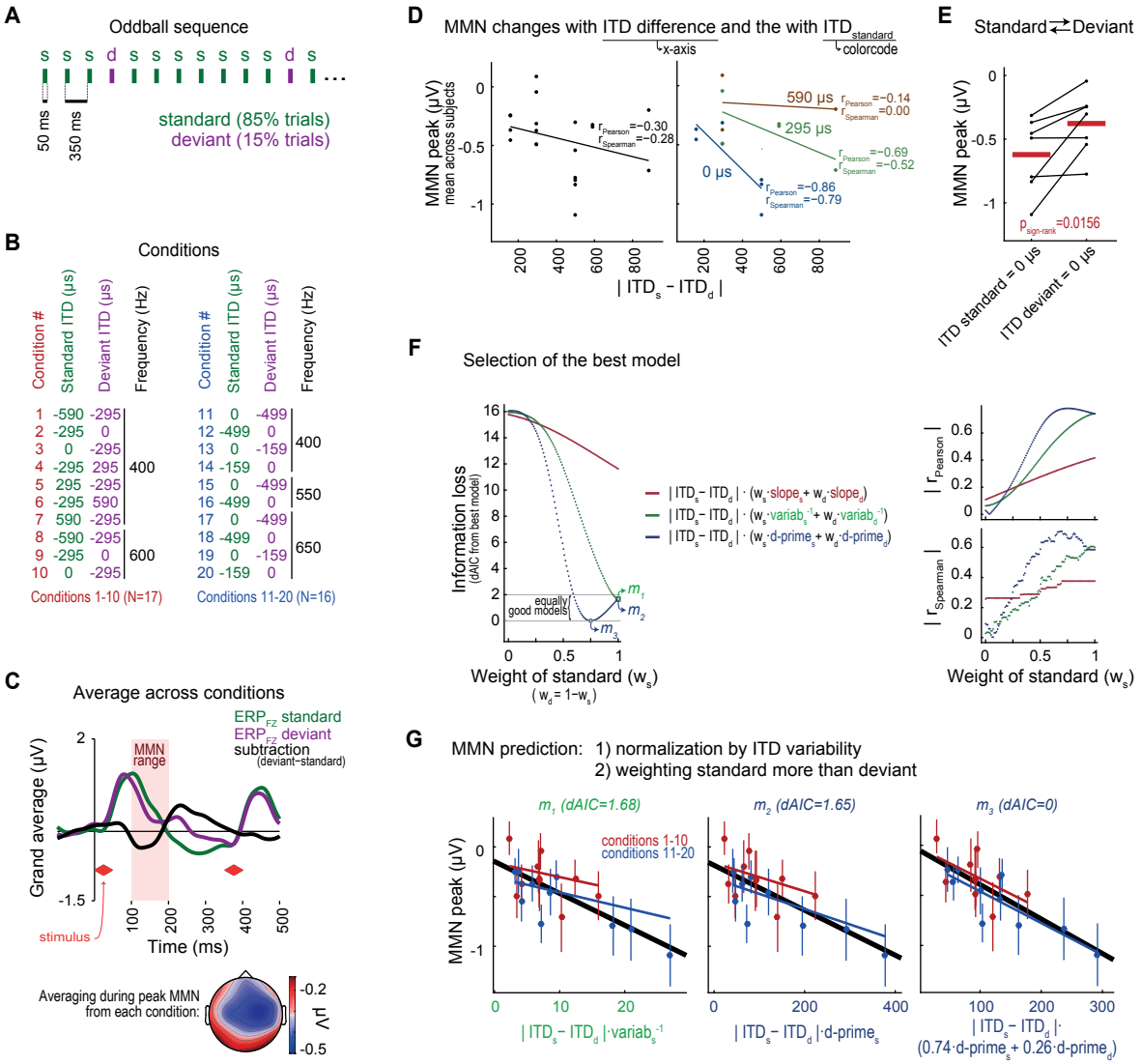

Figure 3. Discriminability of spatial deviants indexed by MMN responses is weighted by expected ITD variability at the standard location. (A) Passive oddball sequence protocol, in which subjects listened to frequent "standard" stimuli embedded with rare "deviants." **(B)** In each condition, two tones of same frequency and distinct ITDs were presented. Two sets of 10 conditions were tested in two different groups of subjects. **(C)** Standard *(green)* and deviant *(purple)* event related potentials (ERP), recorded at the midline frontal electrode (FZ), were averaged across conditions and subjects. Note the negative deflection of the subtraction (deviant-standard) ERP within the 100-200 ms range *(black)*. MMN signals are characteristically more negative within frontal locations. **(D)** While MMN amplitude was weakly correlated to ITD difference when all data were included *(left)*, classifying the data by the ITD of the standard (0 µs, 295 µs and 590 µs), exposed different correlation levels *(right)*. In particular, the stronger correlation occurred when the standard was frontal (0 µs, *blue*). **(E)** Paired comparison interchanging the ITD of the standard and deviant showed significantly more negative MMNs when the standard was in the front (Wilcoxon signed rank test). **(F)** AIC statistical test for selecting the model with best predictive power of MMN amplitude. The models normalized ITD difference by anticipated ITD slope and/or ITD variability, weighting standard (*Ws*) and deviant (*Wd*) differently (left). The good models show also high correlation coefficients with MMN peak *(right)*. **(G)** Prediction of MMN amplitude by three models where ITD difference was normalized by ITD variability of the standard alone *(m1, left)*, ITD d-prime of standard alone *(m2, middle)* and d-prime of standard and deviant with differential weights *(m3, right)*. The best of these models (dAIC=0) was *m3*. dAIC indices are shown on top of each plot.

standard stimuli were located in the front and that deviant detection diminished for standards in the periphery.

Finally, we selected the statistical model that best described the MMN amplitude. Three groups of models were tested, which normalized the ITD difference between standard and deviant by the ITD slope (Figure 3F, red), ITD variability (Figure 3F, green) or ITD d-prime (Figure 3F, blue). For each model, the effect of changing the relative weights of standard and deviant was examined. We found that normalization by ITD variability was required for a good predictive power. Interestingly, models that considered only the ITD statistics of the standard performed better (model "m1" and "m2"; Figure 3G, left and middle). Specifically, the best prediction of MMN peak was obtained by normalizing the ITD difference between standard and deviant by the ITD d-prime, and assigning 74% weight to the standard and 26% to the deviant (model "m3"; Figure 3G, right). The two sets of subjects, represented in red and blue, showed consistent results, indicating that these findings were robust across conditions.

Discussion

Human HRTFs were used to estimate the ITD statistics across subjects, namely the rate of change of ITD over azimuth (ITD slope) and standard deviation of ITD over time (ITD variability), across frequency and ITD. ITD slope and ITD variability, and their ratio (ITD d-prime) were used to test whether natural statistics are built-in the brain processing underlying human sound localization.

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 & Pena, 2016). These studies provided evidence of how sensory reliability could be represented (Fischer & Pena, 2011, 2016; Rich et al., 2015; Cazettes et al., 2016) and integrated into an adaptive behavioral command (Cazettes et al., 2017). Although the coding observed in owls cannot be assumed for all species, the present study shows that ITD statistics also affect human sound localization.

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, 1972; Gelfand, 2017). This interpretation does not consider ITD variability. Here we show that natural ITD variability improves the prediction of these thresholds. The correlation between discriminability and each of the ITD statistics shows that the ratio between ITD slope and ITD variability is the best predictor of discriminability thresholds. These results support the idea that the brain anticipates, and therefore represents, the spatial discriminability as a function of frequency and location.

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

Novelty detection is weighted by anticipated ITD statistics

The novelty detection studies conducted with insert earphones were designed to more directly approach the question of whether anticipated ITD variability is built-in for spatial perception. Consistent with observations on the discriminability thresholds reported in Mills (1958), the ratio between ITD slope and ITD variability (d-prime) was a better predictor of deviant detection responses. Although ITD-change detection thresholds, as probed in Mills (1958), and our novelty detection paradigm required discriminating a change in ITD, the latter also involves identification of a repetitive (standard) pattern. The internal representation of standard patterns may be optimized by integrating predicted variability. Thus, anticipated ITD variability may modulate novelty detection both at the ITD-change and pattern detection levels. Consistently, the models that best described the MMN signals weighted the ITD difference between standard and deviant primarily by the ITD statistics of the standard location. Detecting ITD-changes in a reference location, as in Mills (1958), is influenced by attention and training; the novelty detection protocol allowed for control of these factors. In addition, the novelty detection paradigm showed that the prediction is based mostly on the ITD statistics of the standard stimulus, consistent with the interpretation that anticipated ITD variability is critical for pattern detection.

In sum, we found evidence that anticipated ITD statistics modulate auditory spatial perception. The ITD statistics used in this study are a particular aspect of the statistics of sounds reaching the human ears. The statistics of the natural auditory scene could be learned through individual experience. However, the consistency of these statistics across subjects indicates that this information may be invariant enough to be genetically encoded and preserved. Thus, the ability to predict these statistics by a built-in representation is a potentially adaptive evolutionary mechanism for approaching optimal performance. This study provides evidence supporting the idea that high-order statistics of sensory information relevant to sound localization are built into the human brain processing.

Methods

HRTF measurement

The dataset used in this study consisted of head-related impulse responses recorded 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 recordings were 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 recorded with a pair of small Knowles FG3329 microphones, calibrated using a B&K 4149 microphone. These small microphones were mounted in a silicon structure which occluded the ear canals of the subjects, avoiding resonance and placing the microphone sensor at the entrance of the ear canal. The signal captured by the microphones was driven to a custom amplifier with 40 dB gain, recorded through a RME sound card interface with Max/MSP real time library which deconvolved the microphone signal.

HRTF analysis

Data from 51 subjects were included in the HRTF analysis. Recordings corresponding to 0-degree in elevation and azimuth from -90 to 90 degrees in 15-degree steps were selected. We convolved the recorded sound 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 150 to 1250 Hz with 5 Hz steps, and bandwidth equivalent to the estimated bandwidth for human cochlear filters (Glasberg & Moore, 1990). We used the gammatone filter bank from Malcolm Slaney's Auditory Toolbox, available in https://engineering.purdue.edu/~malcolm/interval/1998-010). The filtered signal was represented in the Hilbert space for extracting its instantaneous phase, using the Signal Processing Toolbox (Mathworks). For each frequency and azimuth, we calculated the circular mean and standard deviation of the interaural phase difference (IPD, in radians), which was converted to ITD (in μs) by the following equation: $ITD=10^6 IPD(2\pi Freq)^{-1}$.

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

We computed the median across subjects of the ITD variability for each azimuth and frequency sampled. These data were then interpolated across azimuth and represented as a function of ITDs (Figure 1B). Although both ITD slope and ITD variability are likely affected by diffraction and interference between neighboring frequencies, their correlation was not strong (rPearson=-0.41; rSpearman=-0.42; data not shown), indicating that they are partially independent. The combined effect of slope and variability was assessed by their ratio, a metric consistent with the Signal Detection Theory's d-prime (difference in means divided by square root of mean variance).This ratio was thus referred to as ITD d-prime (Figure 1C).

Estimation of spatial discriminability thresholds

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

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

For the current study, we extracted the scales, intercept and slopes from the linear fit of threshold IPD against reference ITD across frequency of Mills (1958). The first conversion we did was threshold IPD to threshold ITD, followed by interpolation using the cubic spline method. This conversion was also performed by Mills (1958) and allowed us to confirm that our value extraction was correct (Figure 2A, bottom). Finally, we represented the JND of ITD (an inverse measure of discriminability measure) in the same format in which we represented the HRTF metrics (Figure 2B).

Prediction of spatial discriminability thresholds by ITD statistics

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

Collection and analysis of Mismatch Negativity (MMN) responses

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

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

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

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

To measure the MMN, EEG signals from each subject were processed as follows: (1) 20-Hertz low-pass filtering; (2) Pooling signals obtained during sound stimulation; (3) removal of eye-blink and saccade artifacts by performing Independent Component Analysis and reconstructing the signal without components correlated to the EOG; (4) selection of 600-millisecond epochs around sound presentation (-100 to 500 ms from sound onset); (5) removal of linear portions of EEG signals and exclusion of trials containing voltage deflections larger than 75 mV; (6) re-reference of the EEG recordings to the mastoid recordings by subtracting the average mastoid signal

from the other channels; (7) ERP measurement, averaging separately signals evoked by standard and deviant trials (the first 19 trials were used for subject training and excluded from the analysis); (8) subtraction of standard from the deviant ERPs to compute the MMN; (9) Identification of time bin containing of the MMN peak in average signals from the FZ channel across subjects, for each condition; (10) measurement of MMN peak for each subject and condition within this time bin.

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

Prediction of MMN by ITD statistics

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

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

All data processing was performed in Matlab (Mathworks) using built-in or custom made routines.

Acknowledgements

This work was supported by NIH BRAIN Initiative (Grant NS104911) and by the NIDCD (Grant R01 DC004263). We thank Ruben Coen Cagli for comments on the manuscript.

Author contributions

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

Ethics

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

References

Aaronson, N. L. & Hartmann, W. M. Testing, correcting, and extending the Woodworth model for interaural time difference. *The Journal of the Acoustical Society of America* **135,** 817–823 (2014).

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

Cartmill, M. Pads and claws in arboreal locomotion. *Primate Locomotion* 45–83 (1974).

Cazettes, F., Fischer, B. J. & Peña, J. L. Cue reliability represented in the shape of tuning curves in the owl's sound localization system. *Journal of Neuroscience* **36,** 2101–2110 (2016).

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

Cazettes, F., Fischer, B. J., Beckert, M. V. & Pena, J. L. Emergence of an adaptive command for orienting behavior in premotor brainstem neurons of barn owls. *bioRxiv* 119180 (2017).

Changizi, M. A. & Shimojo, S. "X-ray vision" and the evolution of forward-facing eyes. *Journal of Theoretical Biology* **254,** 756–767 (2008).

Collins, E. T. *Arboreal Life and the Evolution of the Human Eye.* (Lea & Febiger Philadelphia, 1922).

Feddersen, W., Sandel, T., Teas, D. & Jeffress, L. Localization of high-frequency tones. *the Journal of the Acoustical Society of America* **29,** 988–991 (1957).

Fischer, B. J. & Peña, J. L. Optimal nonlinear cue integration for sound localization. *Journal of Computational Neuroscience* **42,** 37–52 (2016).

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

Gelfand, S. A. *Hearing: An Introduction to Psychological and Physiological Acoustics*. (CRC Press, 2017).

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

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

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

Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics*. (John Wiley & Sons, New York, 1966).

Grothe, B., Pecka, M. & McAlpine, D. Mechanisms of sound localization in mammals. *Physiological Reviews* **90,** 983–1012 (2010).

Makous, J. C. & Middlebrooks, J. C. Two-dimensional sound localization by human listeners. *The Journal of the Acoustical Society of America* **87,** 2188–2200 (1990).

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

Middlebrooks, J. C. & Green, D. M. Sound localization by human listeners. *Annual Review of Psychology* **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. *The Journal of the Acoustical Society of America* **30,** 237–246 (1958).

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

Pakarinen, S., Takegata, R., Rinne, T., Huotilainen, M. & Näätänen, R. Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP). *Clinical Neurophysiology* **118,** 177–185 (2007).

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

Sams, M., Paavilainen, P., Alho, K. & Näätänen, R. Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* **62,** 437–448 (1985).

Sussman, E. S. A new view on the MMN and attention debate. *Journal of Psychophysiology* **21,** 164 (2007).

Sussman, E. S., Chen, S., Sussman-Fort, J. & Dinces, E. The five myths of MMN: redefining how to use MMN in basic and clinical research. *Brain Topography* **27,** 553–564 (2014).
Tiitinen, H., May, P., Reinikainen, K. & Näätänen, R. Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature* **372,** 90 (1994).
van Bergeijk, W. A. Variation on a theme of Bekesy: a model of binaural interaction. *The Journal of the Acoustical Society of America* **34,** 1431–1437 (1962).
Woodworth, R. *Experimental Psychology.* (New York: Holt, 1938).
Yin, T. & Chan, J. Interaural time sensitivity in medial superior olive of cat. *Journal of Neurophysiology* **64,** 465–488 (1990).
Yost, W. A. Discriminations of interaural phase differences. *The Journal of the Acoustical Society of America* **55,** 1299–1303 (1974).