On the variation of interaural time differences with frequency

Victor Benichoux[a)b)] and Romain Brette[a)]

Institut de la Vision,

(INSERM U968,

CNRS UMR 7210,

UMR S 968)

17,

rue Moreau

75012 Paris,

France

Marc Rébillat

PIMM,

Arts et Métiers ParisTech - CNRS - CNAM,

151,

Boulevard de l'Hopital

75013 Paris,

France

(Dated: February 15, 2016)

[a)] Also at: Institut d'Etudes de la Cognition, Ecole Normale Supérieure, Paris, France

Abstract

Interaural time difference (ITD) is a major cue to sound localization in humans and animals. For a given subject and position in space, ITD depends on frequency. This variation is analyzed here using an HRTF database collected from the literature and comprising human HRTFs from 130 subjects and animal HRTFs from six specimens of different species. For humans, the ITD is found to vary with frequency, in a way that shows consistent differences with respect to a spherical head model. Maximal ITD values were found to be about 800 μs in low frequencies and 600 μs in high frequencies. The ITD variation with frequency (up to 200 μs for some positions) occurs within the frequency range where ITD is used to judge the lateral position of a sound source. In addition, ITD varies substantially within the bandwidth of a single auditory filter, leading to systematic differences between envelope and fine-structure ITDs. Because the frequency-dependent pattern of ITD does not display spherical symmetries, it potentially provides cues to elevation and resolves front/back confusion. The fact that the relation between position and ITDs strongly depends on the sound's spectrum in turn suggests that humans and animals make use of this relationship for the localization of sounds.

19 PACS numbers: 4366Qp, 4366Pn, 4380Lb

I. INTRODUCTION

In humans and many animals, a major cue to localize sounds in the horizontal plane is the difference in time between the peaks and valleys of the acoustical wave at the two ears, i.e. the interaural time difference (ITD). Remarkably, humans can distinguish ITD differences as low as 20 μs for a wide array of sound spectra and envelope characteristics below about 1.5 kHz (Mills, 1958; Brughera *et al.*, 2013). Furthermore, ITD has been shown to have a stronger influence than interaural level differences (ILD) on the perceived lateral location of sounds with energy in low frequencies (below 2.5kHz, (Wightman and Kistler, 1992)). In animals as well, ITD is used as a cue to sound location: cats, gerbils, birds, reptiles and others have dedicated neural structures in the auditory brainstem to process ITD (Grothe *et al.*, 2010). Understanding the neural processing of ITDs requires a precise knowledge of the nature of the temporal disparities imposed by the head, body and environment.

The relationship between source position and ITD is constrained by morphological parameters including the interaural distance, head, ear position, shape, torso, and (even) hair (Duda *et al.*, 1999; Algazi *et al.*, 2001b). Measuring ITDs for tones using a manikin, Kuhn (1997) found that the ITD also varies systematically with the frequency of the tone, as reproduced in Fig. 1a. The ITD for a 2000 Hz tone presented at 45° is 400 μs (Fig. 1a, dashed line), while the ITD for a 500 Hz tone at the same position is 600 μs — about 50% larger. Conversely, sounds presented at different positions can produce the same ITD, provided they have different frequency contents. It follows then that spatial position cannot be estimated from ITD independently of sound frequency. This physical phenomenon is known, and has been observed in models where the head is a rigid sphere (Kuhn, 1977), or an ellipsoid (Cai *et al.*, 2015). As an illustration we computed ITDs from the spherical head model (details below) using a head radius of 9.3 cm (as reported in (Kuhn, 1977)), and plotted it on Fig. 1b.

Despite these early insights, the dependence of ITD on frequency in humans and animals has not, to our knowledge, been quantitatively examined. We propose here to bridge that

[b)]Electronic address: victor.benichoux@ucdenver.edu

gap by a careful and in-depth assessment of the frequency-dependence of ITD in human and animal acoustical data. Furthermore, we provide a new interpretation of this dependence, showing that it results in different ITDs in the envelope and fine-structure of auditory filters' outputs.

After a review of the physics of the frequency-dependence of ITD using previous reports of the spherical model (section II), we provide a quantitative analysis of this phenomenon in acoustical recordings of 130 subjects from four databases (section III). Further, we show that ITDs between the envelopes and fine-structures of binaural signals are different (section IV). Finally, we analyze the frequency-dependence of ITDs in the HRTFs of six different animal species (section V), and show that the highly non-spherical nature of many animals' heads makes it hard to predict the range of ITD from head size.

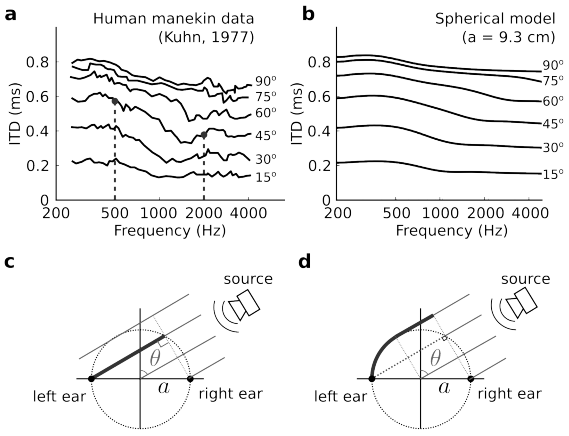

Figure 1: Frequency-dependence of ITD. (a) ITD measured with pure tones of varying frequency for different source positions on a human manikin (replotted from Kuhn (1977)). (b) ITD computed for a spherical head model with head radius 9.3 cm. (c) Propagation of a planar sound wave with an acoustically transparent head. The additional pathlength to the contralateral ear (thick line) is a sine function of the azimuth angle θ. (d) Propagation of a high frequence planar sound wave diffracted by a sphere. The additional path to the contralateral ear is the thick line.

II. SCATTERING AND FREQUENCY-DEPENDENT ITDS

A. Phase ITD

The complete acoustical transformation occurring between a point source and the ears of a subject can be modeled as a pair of filters (H_L, H_R) usually termed Head Related Transfer Functions (HRTFs). The *phase ITD* ($\mathrm{ITD_p}$, as reported in the original (Kuhn, 1977) study), is the ITD of the fine-structure of the binaural signal, defined at any frequency f using the phase response of the HRTFs:

$$\mathrm{ITD_p}(f,\theta,\phi) = \frac{1}{2\pi f}\angle\left[\frac{H_L(f,\theta,\phi)}{H_R(f,\theta,\phi)}\right] \tag{1}$$

where θ is azimuth measured in radians, ϕ is elevation (in standard vertical-polar coordinates: azimuth between -180° and 180°, and elevation between -90° (down) and 90° (up)). The bracket operator $\angle$ is the *unwrapped phase* operator that yields a continuous phase spectrum (not constrained to $[-\pi, \pi]$). By convention, positive azimuth values indicate that the source is to the right of the subject, where $\mathrm{ITD_p}$ is positive.

B. ITD in the spherical head model

In a first, simplified geometrical model of ITD, we can consider a planar acoustical wave incident on an acoustically transparent head (Figure 1c). In this case, the ITD is the difference in path lengths to the two ears (thick line) divided by the speed of sound: $\mathrm{ITD}(\theta) = 2\frac{a}{c}\sin(\theta)$, where θ is the azimuth of the sound source, c the speed of sound in air and a the head radius. In this description, ITD does not depend on frequency: ITD does show frequency dependence because the head is not acoustically transparent.

A more plausible acoustical model of the situation is to assume that the head is a rigid sphere, with ears lying on a diameter. The first observation that ITD is frequency-dependent in this context is attributed to Lord Rayleigh's spherical head model (Rayleigh and Lodge, 1904). Many studies have then used this model to analyze binaural cues (Duda and Martens, 1998; Kuhn, 1977). In particular, the phase ITD, i.e. $\mathrm{ITD_p}$ as defined in Eq. 1, can be

numerically calculated at all frequencies for the spherical head model, as shown on Fig. 1b (for sources at infinite distance from the head, following Bruneau (2010)). In this model, $\mathrm{ITD_p}$ for any given source position generally decreases with increasing frequency (Fig. 1b), which is broadly consistent with the human manikin data reported by Kuhn (1977) and shown in Fig. 1a.

For high frequencies, when the wavelength is small compared to the head radius, the ITD tends to the difference between the shortest path lengths to the two ears divided by the speed of sound (thick line in Figure 1c), which is expressed in Woodworth's formula:

$$\mathrm{ITD_{HF}}(\theta) = \frac{a}{c}(\sin(\theta) + \theta) \tag{2}$$

The low-frequency limit of ITD can be calculated by considering the first terms in the spherical-harmonic development of the acoustical field solution (Kuhn, 1977):

$$\mathrm{ITD_{LF}}(\theta) \approx 3\frac{a}{c}\sin(\theta) \tag{3}$$

The ratio between low and high frequency ITD is then: $\frac{3\sin(\theta)}{\theta+\sin\theta}$, which is always greater than one. Thus the the low-frequency ITD is always greater than the high frequency ITD. For azimuths θ between 0 and $\pi/2$ radians (90°), this ratio is a monotonically decreasing function of θ. For sources near 0°, the ITD is 50% larger at high frequency relative to low frequency, but for those positions (close to the midline) ITD values are close to zero. Conversely, when the ITD is maximal for azimuths near 90°, the low frequency ITD is only about 16% larger than the high frequency ITD. Readers should note that this is hard to see on 1b.

C. Visualization of the scattering phenomenon

At low frequency, the head is small compared the wavelength of the sound, and one might infer that the ITD should be close to the situation when the head is acoustically transparent (Fig. 1c). This would predict a smaller ITD than in high frequency, yet the opposite occurs. The reason for this counter-intuitive phenomenon is that the sphere is not an obstacle between the source and the ears, but rather the ears are *on* the sphere and diffraction phenomena are at play (Kuhn, 1977).

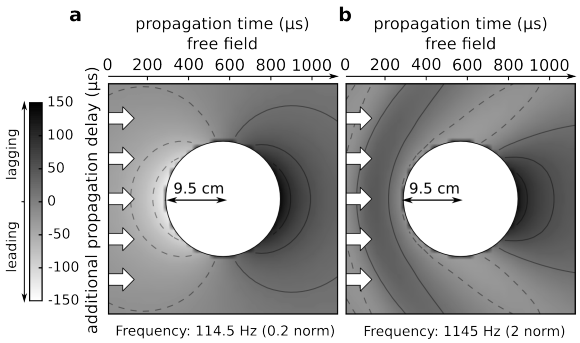

Figure 2: Propagation time of a planar sound wave in the presence of a sphere, relative to the propagation time in free field, for tone frequencies 114.5 Hz (a) and 1145 Hz (b). Propagation time in free field (no head) is shown on top. Negative values (lighter shades) indicate regions where phase is leading, and positive values (darker shades) indicate regions where the sound phase is lagging.

To get a better grasp of the situation, we calculated and represented the acoustical field in the vicinity of the head, using the formula in Bruneau (2010, paragraph 5.2.3). In Fig. 2, we show the steady-state propagation time of a monochromatic planar wave emanating from an infinite-distance source to the left ($\theta = 90°$), relative to the propagation time for an acoustically transparent head (free field). For head radius of 9.5 cm, the free field ITD is about 550 μs. For a low frequency tone (about 115 Hz, Fig. 2a), a phase lead appears on the ipsilateral side of the head (the acoustical wave is "compressed" against the head), and a phase lag appears on the contralateral side (the wave must turn around the head). As a result, the ITD is $550 + 150 + 150 = 850\mu$s. For a high-frequency tone (1145 Hz, Fig. 2b), the propagation time to the ipsilateral ear (left) is not affected by the head but there is still a phase lag at the contralateral ear, corresponding to the additional path length. As a result, the ITD is $550 + 150 = 700\mu$s, which is smaller than at low frequency.

The physical phenomenon is entirely specified by the wavelength ($\lambda = c/f$) of the acoustical wave relative to the size of the sphere. To account for the effect of the size of the head, it is thus convenient to introduce a normalized frequency scale, where unit normalized

frequency ($f_{norm} = 1$) corresponds to the physical frequency of a wave with wavelength equal to one sphere circumference: $f = c/2\pi a$. Scaling head size also scales ITD (it is inversely related), and therefore ITD can also be normalized, so that a scaled ITD of unity ($ITD_n = 1$) corresponds to a physical ITD of a/c, the time for sound to propagate one radius of the spherical head. Assuming a head radius of $a = 9.5$ cm, the low-frequency condition of Fig. 2 corresponds to a normalized frequency of 0.2, and the high-frequency condition to a normalized frequency of 2.0.

From this description, in terms of normalized time and frequency, the effect of changing the head size on $\mathrm{ITD_p}(f)$ is easier to grasp. It is two-fold: for a given normalized frequency, the ITD depends linearly on head size; and the frequency scale on which ITD varies depends linearly on the inverse of head size. In particular, the transition between the low and high frequency regimes occurs at lower frequencies for bigger head sizes.

III. FREQUENCY-DEPENDENCE OF ITD IN HUMAN HRTFS

Human head morphology is more complex than a sphere, and other parts of the human body also influence the ITD (Kuhn, 1977). Furthermore, comparing the human-manikin ITDs to the spherical-head ITDs in Fig. 1a-b reveals that ITDs exhibit a more complex frequency-dependence in humans than in the spherical head model. Therefore it is necessary to analyze the frequency-dependence of ITD from HRTF measured in real human subjects.

A. HRTF databases

HRTF data were obtained from three publicly available databases (ARI (ARI, 2010), CIPIC (Algazi *et al.*, 2001a) and LISTEN-V1 (Warusfel, 2002)). Another set of data was specifically recorded for this project (LISTEN-V2), following the protocol of the LISTEN-V1 database, and has not been made public yet. All data are available from the corresponding author on request. Overall, this combined dataset includes 130 subjects (Table I). Because these databases were obtained in slightly different conditions (in particular spatial measure-

ment grids and number of samples), ITDs were evaluated separately in each database, and then all statistics were interpolated on a common space-frequency grid (that of LISTEN-V2) using a natural neighbor interpolation. Results are therefore always presented with a spatial resolution of 5° in azimuth and about 15° elevation.

B. Frequency dependence of human ITDs

1. Acoustical head radius estimation and normalization

As discussed above, the subject's head size affects ITD across positions and frequencies in a way that is completely predicted by the acoustics of sound propagation. We are not, however, interested in the variability of ITD cues across the population of subjects that is explained by head size. Rather, we are interested in how this variability reflects the variability in head morphologies. Furthermore, increasing headsize systematically shrinks the frequency axis, thus averaging different subjects' ITD at each frequency will spuriously smooth out ITD variations. Therefore, in order to account for the effect of head size across the population of subjects, we normalize the time and frequency axes of each subject's ITD data using a measure of the head of the subject derived from its HRTFs (usually termed *acoustical head radius*, e.g. in (Algazi *et al.*, 2001b)). Similar normalization methods have been proposed in the context of reducing variability in spectral notches position and amplitude (Middlebrooks, 1999).

We define the "acoustical head radius" for each subject as the radius of the spherical model which best matches the subject's high-frequency ITD. More precisely, for each subject we compute the high frequency ITD at all positions as the average phase ITD between 3kHz and 5kHz. We compute the spherical-model high-frequency ITDs in the exact same way and for the same positions. The radius of the sphere is then adjusted so as to minimize the squared differences between the subject's ITDs and the sphere's. This regression is performed using a standard gradient-descent algorithm. The best-fitting sphere radius resulting from this procedure is the *acoustical head radius* of the subject. We validated this method by

simulating HRTFs using the spherical head model, and recovering the radius of the simulated spherical head. As expected, we found that this method estimated accurately the radius of the sphere within 0.01 mm (over a range of sphere radii from 5cm to 15 cm).

We then estimated the acoustical head radius of each subject in the population. We found the average head radius over all subjects to be 9.5 cm (N=130, $\pm$ 0.48 cm standard deviation, 8.3 to 10.8 cm range). For each subject, we computed the mean squared error between ITDs derived from the optimal spherical model and acoustically measured ITDs. The mean squared error was on average 67 μs $\pm$ 22 μs STD across the population, indicative of a consistently good fit. The average head radius value we found is slightly higher than the one reported in the (Algazi *et al.*, 2001b) study, which was obtained using another ITD estimation method (onset-time differences), and only on positions on the horizontal plane. The same estimation and fitting method on our data yields a radius of 8.43 cm ($\pm$ 1.14 cm), consistent with the value reported by (Algazi *et al.*, 2001b).

The goal of this normalization procedure was to account for the systematic effect of head size on the variability of the ITD measures across the population. We computed the standard deviation of ITD at each frequency and position across the population of subjects before and after normalization. The average standard deviation before normalization was 54 μs and reduced by about 20% with normalization (42 μs STD, using the average head size of 9.5 cm to convert to physical units). This reduction was quantitatively different across databases, and more pronounced in databases with more subjects (45% in LISTEN-V2).

Most ITDs and derived statistics are reported in the rest of the manuscript on *normalized* frequency and ITD scales with a single pair of normalization factors (frequency and ITD) for the population. For ease of reading, when representing human ITD data, ITDs and frequencies are represented both in normalized and direct physical units, with a conversion factor assuming a head radius equal to the average over the population ($a = 9.5$ cm). For this value of the head radius, a normalized frequency of 1 corresponds to a frequency of 573 Hz, and a normalized ITD of 1 corresponds to an ITD of 278 μs. Therefore, to convert from normalized to physical ITD in μs, multiply the normalized value by 278 μs. To convert from

normalized to physical frequency, multiply the normalized value by 573 Hz.

2. Asymptotic ITD in the horizontal plane

We estimated the asymptotic low and high frequency values of the ITD to compare them with theoretical predictions from the spherical model (Figure 3).

The high-frequency ITD, $\mathrm{ITD}_{\mathrm{HF}}(\theta)$ is estimated as the mean ITD between $f_{norm} = 7$ (4010 Hz) and $f_{norm} = 8$ (4600 Hz). As per construction, the Woodworth formula (Figure 3a, black) fits the data (shaded area: mean $\mathrm{ITD}_{\mathrm{HF}}(\theta)$ over subjects $\pm 1/2$ s.d.), except when the azimuth is in the $70 - 110°$ range, which has been previously documented (Aaronson and Hartmann, 2014). Figure 3B shows the low frequency prediction with Kuhn's formula (Eq. 3) alongside mean ITD estimated from HRTFs between $f_{norm} = 0.5$ (290 Hz) and $f_{norm} = 0.6$ (350 Hz), termed low frequency $\mathrm{ITD}_{\mathrm{LF}}(\theta)$. Consistent with previous reports, the approximation of the low-frequency ITD is reliable.

On Figure 3c, the values of ITD across frequency are reported for seven positions on the horizontal plane (0, 30, 60, 90° and the symmetrical positions). Curves represent the average normalized ITD across subjects, and shaded areas are ± 1 s.d.. The standard deviation of the normalized ITDs at each frequency and position was on average 46 μs (0.16 normalized). This variability is relatively small: it is about 6% of the maximal ITD value observed, and is about a factor of two larger than a human just-noticeable difference (JND) in ITD (the smallest ITD difference perceptible by human subjects (Mills, 1958)).

In addition, the plots on Figure 3c reveal some fine variations of ITD with frequency that are not accounted for by the spherical model (e.g. the low frequency bump for azimuth $\pm 30°$). This reveals that deviations of the human head morphology from a sphere contribute systematically to the ITD versus frequency relationship.

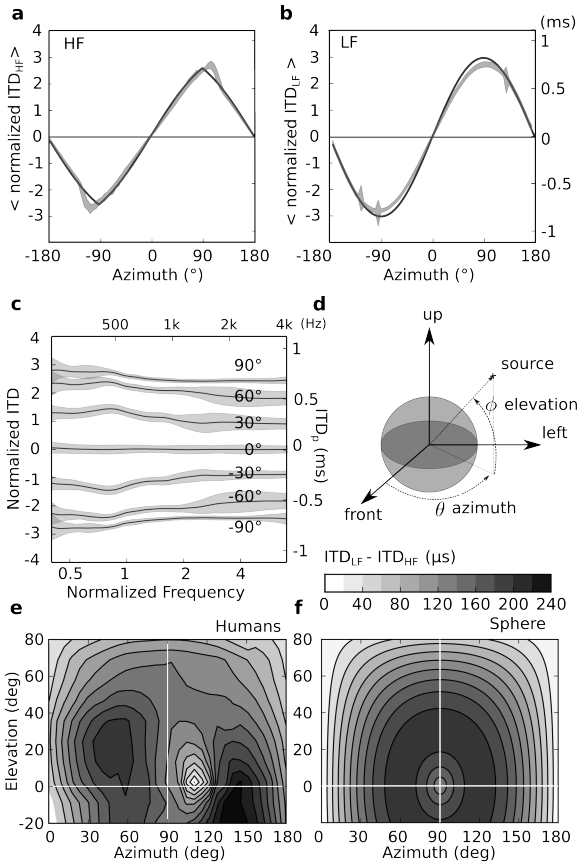

Figure 3: Frequency dependence of ITD in human subjects. (a) Inter-individual average normalized ITD in high frequency as a function of azimuth $\pm 1/2$ s.d. (shaded area). Black line indicates the theoretical value from the Woodworth model (Eq. 2). Corresponding ITD values for a head radius of 9.5 cm are shown on the right of panel b. (b) Average normalized ITD in low frequency $\pm 1/2$ s.d. (shaded area), black line indicates the theoretical value from Kuhn's formula (Eq. 3). (c) Average normalized ITD (black lines) as a function of frequency for seven source positions (shaded area: $\pm 1/2$ s.d.). (d) Azimuth θ and elevation ϕ are defined in a standard vertical-polar coordinate system (see text). (e) Difference between high- and low-frequency normalized ITD as a function of elevation and azimuth. Physical ITD is calculated for a head radius of 9.5 cm. (f) Same as (e) for the spherical model. Normalized units correspond to a head radius of 9.5cm.

3. Frequency variation of ITD as a function of azimuth and elevation

Consistent with previous reports, our data show that ITD is frequency-dependent in human HRTF, with similar differences between low and high frequency values as in the spherical models in the horizontal plane. We now quantify this difference as a function of both azimuth θ and elevation ϕ. Recall that we used a vertical-polar coordinate system (Fig. 3d). Fig. 3e shows the average difference $\mathrm{ITD}_{\mathrm{LF}} - \mathrm{ITD}_{\mathrm{HF}}$ across subjects for all positions on the spatial grid, converted to physical ITD values assuming head radius of 9.5 cm. For comparison, the same quantity is shown for the sphere on Fig. 3f.

The difference between high and low frequency ITD exceeds 50 μs for most of the positions on the sphere, and can reach more than 200 μs. As a comparison, human subjects can discriminate ITDs differing by a JND of only 20 μs (Mills, 1958). Therefore, the variation of ITD with frequency should be perceptually significant for most source positions away from the midline.

For large enough source distances, the pressure at any point on the surface of a sphere depends only on the angle between the ray from the center of the sphere to the source, and the ray to the measurement point on the surface of the sphere (Duda and Martens, 1998). Because of this symmetry property, binaural cues are constant for sources lying on so-called cones of confusion, centered on the interaural axis. In other words, acoustical cues in a spherical head model only depend on the angle of the source acoustical wave and the medial-sagittal plane: the *incidence angle* $\beta = \arcsin(\cos(\phi)\sin(\theta))$. Cones of confusion are then the set of points of equal incidence angle β. This aspect makes it hard to differentiate sound sources positioned symmetric to the interaural axis, which includes front and back positions.

Consistent with previous reports in humans, our data show that cones of confusion are centered around source positions directly facing the ears (that is, at the same azimuth and elevation as the ears, Aaronson and Hartmann, 2014). Furthermore, they appear distorted (around $(\theta, \phi) = (110°, 5°)$, Fig. 3e). In particular, the variation of ITD with frequency

quantitatively differs between front and back (and up and down) positions.

C. Maximal ITD

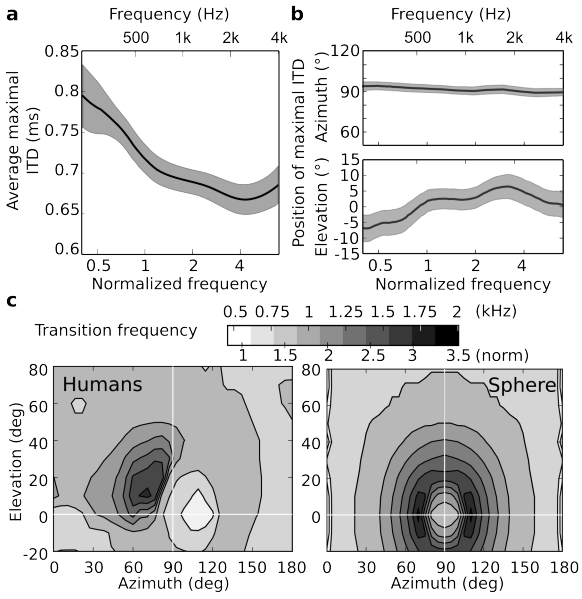

Figure 4: Maximal ITD and transition frequency in human subjects. (a) Maximal ITD across subjects as a function of frequency. (b) Azimuth (top) and elevation (bottom) where ITD is maximal as a function of frequency Shaded areas of (a) and (b) are the mean $\pm 1/2$ s.d. (c) Transition frequency (see text) as a function of azimuth and elevation in humans (left) and in the spherical model (right). Normalized units converted assuming a head radius of 9.5cm.

We estimated, from the *unnormalized* data, the maximal ITD over positions for each frequency and subject (reported here on a normalized frequency scale). By nature, ITD estimation can be unstable in high frequency because of the ambiguity inherent when unwrapping a phase response, and in low frequency because of the low frequency resolution. This is especially prominent when automatically processing a large number of recordings (several hundred positions, and subjects), and creates many outlier datapoints, which positively biases the estimation of a maximal ITD.

The 95% percentile (see, e.g., Papadatos, 1995) is a more robust estimator of sample maximum, which we define here as the maximal ITD. Consistent with the theory and previous reports, we found that the maximal low-frequency ITD value is 813 μs $\pm$ 70 μs (s.d., see Fig. 4a). The maximal broadband ITD – computed as the peak lag of the cross-correlated impulse responses, was found to be 612 μs $\pm$ 34 μs (s.d.). This value is very close to the value of the maximal high frequency ITD, 688 μs $\pm$ 47 μs (s.d., see Fig. 3b).

The maximal ITD occurs for azimuth around 95° (Fig. 4b), for which the source is directly facing one of the two ears, which is consistent with previous reports (Aaronson and Hartmann, 2014). There are systematic variations of the position of maximal ITD with frequency, but it remains near eccentric azimuths ($\pm$90°) and close to the horizontal plane (-10° to 10°). In the spherical model computations, the maximal ITD is reached at $\theta = 90°$ (Eq. 3), in humans it occurs for positions slightly more to the back (Fig. 4b), and for sources originating from below or above the horizontal plane, depending on the frequency of the signal (Fig. 4b).

D. Transition between ITD regimes

It could be argued that even though the ITD varies across frequency, this variation does not occur in the range where ITD is used as a cue to azimuth (i.e. the ITD is constant below 1.8 kHz). To assess this, we examined the shape of the ITD versus frequency curves, specifically trying to get at the frequency at which ITD effectively transitions between its low and high frequency regimes. We define the *transition frequency* as the frequency at which the ITD equals the average between its high frequency ($\mathrm{ITD}_{\mathrm{LF}}(\theta, \phi)$) and low-frequency ($\mathrm{ITD}_{\mathrm{HF}}(\theta, \phi)$) values for a given position. Because in general ITD is a decreasing function of frequency with a relatively narrow transition, this transition frequency allows us to separate frequency regions of high and low ITD values for any position.

Figure 4C shows the transition frequency in humans and in the spherical head model (for an infinitely distance source), as a function of azimuth and elevation. The transition

frequency increases as the source is moving away from the median sagittal plane, up to an azimuth angle $\simeq 70°$ where a maximum value $f_{\text{tran}} \simeq 2.8$ (3.2 in the spherical model) is reached. It then decreases until a minimum is reached at $\theta \simeq 110°$ and $\phi \simeq 0°$ ($\theta = 90°$ and $\phi = 0°$ in the spherical model).

In conclusion, for all positions, transition frequencies are between 1 and 3 (normalized scale), which corresponds to physical frequencies between approximately 600 and 1700 Hz. The ITD thus varies substantially at frequencies within the range where ITD is the dominant cue for sound laterality in the horizontal plane (Wightman and Kistler, 1992). The magnitude of transition frequencies in humans is overall similar to the predictions of the spherical model, yet as previously mentioned symmetries seen in the spherical model do not appear in the human HRTF data.

IV. ENVELOPE AND FINE-STRUCTURE ITD

When a sound wave excites the cochleae, different points on the basillar membranes are preferentially excited by energy in different frequency bands. Many neurons in the auditory system, in particular neurons in the midbrain that are sensitive to ITD, are also tuned to different frequencies and are tonotopically organized. We have shown that ITD varies substantially *across* different frequency bands, that is, between distant auditory filters (Fig. 5a). An interesting question is whether ITD also varies substantially *within* a single auditory filter (Fig. 5b), as it would then have direct physiological relevance. We relate this question to the difference between envelope and fine-structure ITD.

A. Variation of ITD within a auditory filter

We first analyzed the variation of ITD within single channels for each position and subject (same HRTF database as in Section III). The variation of ITD is defined as the difference between the maximum and minimum ITD_p in a frequency band with constant $Q = 4.3$ (one-third octave) or with equivalent-rectangular bandwidths (ERB) (Glasberg

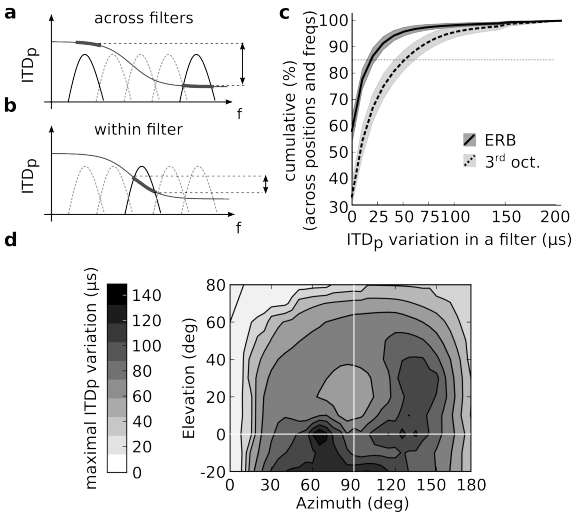

Figure 5: Variation of ITD within single auditory filters. (a) Schematics of the global variation of ITD across different auditory filters. (b) Schematics of the variation of ITD within a single auditory filter. (c) Proportion of positions and center frequencies where ITD variation within a single channel is smaller than a specified value. Lines are averages across population, ± s.d.. Channel width is either 1 ERB or 1/3 octave. (d) Maximal ITD variation within single channels, as a function of azimuth and elevation, with ERB-wide channels.

and Moore, 1990) (Fig. 5b). In the ERB scale, the Q factor value varies between 5 and 9, indicative of the relatively narrow band filtering imposed by the auditory periphery. We computed the variation of ITD for center frequencies between 350 and 3000 Hz and for positions close to the horizontal plane ($|\phi| \leq 20°$). We report the cumulative distributions of ITD variation on Fig. 5c: curves display the percentage of positions and center frequencies for which the ITD variation is lower than a given amount. In both conditions, for more than 15 % of the channels and positions the magnitude of the ITD variation is larger than 25 μs (Fig. 5c). In Fig. 5d, we show the maximal variation of ITD in single channels as a function of azimuth and elevation (assuming ERB-wide channels). That is, for every position, we report the variation of ITD in the channel where it varies the most. At specific source positions, very large variations of ITD can occur within channels (up to 150 μs): the

311 variation of ITD within a single channel is therefore quite substantial.

The fact that ITD varies within a frequency band means the signal undergoes more than a simple delay when passed through the HRTF. Mathematically, the phase responses of the monaural filters are nonlinear functions of frequency. We can approximate the IPD by an affine (i.e. linear with a non-zero intercept) function of frequency around the center frequency f_0 of a cochlear filter (Fig. 6a):

$$\begin{aligned} \mathrm{IPD}(f) &\stackrel{def}{=} \mathrm{ITD}_p(f) f \\ &\approx \mathrm{ITD}_g(f_0)(f - f_0) + \mathrm{IDI}(f_0) \end{aligned} \quad [1]$$

312 where phases are expressed in cycle. The slope of this fit is the *group* ITD, which is the ITD
313 of the envelope (Marple Jr, 1999).

314 The intercept IDI is an additional shift in the phase of the fine-structure of the signal
315 (Fig. 6b). This shift only occurs when the phase ITD varies with frequency, i.e., when
316 propagation does not result in a pure delay. For this reason, we termed this binaural cue
317 the *Interaural Diffraction Index* (IDI, see Rebillat *et al.* (2014)). The IDI can be seen as a
318 measure of the difference between group and phase ITDs at any frequency, converted into a
319 phase value: $\mathrm{IDI} = (\mathrm{ITD}_p - \mathrm{ITD}_g) f$. If $\mathrm{IDI} = 0$ cycles, phase and group ITD are equal, and
320 locally frequency-independent. When the IDI is positive, by convention the phase ITD has
321 a higher absolute value than the group ITD, and vice versa when IDI is negative.

322 B. Estimating ITD in the envelope and fine-structure of binaural signals

323 The group ITD is classically defined as the derivative of the unwrapped IPD curve with
324 respect to frequency as represented in Fig. 6a. Because of occasional errors of the phase
325 unwrapping operation, estimating the derivative from large sets of unwrapped IPD curve is
326 unreliable. Instead we use an equivalent approach wherein we perform circular-linear fits on
327 the *wrapped* IPD. The estimation can be formulated as a non linear least square problem,
328 where IDIand $\mathrm{ITD_g}$are chosen to minimize the fit error $\sum_f \|\mathrm{IPD}(f) - (\mathrm{IDI} + \mathrm{ITD_g} f)\|^2$ over
329 a given frequency band, where the norm $\|.\|$ is a norm on phases. Because wrapped phase

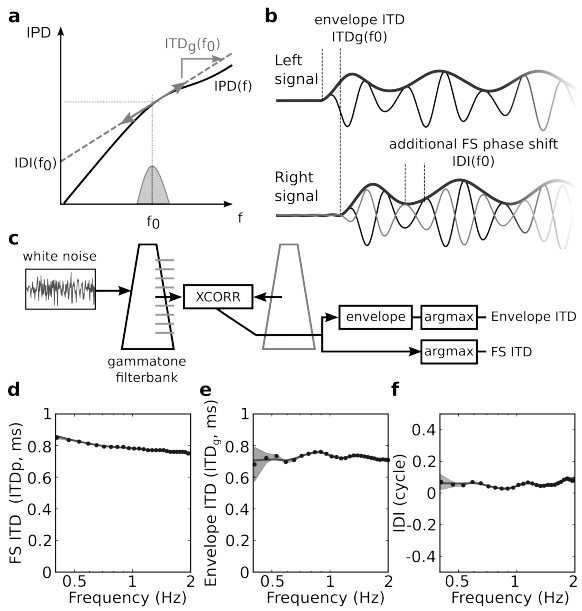

Figure 6: Envelope and fine-structure ITD. (a) The IPD for one position is unwrapped and an affine fit is taken locally around f_0. The intercept of the fit is the IDI and the slope the group ITD. (b) When the IDI is zero, the delay is frequency-independent and both envelope and fine-structure are delayed by the ITDg (bottom, black signal). When IDI is non zero, the fine-structure undergoes an additional phase shift equal to the IDI (bottom, grey signal). (c) Simulation: white noise is passed through HRTF filters for one position (spherical head model, azimuth = 70°). The resulting signals are then fed into gammatone filterbanks. The responses in the two banks are then cross-correlated, and the result is separated in envelope and fine-structure components. The time lag of the maximum of the cross-correlation is the phase ITD, and that of the maximum of the envelope of the cross-correlation is the envelope ITD (see text). (d) Results of estimating phase ITD, $\mathrm{ITD_p}$, from the IPD (plain line), and from simulations (mean: dots, shaded area: 95% confidence interval). (e) Same as (d) for $\mathrm{ITD_g}$. (f) Same as (d) for IDI.

values are a circular quantity, so is the norm we use in the fit. It can be expressed as a cosine function of its argument: $\|x\|^2 = 1 - \cos(2\pi x)$. The precise algorithm used is described in more details in the Methods section of Luling *et al.* (2011).

To show in practice that group ITD and IDI can be extracted from the envelope and fine-structure of band-limited binaural signals, we simulated a simple model of the auditory periphery. We then computed the envelope and fine-structure ITD in different frequency bands of our auditory model (Fig. 6c) using a standard cross correlation approach. The auditory periphery model consisted of two gammatone filter banks receiving 100-ms-duration white-noise inputs convolved with the spherical HRTFs. Each filter's response was then cross-correlated with the opposite frequency band. This operation is a good approximation of the response of binaural neurons in the medial superior olive of mammals (Yin and Chan, 1990). The fine-structure ITD is obtained by computing the position of the maximum of the cross correlation function. Then, the envelope of the cross-correlation is extracted using a Hilbert transform, and the maximum computed. This procedure yields an estimate of the delay in the envelopes (i.e. it is equivalent to computing the maximum of the cross-correlation of the envelopes Marple Jr (1999)). The results of this simulation are plotted on Fig. 6d-f. Dotted points (with 95% confidence intervals over repeated trials) represent the ITDs estimated using the simplified auditory model. The theoretical predictions (plain lines) were obtained by taking circular-linear fits of the IPD of the HRTF pair used. The match between theory and simulation is excellent, which shows that the group ITD indeed corresponds to the envelope ITD, which appears in the cross-correlation of the monaural signals, and the IDI indeed corresponds to the additional shift of the fine-structure seen in the cross-correlation function.

C. Analysis of envelope and fine-structure ITDs in human HRTFs

We computed ITDs in the envelope and fine-structure of the human data presented in Section III, according to the methods presented above. On Figure 7 we present the average $\mathrm{ITD_p}$, $\mathrm{ITD_g}$ and IDI over the whole population. We observe that $\mathrm{ITD_p}$ and $\mathrm{ITD_g}$ can be dramatically different in some frequency bands (typically around 1kHz). As a result the IDI is non-zero in that range (Figure 7e), which correponds to the frequency range just above

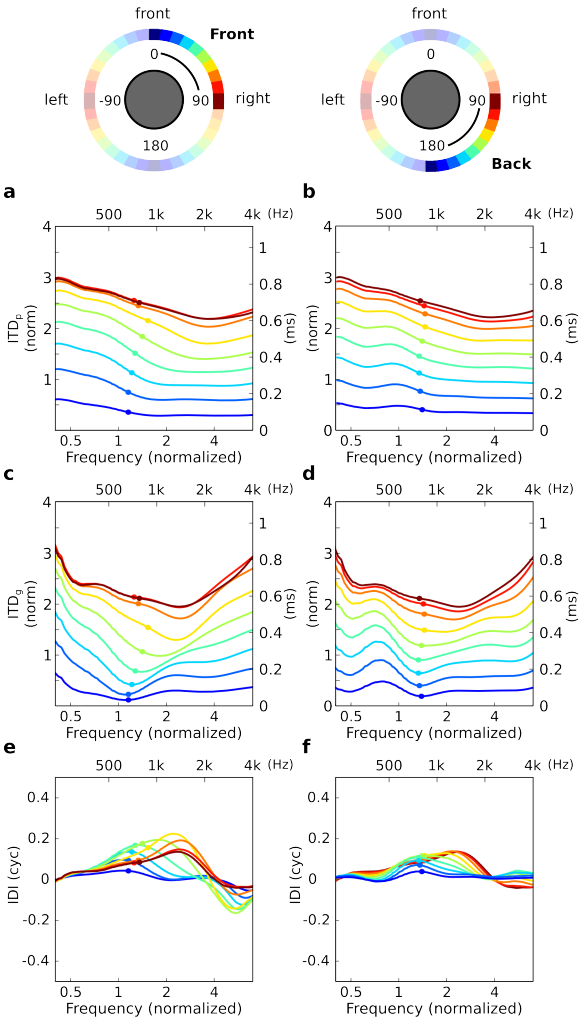

Figure 7: Averages over the whole population of normalized ITD_p (a,b), ITD_g (c,d), and IDI (e,f) for horizontal plane positions as a function of frequency. Top part of the figure depicts the lines color codes (positive azimuths, separated by 10°). Blue lines are more medial positions, and red more eccentric. Left column (a,c,e) displays data from the front positions, right column (b,d,f) from the back positions. Dots overlaid on the line plot represent the position of the transition frequency.

the "transition frequency" introduced above (data not shown).

For all positions, we find that the IDI is close to zero for lower frequencies. It then positively increases in low frequencies (below 2-3 kHz) and then drops to negative values for higher frequencies (above about 4kHz). Therefore, generally the fine-structure ITD is higher

than the envelope ITD for frequencies between 1 and 3 kHz, and smaller in high frequencies (around 4kHz). While both ITDs are monotonically increasing functions of eccentricity (Figure 7a-b, c-d), the relationship is more complex for IDI. In addition, as noted in the above statistical analyses, the frequency-dependence of ITD is different for front and back positions (Figure 7e,f), which is potentially a cue to disambiguate between them.

V. FREQUENCY-DEPENDENCE OF ITD IN ANIMALS

In many animal species, neurons tuned to ITD have been identified, e.g. in the Medial Superior Olive (MSO) or Inferior Colliculus (IC) of mammals (Yin and Chan, 1990). A recent debate has emphasized the importance of the natural distribution of binaural cues in our understanding of electrophysiological data (Grothe *et al.*, 2010). In this context, and more generally in neurophysiological studies of sound localization, binaural cues (ITD and Interaural Intensity Differences, IID) have been measured for a number of animal species, including mammals, but also birds and reptiles.

As seen in humans, a strong dependence on morphological features of the animals is found in many instances. For example, the owl's facial ruff (Campenhausen and Wagner, 2006), or the cat's pinnae (Tollin and Koka, 2009) increase the magnitude of ITD at a given position. However, the variation of ITD with frequency has received little attention until recently (Benichoux *et al.*, 2015). We applied the same analysis as above to measured HRTFs of different animal species.

A. Animal HRTF recordings

We measured the HRTFs of six taxidermist animal models, using the same setup as for the human recordings (LISTEN-V2, see Table I): rat, rabbit, guinea pig, chinchilla, cat and macaque. All animals had their ear canals obstructed by the taxidermy, which means that recordings are taken in a blocked-meatus configuration. Animal models were chosen according to the well-preserved quality of their pinnae. We previously showed that

HRTFs measured on taxidermist models agree closely with acoustical simulations based on 3D models of the animal with rigid boundaries (Rebillat *et al.*, 2014).

In general, animal models were in a natural-looking position, in which the head of the animal is not aligned with the body. Therefore, the coordinate system is rotated so that the head points in the 0° direction. This is achieved by computing the *head angle* relative to the body: the azimuth that minimizes the magnitude of the low-frequency ITD value. Head angles were generally non-zero for all models (rat 10°, rabbit -10°, guinea pig 20°, chinchilla -20°, cat -55°, macaque -30°). It should also be noted that the interaural distance of the rabbit, guinea pig and macaque models in the present study are noticeably smaller than those of animals of the same species whose recordings are reported in the literature (see Table II).

B. Frequency-dependent ITD in animals

A difference between low- and high-frequency ITD in animals has been previously shown in a number of animal species: in the rat (Koka *et al.*, 2008), rabbit (Kim *et al.*, 2010), guinea pig (Greene *et al.*, 2014), chinchilla (Lupo *et al.*, 2011; Koka *et al.*, 2011), cat (Roth *et al.*, 1980; Tollin and Koka, 2009) and macaque monkey (Spezio *et al.*, 2000). Yet, only very few studies reported the frequency-dependent ITD curves for many azimuth positions. The phase ITD for all animal models and frequencies between 350 and 3000 Hz is reported on Fig. 8, for positions in the horizontal plane. Consistent with physical acoustics, and the above results in humans, the phase ITD is frequency-dependent in all species.

Because humans are bipeds, no part of the body normally finds itself on the way between the source and the ears. In many quadrupeds, for example in the cat (Fig. 8e), sounds coming from the back can be reflected or scattered by the body before reaching the head and ears. This morphological asymmetry results in large differences between the frequency-dependent ITDs of sources in the front (solid curves) and in the back (dashed). Thus in principle, front and back positions can be distinguished on the basis of the ITD at different

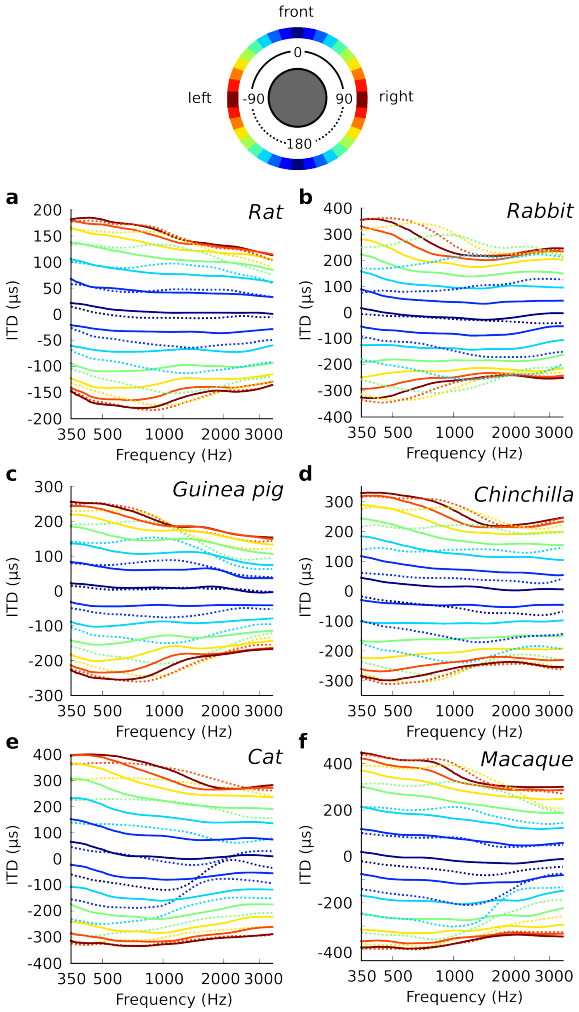

Figure 8: Animal ITDs in the horizontal plane, for 24 positions around the head (separated by 15°). Top: line color code; front positions, solid lines; back, dashed lines. (a) Rabbit, (b) Guinea pig, (c) Chinchilla, (d) Cat, (e) Rat, (f) Macaque.

414 frequencies. Similarly, it was shown using acoustical measures and simulations that the
415 posture of the animal influences the frequency-dependent pattern of ITD (Rebillat *et al.*,
416 2014).

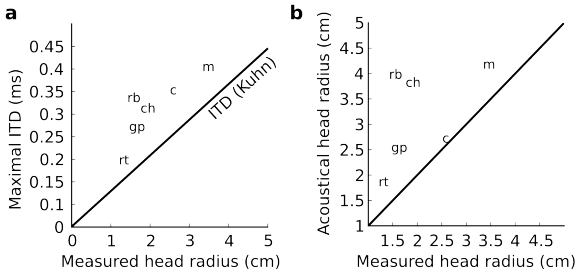

Figure 9: Comparison of measured ITD range with anatomy. (a) Maximal measured low-frequency ITD as a function of half the interaural distance measured on the taxidermist models. Predictions are shown for Kuhn's formula (solid). (b) Acoustical head radius estimated with Woodworth's formula (see Text) vs. half the interaural distance measured on the taxidermist models (dashed line: diagonal). Legend: rb: rabbit; ch: chinchilla; gp: guinea pig; m: macaque; rt: rat; c: cat.

C. Estimating animal ITD from head size

For electrophysiological studies, a way to estimate the maximal ITD for a given animal species is to measure the interaural distance and then use one of Woodworth's (Eq. 2) or Kuhn's formula (Eq. 3). The maximal ITD is thus obtained in low frequencies (via Eq. 3) for the most eccentric position, that is $\mathrm{ITD}_{\max} = 3a/c$. However, in some species there is evidence that this method yields an underestimation of the maximal ITD (see, e.g. (Tollin and Koka, 2009) in the cat). " In Fig. 9a, the maximal ITD measured in the horizontal plane is reported as a function of the half interaural distance of the animal, measured as the half distance between the entrances of the (occluded) ear canals. In all cases, the maximal ITD is well correlated with the physical head size, but substantially larger than predicted using the Kuhn formula. We further computed the *acoustical head radius* for each animal model, as we did in humans (see Section III). Consistent with previous observations (e.g. (Tollin and Koka, 2009)), the *acoustical* head radius of animals is substantially larger than their physical "head radius" (the half interaural distance of the animal model, Fig. 9b).

Overall, our animal results suggest that the maximal ITD should not be estimated from

a crude measure of the morphology of the animal (here interaural distance), because this leads to a systematic underestimation of the magnitude of ITD for all species.

VI. SUMMARY AND DISCUSSION

A. Summary

In this paper, we have quantified the variation of ITD with frequency in humans and animals, measured in anechoic space. First, we confirmed that ITD does vary significantly with frequency, as predicted by a spherical head model (Kuhn, 1977) and mentioned in previous studies (Wightman and Kistler, 1989). Specifically, maximal ITD values were found to be about 800 μs in low frequencies and 600 μs in high frequencies. Therefore, the low frequency ITD can be larger than the high frequency ITD by as much as 200 μs, which is an order of magnitude larger than human JNDs in ITD discrimination tasks (of the order of 20 μs), even for pure tones (Brughera *et al.*, 2013) (10-40 μs for frequencies below 1250 Hz). The transition between low- and high-frequency ITDs occurs at frequencies between 600 and 1800 Hz, within the range where ITD is a dominant cue for localization in the horizontal plane (Wightman and Kistler, 1992).

Additionally, we observe that the frequency-dependence of ITD does not exhibit simple spherical symmetries. In particular, symmetrical front and back (and up and down) positions, both in humans and animals, have different frequency-dependent ITDs. The frequency dependence of ITD provides, in addition to azimuth, a cue to elevation including information about front versus back.

We also show that for multiple of source positions, ITD varies not only globally across the spectrum, but also locally within the bandwidth of a single auditory filter. This causes different ITDs for envelope and fine-structure, which can provide additional information about the position of the sound source. Furthermore, those cues can be estimated from binaural signal using cross-correlation. The difference in group and fine-structure ITD is quantified by the *interaural diffraction index* (IDI).

B. Relation with psychoacoustical experiments

A few studies have examined the sensitivity of human subjects to the frequency-dependence of ITD. Kistler and Wightman (1992) showed that localization errors for bursts of white noise are similar with individual HRTFs compared to HRTFs in which the monaural phase information was degraded. However, these manipulated HRTFs — minimum-phase filters, do in fact preserve the frequency-dependence ITDs, in a way known to be close to those of measured HRTFs (Kulkarni *et al.*, 1999).

In a study on the cues for externalization of sounds, Hartmann and Wittenberg (1996) showed that human subjects are unable to detect the substitution of the phase information of HRTFs by a properly adjusted frequency-independent ITD. The test was done for a source location at 37° on the horizontal plane, in an anechoic room. According to our analysis, at that location the ITD varies by about 130 μs across frequency and the transition frequency is 900 Hz (about 1.6 in normalized frequency). Kulkarni *et al.* (1999) also found that human subjects were unable to discriminate individual HRTFs from linear phase HRTFs, as long as the average low-frequency ITD was correct. Constan and Hartmann (2003) also showed that subjects cannot determinate whether binaural sounds have frequency-independent ITDs or frequency-dependent ITDs as in the spherical model — however, neither of these two cases is entirely realistic.

The fact that human subjects cannot perceive the difference is puzzling, because they can detect ITD changes of 10-40 μs in pure tones below 1250 Hz (Brughera *et al.*, 2013). Furthermore, in a two-dimensional absolute localization task the mean error is about 5° in the frontal hemifield (for broadband noise bursts, Makous and Middlebrooks (1990)), which corresponds to about 50 μs ITD. As the ITD variation across the spectrum can reach 200 μs for some positions, systematic frequency-dependent errors should be observed if the ITD variation were discarded.

Together, these studies suggest that human subjects can detect small ITD changes in tones when they are presented in isolation, but they cannot detect them when they are

embedded in a complex sound, as long the average ITD is unchanged. This is consistent with the notion that source location is inferred from the pattern of ITD, but that only that inferred location, rather than the acoustical cues, is available to conscious perception and behavior, and in particular is used in discrimination tasks. Thus, two sounds with different patterns of frequency-dependent ITD are indistinguishable if they yield the same estimated location. This is consistent with other aspects of binaural hearing. In particular, it has been shown that the sensitivity to interaural intensity differences (IID) is substantially degraded when the use of intracranial position as a cue is eliminated by roving the the ITD of the stimuli (Bernstein, 2004).

A possible experiment to determine whether ITD information is discarded in estimating the location of the source is to include judgements of the position of sounds with different frequency contents. For example, localization performance could be tested as in measured HRTFs with linear phase HRTFs, but with band-pass filtered noises in different frequency regions. If the frequency-dependence of ITD is discarded, then results should be identical in the two conditions (provided the ITD of linear phase HRTFs is adjusted). On the other hand, if the variation of ITD is indeed taken into account to estimate source position, we should observe systematic errors depending on frequency and position.

C. Binaural coherence

Binaural coherence is defined as the maximal value of the cross-correlation of monaural signals (Gabriel and Colburn, 1981). Humans are very sensitive to small changes in binaural coherence, usually modeled by adding a small amount of independent noise at each ear (usually below 3-4% for noise (Gabriel and Colburn, 1981)). In HRTF recordings, binaural coherence is found to be mainly affected by the amount reverberation in the room: binaural coherence is very high in anechoic environments, and dramatically goes down as the environment gets more reverberant (Hartmann *et al.*, 2005). It can be argued that the effect of the variation of ITD within an auditory filter is a decreased coherence (Constan

and Hartmann, 2003). Yet, in anechoic conditions this effect remains marginal, especially for the narrow bands of noise resulting from filtering by the auditory periphery (less than 0.1% (Constan and Hartmann, 2003)).

We argue here for a different interpretation of the frequency-dependence of ITD. Decoherence due to reverberation is intrinsically non-deterministic: different wavefronts reach the listener at different times depending on the unknown geometry of the room. On the other hand, we have shown above that the frequency-dependence of ITD has a fully deterministic effect: envelope and fine-structure ITD cues are affected in a way that is predicted by the morphology of the subject. Decoherence, insofar as it is non-deterministic, objectively makes the task of recovering the ITD from the cross-correlation function harder. It is unclear, however, why imposing different ITDs in the envelope and fine structure of the monaural signals would make the recovering of ITDs harder, because it is fully deterministic. Therefore, we argue that the variation of ITD in small frequency bands is best thought of as imposing different ITDs in the envelope and fine-structure of monaural inputs, rather than as causing binaural decoherence, as imposed by adding independent white noise to monaural inputs.

D. Signal processing of binaural sounds

Our results are relevant to two classes of signal processing applications: reproduction of binaural sounds and sound localization algorithms. The large variation of ITD with frequency suggests that it is important for proper reproduction of binaural sounds. However, it could be that humans can adapt to non-natural ITD patterns, as they do to spectral cues (Wanrooij and Opstal, 2005). In either case, we note that replacing frequency-dependent ITDs with fixed ITDs removes some potential cues to elevation.

State-of-the-art sound localization algorithms using HRTF-filtered inputs do use the frequency-dependence of ITD to estimate source location. In the algorithm described by May *et al.* (2011), sounds are divided into frequency bands, and position is estimated with a

maximum likelihood approach from the overall ITDs in these bands. Because ITD likelihood for each position is measured with KEMAR HRTFs, this algorithm uses the variation of ITD across channels. However, it does not use the variation of ITD within channels.

Other algorithms use HRTF data with the within-channel ITD variations preserved, (Durkovic *et al.*, 2011; Macdonald, 2008) and were shown to perform well in realistic conditions. In each frequency band, monaural signals are convolved with the contralateral HRTF of a candidate source position (i.e., left signal with right HRTF), and the position giving the highest cross-correlation is picked. A spiking neural model relying on similar ideas was also previously presented (Goodman and Brette, 2010): it used cross-correlation, biophysically modeled with coincidence detection between spike trains, and performed better when the variation of ITD within channel was taken into account.

E. Electrophysiology

The firing rates of neurons in several auditory brainstem nuclei, in particular the medial superior olive (MSO) and inferior colliculus (IC) of mammals, is sensitive to the ITD of binaural sounds (Grothe *et al.*, 2010). Similar to humans, we have shown that ITD is frequency-dependent in animals, in the frequency range where it is used for sound localization (Fig. 8). Furthermore, we showed that asymmetries in this frequency-dependence exist between front and back positions, presumably due to reflections on the back of the animals (Fig. 8). Finally, we have noted that the maximal ITD is generally larger than when estimated from simple morphological considerations (Fig. 9). All these observations should be taken into account when interpreting electrophysiological measurements.

In the physiological literature, two types of frequency-dependent properties have been discussed (Grothe *et al.*, 2010; Day and Semple, 2011; Benichoux *et al.*, 2015). The preferred ITD of binaural neurons, i.e., the ITD that elicits the largest firing rate, depends on their preferred frequency: at the level of the population those quantities are inversely correlated. This observation has been seen as a challenge to the mainstream theory, according to which

neurons are tuned to the ITD of particular source locations, which should cover all possible locations independently of the frequency band (Grothe *et al.*, 2010). In our animal measurements, ITD is also larger in lower frequencies than in high frequencies — although to a smaller extent than in electrophysiological recordings. An additional contribution to large low-frequency ITDs in animals is early reflections on the ground, which produce arbitrarily large ITDs in low frequencies (Gourevitch and Brette, 2012).

Many binaural neurons also display a second type of frequency-dependence: for a given neuron, the preferred ITD depends on the frequency of the sound (Day and Semple, 2011). We have shown that ITD varies also with frequency within an auditory filter, which provides a potential ecological explanation of this variation (Benichoux *et al.*, 2015). The present analysis suggests that cells with frequency-dependent best delays should be differentially sensitive to envelope and fine-structure delays.

Acknowledgments

This work was supported by the European Research Council (ERC StG 240132). We thank the Museum d'Histoire Naturelle de la Ville de Paris for lending the taxidermist animal models, as well as Olivier Warusfel at IRCAM (Paris) for the anechoic chamber.

References

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

Algazi, V., Duda, R., Thompson, D., and Avendano, C. (**2001**a). "The CIPIC HRTF database", in *Proceedings of the IEEE Worshop on the Applications of Signal Processing to Audio and Acoustics*, 99–102.

Algazi, V. R., Avendano, C., and Duda, R. O. (**2001**b). "Estimation of a Spherical-Head Model from Anthropometry", Journal of the Audio Engineering Society **49**, 472–479.

ARI (**2010**). "HRTF Database http://www.kfs.oeaw.ac.at/content/view/608/606/lang,8859-1/ (date last viewed 13/8/2015)", .

Benichoux, V., Fontaine, B., Franken, T. P., Karino, S., Joris, P. X., and Brette, R. (**2015**). "Neural tuning matches frequency-dependent time differences between the ears", eLife **4**, e06072.

Bernstein, L. R. (**2004**). "Sensitivity to interaural intensitive disparities: listeners' use of potential cues", The Journal of the Acoustical Society of America **115**, 3156–3160.

Brughera, A., Dunai, L., and Hartmann, W. M. (**2013**). "Human interaural time difference thresholds for sine tones: The high-frequency limit", The Journal of the Acoustical Society of America **133**, 2839–2855.

Bruneau, M. (**2010**). *Fundamentals of acoustics* (Wiley-ISTE, London, 636 pages).

Cai, T., Rakerd, B., and Hartmann, W. M. (**2015**). "Computing interaural differences through finite element modeling of idealized human heads", The Journal of the Acoustical Society of America **138**, 1549–1560, URL `http://scitation.aip.org/content/asa/journal/jasa/138/3/10.1121/1.4927491`.

Campenhausen, M. v. and Wagner, H. (**2006**). "Influence of the facial ruff on the sound-receiving characteristics of the barn owl's ears", Journal of Comparative Physiology A **192**, 1073–1082.

Constan, Z. and Hartmann, W. (**2003**). "On the detection of dispersion in the head-related transfer function", The Journal of the Acoustical Society of America **114**, 998–1008.

Day, M. L. and Semple, M. N. (**2011**). "Frequency-dependent interaural delays in the medial superior olive: implications for interaural cochlear delays", Journal of Neurophysiology **106**, 1985–1999.

Duda, R. and Martens, W. (**1998**). "Range dependence of the response of a spherical head model", The Journal of the Acoustical Society of America **104**, 3048–3058.

Duda, R. O., Avendano, C., and Algazi, V. R. (**1999**). "An adaptable ellipsoidal head model for the interaural time difference", in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 965–968.

Durkovic, M., Habigt, T., Rothbucher, M., and Diepold, K. (**2011**). "Low latency localization of multiple sound sources in reverberant environments", The Journal of the Acoustical Society of America **130**, EL392–EL398.

Gabriel, K. J. and Colburn, H. S. (**1981**). "Interaural correlation discrimination: I. Bandwidth and level dependence", The Journal of the Acoustical Society of America **69**, 1394–1401.

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

Goodman, D. F. M. and Brette, R. (**2010**). "Spike-Timing-Based Computation in Sound Localization", PLoS Comput Biol **6**, e1000993.

Gourevitch, B. and Brette, R. (**2012**). "The impact of early reflections on binaural cues", The Journal of the Acoustical Society of America **132**, 9–27.

Greene, N. T., Anbuhl, K. L., Williams, W., and Tollin, D. J. (**2014**). "The acoustical cues to sound location in the guinea pig (Cavia porcellus)", Hearing Research **316C**, 1–15.

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

Hartmann, W. and Wittenberg, A. (**1996**). "On the externalization of sound images", The Journal of the Acoustical Society of America **99**, 3678–3688.

Hartmann, W. M., Rakerd, B., and Koller, A. (**2005**). "Binaural coherence in rooms", Acta acustica united with acustica **91**, 451–462.

Kim, D. O., Bishop, B., and Kuwada, S. (**2010**). "Acoustic Cues for Sound Source Distance and Azimuth in Rabbits, a Racquetball and a Rigid Spherical Model", Journal of the Association for Research in Otolaryngology **11**, 541–557.

Kistler, D. J. and Wightman, F. L. (**1992**). "A model of Head-Related Transfer Functions based on Principal Components Analysis and Minimum-Phase Reconstruction", Journal of the Acoustical Society of America **91**, 1637–1647.

Koka, K., Jones, H. G., Thornton, J. L., Lupo, J. E., and Tollin, D. J. (**2011**). "Sound pressure transformations by the head and pinnae of the adult Chinchilla (Chinchilla lanigera)",

644 Hearing Research **272**, 135–147.

645 Koka, K., Read, H. L., and Tollin, D. J. (**2008**). "The acoustical cues to sound location in
646 the rat: Measurements of directional transfer functions", Journal of the Acoustical Society
647 of America **123**, 4297–4309.

648 Kuhn, G. (**1977**). "Model for interaural time difference in azimuthal plane", Journal of the
649 Acoustical Society of America **62**, 157–167.

650 Kulkarni, A., Isabelle, S., and Colburn, H. (**1999**). "Sensitivity of human subjects to head-
651 related transfer-function phase spectra", Journal of the Acoustical Society of America **105**,
652 2821–2840.

653 Luling, H., Siveke, I., Grothe, B., and Leibold, C. (**2011**). "Frequency-Invariant Represen-
654 tation of Interaural Time Differences in Mammals", PLoS Comput Biol **7**, e1002013.

655 Lupo, J. E., Koka, K., Thornton, J. L., and Tollin, D. J. (**2011**). "The effects of experimen-
656 tally induced conductive hearing loss on spectral and temporal aspects of sound transmission
657 through the ear", Hearing Research **272**, 30–41.

658 Macdonald, J. A. (**2008**). "A localization algorithm based on head-related transfer func-
659 tions", The Journal of the Acoustical Society of America **123**, 4290–4296.

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

662 Marple Jr, S. L. (**1999**). "Estimating group delay and phase delay via discrete-time "ana-
663 lytic" cross-correlation", Signal Processing, IEEE Transactions on **47**, 2604–2607.

664 May, T., van de Par, S., and Kohlrausch, A. (**2011**). "A Probabilistic Model for Robust Lo-
665 calization Based on a Binaural Auditory Front-End", IEEE Transactions on Audio, Speech,
666 and Language Processing **19**, 1–13.

667 Middlebrooks, J. C. (**1999**). "Individual differences in external-ear transfer functions reduced
668 by scaling in frequency", The Journal of the Acoustical Society of America **106**, 1480–1492.

669 Mills, A. W. (**1958**). "On the minimum audible angle", The Journal of the Acoustical Society
670 of America **30**, 237.

671 Papadatos, N. (**1995**). "Maximum variance of order statistics", Annals of the Institute of

672 Statistical Mathematics **47**, 185–193.

673 Rayleigh, L. and Lodge, A. (**1904**). "On the Acoustic Shadow of a Sphere. With an Ap-
674 pendix, Giving the Values of Legendre's Functions from P0 to P20 at Intervals of 5 Degrees",
675 Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of
676 a Mathematical or Physical Character **203**, 87–110.

677 Rebillat, M., Benichoux, V., Otani, M., Keriven, R., and Brette, R. (**2014**). "Estimation
678 of the low-frequency components of the head-related transfer functions of animals from
679 photographs", The Journal of the Acoustical Society of America **135**, 2534–2544.

680 Roth, G., Kochhar, R., and Hind, J. (**1980**). "Inter-aural Time Differences - Implications
681 regarding the neurophysiology of sound localization", Journal of the Acoustical Society of
682 America **68**, 1643–1651.

683 Spezio, M., Keller, C., Marrocco, R., and Takahashi, T. (**2000**). "Head-related transfer
684 functions of the Rhesus monkey", Hearing Research **144**, 73–88.

685 Tollin, D. J. and Koka, K. (**2009**). "Postnatal development of sound pressure transformations
686 by the head and pinnae of the cat: binaural characteristics", The Journal of the Acoustical
687 Society of America **126**, 3125.

688 Wanrooij, M. M. V. and Opstal, A. J. V. (**2005**). "Relearning Sound Localization with a
689 New Ear", The Journal of Neuroscience **25**, 5413–5424.

690 Warusfel, O. (**2002**). "LISTEN HRTF database, http://recherche.ircam.fr/equipes/salles/listen
691 (date last viewed 13/8/2015)", .

692 Wightman, F. L. and Kistler, D. J. (**1989**). "Headphone Simulation of Free-Field Listening.
693 1. Stimulus Synthesis", Journal of the Acoustical Society of America **85**, 858–867.

694 Wightman, F. L. and Kistler, D. J. (**1992**). "The Dominant Role of Low-frequency Inter-
695 aural Time Differences inf Sound Localization", Journal of the Acoustical Society of America
696 **91**, 1648–1661.

697 Yin, T. C. and Chan, J. C. (**1990**). "Interaural time sensitivity in medial superior olive of
698 cat", Journal of Neurophysiology **64**, 465–488.

Database	N_s	$\Delta\theta$	$\Delta\phi$	N_{sub}	Room type
LISTEN-V1	8192	15°	$\simeq 15^\circ$	49	Anechoic
LISTEN-V2	8192	5°	$\simeq 15^\circ$	35	Anechoic
CIPIC	200	$\simeq 10^\circ$	5.6°	36	Anechoic
ARI	2400	5°	10°	10	Semi-anechoic

Table I: Overview of the different human HRTF databases used in this study. For each database, the sampling frequency is 44.1 kHz. N_s: length of the head-related impulse responses in samples. $\Delta\theta$, $\Delta\phi$: approximate spatial resolution in azimuth and elevation. N_{sub}: number of subjects from each database included in the present study. The LISTEN database consists of the 49 subjects freely available on the IRCAM website (LISTEN-V1) and of 35 subjects measured later with an increased spatial resolution in azimuth (LISTEN-V2). Measurements for the ARI database have been performed under semi-anechoic conditions and because of measurement artifacts, only 10 subjects have been retained and the spatial resolution in elevation has been decreased to 10°.

	Max ITD		Interaural distance		
Animal	LF	HF	Tax. models	Reported	Acoustical
Rat	165 μs	134 μs	2.7 cm	2.96 cm[1]	3.78 cm
Rabbit	319 μs	246 μs	3.2 cm	5.6 cm[2]	8.02 cm
Guinea pig	242 μs	184 μs	3.35 cm	4.94 cm[3]	5.02 cm
Chinchilla	293 μs	240 μs	3.9 cm	3.6 cm[4]	7.68 cm
Cat	335 μs	276 μs	5.2 cm	5.6 cm[5]	5.44 cm
Macaque	393 μs	310 μs	7.0 cm	10.4 cm[6]	8.36 cm

Table II: Overview of the animal ITD data. Maximal ITDs measured in low and high frequencies for the animal HRTFs. Interaural distances are the distances measured between the ear canal entrances of the taxidermized models (Tax. models), or the value as reported in previous studies (Reported), or twice the acoustical head radius (Acoustical, estimated from ITDs). References: [1](Koka *et al.*, 2008), [2](Kim *et al.*, 2010), [3](Greene *et al.*, 2014), [4](Lupo *et al.*, 2011), [5](Roth *et al.*, 1980), [6](Spezio *et al.*, 2000) .

List of Figures

765 Figure 9 Comparison of measured ITD range with anatomy. (a) Maximal measured
766 low-frequency ITD as a function of half the interaural distance measured on
767 the taxidermist models. Predictions are shown for Kuhn's formula (solid). (b)
768 Acoustical head radius estimated with Woodworth's formula (see Text) vs.
769 half the interaural distance measured on the taxidermist models (dashed line:
770 diagonal). Legend: rb: rabbit; ch: chinchilla; gp: guinea pig; m: macaque;
771 rt: rat; c: cat. 25