Timing cues for azimuthal sound source localization

Victor Benichoux

Timing cues for azimuthal sound source localization

Victor Benichoux

Equipe Audition
Laboratoire de Psychologie de la Perception
Ecole Normale supérieure, Paris Descartes, CNRS
Paris, France

Pour obtenir le grade de
Docteur de l'université Paris Descartes

Jury:

Steve Colburn	Reviewer
Daniel Tollin	Reviewer
Alan Palmer	Examiner
Alain Destexhe	Examiner
Jean-Pierre Nadal	Examiner
Frederic Alexandre	Examiner

Supervisor:
Romain Brette

Remerciements

Mes premiers remerciements s'adressent tout naturellement à mon directeur, Romain, pour m'avoir pris sous son aile il y a quatre ans déjà. Il a partagé avec moi sa science, ses opinions, et quelques verres, il m'a emmené tout autour du monde, et m'a toujours soutenu. Je n'ai sûrement pas réussi à aller au bout de tout ce qu'il voulait que l'on fasse. Mais ce manuscrit c'est d'abord pour lui que je l'ai écrit.

Je tiens aussi a remercier tous les membres du jury qui ont accepté de venir depuis parfois très loin, Alain Destexhe, Jean-Pierre Nadal, Frédéric Alexandre, Alan Palmer, Steve Colburn et Daniel Tollin. Tout particulièrement à ces deux derniers, à qui incombe la lourde tâche de lire ce manucrit. Un grand merci va aussi à Philip Joris, car sa collaboration et ses conseils ont été un élément déterminant de ce projet.

Une thèse est un travail d'équipe, dans mon cas structurée autour de Romain, dont les membres qui méritent ces remerciements tout autant que lui. A Bertrand, mon ami, pour tes chants, tes blagues, ta bonne humeur, ton soutien, ta femme, ta musique, ton fils, tes conseils, t'es gentil. Pour Marc aussi, qui a bien voulu rester pas loin de nous le temps de m'aider à finir ma thèse. A Jonathan, qui a eu le bon gout d'importer "fat friday" à l'Équipe Audition. A Marcel, même s'il a failli à m'apprendre l'Allemand, pour les innombrables moments passés ensemble à discuter. A tous les autres de l'*entourage*: Yi, Martin, Nelson, mais aussi les plus anciens que l'on oublie pas, Dan bien sûr, mais aussi Jonathan et Cyrille. Aux *petits* qui ont partagé notre dur quotidien en stage: Marie, Betty, Francois, Aloïs, et bien sûr Vincent.

Cette thèse n'a aussi été possible que grâce à la totalité de l'Équipe Audition. Merci donc à Christian, *le patron*, pour la philosophie déclamée le matin au labo, l'épigraphe de ce manuscrit, et les conseils donnés sur des trottoirs ou entre deux couloirs. Je tiens aussi à remercier les autres chercheurs du laboratoire qui sont toujours disponibles et de bon conseil: Daniel, Alain, Shihab. À Marion et Agnès, pour m'avoir donné l'esprit de la Love Team, que j'essaye maintenant de transmettre. Merci à nos anglo-saxons, Trevor, my very special kind of geek, Tim, for the Branson pickle and the "Perrier". Merci aussi à Bernhard, Claire, Dorothée, et tout particulièrement à Claire et Axelle pour leur bonne humeur, leur soutien, et m'avoir écouté râler envers et contre tout.

Cette thèse est aussi une tentative désespérée de rendre les parents fiers du travail de leur fils. Pour les remercier de m'avoir toujours soutenu dans ma voie et dans ma vie. Un grand merci à eux donc: ainsi qu'à ma famille en gńéral, qui m'ont toujours entouré et soutenu, comme je sais qu'il le feront toujours. Voilà ce que j'ai fait pendant tout ce temps. Merci à ma soeur Fanny, Pierre, les p'tis n'veux, Mathieu et Romain. À Alexis, parceque ta presence me manque souvent, et parcequ'on à finalement réussi

à pas avoir exactement le meme titre de thèse.

Ces trois ans de thèse, c'est trois ans d'une autre histoire, écrite avec mon autre famille: la coloc. C'est trois ans d'amitié en plus avec Edouard, Mehdi, Maxence, leurs dulcinés et tous les autres. C'est trois saisons à porter le maillot du FC DEC, avec un mot tout particulier pour Alexandre (pour qui-tu-sais). C'est trois ans de glaère partagée encore avec mes amis du Master, Baptiste, Guillaume, Jean-Remy. Enfin, merci à ma Mathilde de m'avoir soutenu, cette thèse c'est aussi un peu la sienne. Vivement demain!

Résumé

La localisation des sources en azimuth repose sur le traitement des différences de temps d'arrivée des sons à chacune des oreilles: les diffrences interaurales de temps ("Interaural Time Differences" (ITD)). Pour certaines espèces, il a été montré que cet indice dépendait du spectre du signal émis par la source. Pourtant, cette variation est souvent ignorée, les humains et les animaux étant supposés ne pas y être sensibles. Le but de cette thèse est d'étudier cette dépendance en utilisant des méthodes acoustiques, puis d'en explorer les conséquences tant au niveau électrophysiologique qu'au niveau de la psychophysique humaine.

A la proximité de sphères rigides, le champ sonore est diffracté, ce qui donne lieu à des régimes de propagation de l'onde sonore différents selon la fréquence. En conséquence, quand la tête d'un animal est modélisée par une sphère rigide, l'ITD pour une position donnée dépend de la fréquence. Je montre que cet effet est reflété dans les indices humains en analysant des enregistrements acoustiques pour de nombreux sujets. De plus, j'explique cet effet à deux échelles: localement en fréquence, la variation de l'ITD donne lieu à différents délais interauraux dans l'envelope et la structure fine des signals qui atteignent les oreilles. Deuxièmement, l'ITD de sons basses-fréquences est généralement plus grand que celui pour des sons hautes-fréquences venant de la même position.

Dans une seconde partie, je discute l'état de l'art sur le système binaural sensible à l'ITD chez les mammifères. J'expose que l'hétérogénéité des réponse de ces neurones est prédite lorsque l'on fait l'hypothèse que les cellules encodent des ITDs variables avec la fréquence. De plus, je discute comment ces cellules peuvent être sensibles à une position dans l'espace, quel que soit le spectre du signal émis par la source. De manière générale, j'argumente que les données disponibles chez les mammifères sont en adéquation avec l'hypothèse de cellules selectives à une position dans l'espace.

Enfin, j'explore l'impact de la dépendence en fréquence de l'ITD sur le comportement humain, en utilisant des techniques psychoacoustiques. Les sujets doivent faire correspondre la position latérale de deux sons qui n'ont pas le même spectre. Les résultats suggèrent que les humains perçoivent des sons avec différents spectres à la même position lorsqu'ils ont des ITDs différents, comme prédit part des enregistrements acoustiques. De plus, cet effet est prédit par un modèle sphérique de la tête du sujet.

En combinant des approches de différents domaines, je montre que le système binaural est remarquablement adapté aux indices disponibles dans son environnement. Cette stratégie de localisation des sources utilisée par les animaux peut être d'une grande inspiration dans le dévelopment de systèmes robotiques.

Summary

Azimuth sound localization in many animals relies on the processing of differences in time-of-arrival of the low-frequency sounds at both ears: the interaural time differences (ITD). It was observed in some species that this cue depends on the spectrum of the signal emitted by the source. Yet, this variation is often discarded, as humans and animals are assumed to be insensitive to it. The purpose of this thesis is to assess this dependency using acoustical techniques, and explore the consequences of this additional complexity on the neurophysiology and psychophysics of sound localization.

In the vicinity of rigid spheres, a sound field is diffracted, leading to frequency-dependent wave propagation regimes. Therefore, when the head is modeled as a rigid sphere, the ITD for a given position is a frequency-dependent quantity. I show that this is indeed reflected on human ITDs by studying acoustical recordings for a large number of human and animal subjects. Furthermore, I explain the effect of this variation at two scales. Locally in frequency the ITD introduces different envelope and fine structure delays in the signals reaching the ears. Second the ITD for low-frequency sounds is generally bigger than for high frequency sounds coming form the same position.

In a second part, I introduce and discuss the current views on the binaural ITD-sensitive system in mammals. I expose that the heterogenous responses of such cells are well predicted when it is assumed that they are tuned to frequency-dependent ITDs. Furthermore, I discuss how those cells can be made to be tuned to a particular position in space irregardless of the frequency content of the stimulus. Overall, I argue that current data in mammals is consistent with the hypothesis that cells are tuned to a single position in space.

Finally, I explore the impact of the frequency-dependence of ITD on human behavior, using psychoacoustical techniques. Subjects are asked to match the lateral position of sounds presented with different frequency content. Those results suggest that humans perceive sounds with different frequency contents at the same position provided that they have different ITDs, as predicted from acoustical data. The extent to which this occurs is well predicted by a spherical model of the head.

Combining approaches from different fields, I show that the binaural system is remarkably adapted to the cues available in its environment. This processing strategy used by animals can be of great inspiration to the design of robotic systems.

La science est affaire d'ennui et de bon goût.

C. Lorenzi

Contents

Introduction

Biological systems constantly interact with a complex, fluctuating environment. The environment itself is govern by a number of *physical rules*, for example on how acoustical pressure waves propagate in space. Those rules impose some structure on the flow of sensory inputs that an animal perceives. Gibson proposed that perception consists in the detection of *invariant* structures in sensory flows: it is the concept of *ecological computation* [Gibson, 1986]. For example, when a sound source is present at a given position in space, the acoustic signals reaching the ears (the two monaural signals) are slightly different. That is, both contain a version of the source signal, only filtered in different ways by the body, head and ears of the subject. In particular, whatever sound is presented at the right of a subject, it will arrive at the left ear *delayed* by a certain amount, that depends on the position of the source. The *position* of a sound source therefore is a *relationship* between the two sensory inputs, here the sounds at the ears. It is a *structure* invariant on the signal emitted by the sounds: the position of a speaker does not depend on what he/she says. In this context, the Interaural Time Difference (ITD) is a *cue* to source location because it can be mapped to a position. In reality, the structure imposed by our body and environment on the sounds reaching our ears is much more complex. That is to say that the signals reaching our ears are not exactly delayed versions of one another. Rather, because the sound waves reflect and diffract on our head, the monaural signals are linked by a more complex structure. This structure includes differences in level: the sound at the contralateral ear is also attenuated with respect to the ipsilateral ear. The present thesis, however, is only concerned with the temporal structure of the binaural inputs, the ITDs.

It is commonly assumed that, for a given position in space, the ITD is the same quantity whatever signal is emitted by the source. Yet, the presence of the head in between the ears make the temporal structure of binaural inputs much more complex. As a consequence, the ITD for a given position depends on the frequency of the signal emitted by the source. While this additional complication is oftentimes regarded as noise, or unwanted variability, I argue that it is added richness. Indeed, because ITD varies systematically across frequencies and positions in a way that is imposed by *phys-*

ical laws, those variations provide *more information* about the position of the sound source. Therefore, the binaural structure is both more complex and more informative than was previously believed. This observation is the starting point of this project, which will be dedicated to understanding the consequences of this added complexity on our understanding of the binaural system. The problematics of this thesis are therefore: how do mammalian neural systems integrate frequency-dependent ITD cues so as to produce a sense of sound location that does not depend on the source signal? An important insight of this work may be formulated as follows: if a sensory system is tuned to a complex stimulus, then its response to simple stimuli may appear complex. If, on the other hand, the system is probed with the complex stimulus it expects, then responses will appear much simpler. Necessarily, such an endeavor requires the use of three complementary approaches:

- *Acoustical* analysis of the temporal structure of binaural inputs. This includes understanding the *physical* basis of this structure, assessing its effect in terms of *signal processing*, and *measuring* this structure across different species.
- *Neurophysiological* study of the binaural system that is involved in the processing of ITDs, and *modeling* of this function.
- *Psychoacoustical* evaluation of human listeners' ability to detect a frequency-dependent temporal structure in sounds presented over headphones.

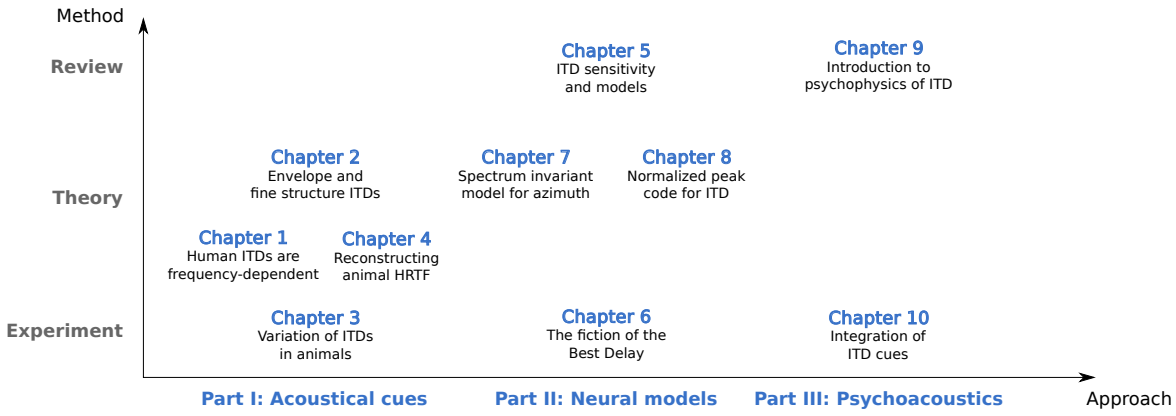

Figure 1: Organization of the manuscript. The chapters are represented according to their approach, and the method used (state-of-the-art, theoretical and experimental chapters).

Organization of the manuscript

The thesis is divided into three parts that approach the problem of ITD-based sound source localization through different angles. The organization of the thesis is pictured on Fig. 1, wherein the different chapters are represented according to their approach (acoustics, modeling, behavioral studies), and their method (review chapters, experimental chapters, theoretical chapters). First, acoustical timing cues are computed and discussed in the first four chapters (1, 2, 3 and 4). The neural basis of the ITD processing system in the mammalian brainstem is first introduced in Chapter 5, and then three modeling studies are presented (Chapters 6, 7 and 8). Finally, the psychoacoustics of sound localization is reviewed in Chapter 9, and a behavioral study of lateralization is presented on Chapter 10.

Part I: Studies of acoustical timing cues

Chapter 1 starts with the presentation of acoustical models of ITD. This provides insight into the physical rules that create frequency-dependent ITDs for the same position. Then, I record and compute ITDs for human subjects as a function of the frequency of the input for many positions on the sphere. Those recordings confirm that human ITDs display a coupled frequency- and position dependence, that can be predicted on the basis of morphological parameters. The concept of a frequency-dependent ITD is then developed in Chapter 2, where a signal processing approach is used to explain the effect of such ITDs. At this point, it is shown that envelope and fine structure ITD were different when the ITD varies, providing two cues to localization. An estimation method for those cues is provided, and they are reported using human data of the

first chapter. Chapter 3 uses the same approach to report ITD cues available to a wide range of naturalized animals of different species (cat, rabbit, owl, guinea pig, chinchilla, macaque, and rat). The range of possible envelope and fine structure ITDs is then discussed. Finally, Chapter 4 presents a numerical method to simulate animal HRTFs from raw photographs alone.

Part II: Models of sound localization

This part is introduced by a review of the current knowledge on the binaural ITD sensitive system in mammals (Chapter 5). It develops both how single binaural neurons can be made to be *tuned* to a particular ITD, and how this can be understood in terms of computational models. Then, the current views on how the azimuthal position of the sound source is represented across populations of heterogenously-tuned neurons are discussed. The next chapter, Chapter 7, is the main result of this thesis. Using electrophysiological data gathered by Philip Joris, I argue that binaural ITD-sensitive cells in the cat IC are in fact tuned to a frequency-dependent ITD, rather than a constant ITD as is commonly assumed. Furthermore, I show that this complex ITD sensitivity reflects the variation of ITD found in acoustical data. Next, Chapter 7 develops on the insights of the previous chapter to formulate a complete model of spectrum-invariant sound localization through ITD. The acoustical data is used to fit a model consisting of cells receiving inputs with mismatched frequency from both sides, and different axonal propagation delays. Finally, a study is presented that deals with the problem of coding space with ITD-sensitive neurons (Chapter 8). I argue that the available data in mammals is in fact consistent with the hypothesis that cells encode single positions in space.

Part III: Psychoacoustics of ITD processing

In this last part, the psychoacoustics of sound localization is introduced, and discussed in the context of frequency-dependent ITD cues Chapter 9. Current arguments against the perceptual salience of frequency-depedent ITDs in binaural inputs are reviewed and discussed. Finally, preliminary results from a behavioral study of ITD perception humans are presented. I argue that they suggest that the frequency-dependence of ITD is relevant to human localization.

Part I

Quantifying binaural timing cues

This part of the manuscript is concerned with the notion of *binaural disparities*, the differences between the signals that reach the ears. Intuitively, a sound to the left of a listener will be less intense at the right ear than at the left ear, and also arrive later. Because these disparities depend on the position of the sound source, they are essential to sound localization. Using experimental and theoretical models of the acoustics of the animal, we study the binaural *timing* disparities in animals and humans alike. They correspond, intuitively, to the additional time it takes the sound wave to reach the *contralateral* ear (the ear opposite to the source) with respect to the *ipsilateral* ear.

Put in other words, the acoustical environment and the body of the animal impose a certain *structure* on the acoustical signals reaching the ears. Namely, the vibration pattern of the contralateral tympanon is a *delayed* and *filtered* version of the ipsilateral vibration pattern. Compelling evidence exists showing that both humans and other mammals use these structure to localize sound sources. Moreover, the Interaural Time Differences (ITDs), seem to play an important role in the perception of position of low-frequency sounds (below 3 kHz) [Wightman and Kistler, 1992]. In order to make sense of the data acquired in behavioral of electrophysiological studies, it is important to first quantify which structure is imposed by the acoustical environment and body on the signals entering the ears.

In this part of the dissertation, I measure ITDs as a function of position and frequency, over a wide array of animal and human subjects. In Chapter 1, I show that ITD depends on the frequency of the signal using acoustical recordings in human subjects. I explore the acoustics of this phenomenon, and explain that this arises because of the diffraction the sound wave in the vicinity of the head. In Chapter 2, implications of the frequency-dependence of ITD on the structure of signals reaching the ears are studied. I show that in fact the ITDs in the fine structure and the envelope of the signal are different, and bear information about the location of the sound source. In Chapter 3, the findings of the first chapter are replicated on naturalized animal subjects, and the same analysis is applied. Frequency-dependent envelope and fine-structure ITDs are also found in all recordings, which is quantified using the metrics of the previous chapter. Finally, Chapter 4, deals with the simulation of animal directional acoustic filters for reseach. We expose a method that relies solely on raw photographs and affords efficient computation of acoustical head-related directional filters. The method is then validated using recordings and measures of binaural cues. As an example application of this method, we study the posture-dependence of binaural cues in the cat.

Chapter 1

Human ITDs are frequency-dependent

Contents

In this chapter I characterize the coupled frequency- and position-dependence of ITD in humans subjects, and understanding its physical basis. The frequency-dependence of human ITDs for a given position has already been measured [Kuhn, 1977], and modeled using a spherical model of the head [Duda and Martens, 1998]. I present recordings of directional acoustic filters performed on many human subjects, alongside analysis of the spherical head model to assess the coupled dependence in frequency and position of the ITD.

The variation of ITD across frequencies can be understood as the result of sound wave diffraction around the head, due to the head's spherical shape [Duda and Martens, 1998]. I review known approximations of broadband ITD [Kuhn, 1977], and explain their meaning in the context of physical acoustics. I show that diffraction of sound waves around the head is the primary source of variability of ITD for a given source position. Therefore, for a given position the ITD is in always smaller for high-frequency sounds than for low frequency sounds. This is a quasi-monotonic trend where the transition occurs around a frequency related to the size of the head.

Insights from the spherical head model help understand the effect morphological differences amongst subjects have on the magnitude and variation of ITD cues. Compensating for those differences, I was able to study the ITDs across a population of subjects with different head sizes. I compute ITDs from existing and specifically-recorded acoustic data on human subjects in anechoic settings. I show that the variations of ITD across frequency for a given position is not attributable to measurement noise, and is consistent across subjects.

For a given position, we argue that ITD variation over the spectrum takes place at two scales. Because the difference between the low- to high-frequency ITDs are big, when compared to the human ability to discriminate ITDs, we argue that global variations of ITD must be relevant to the perception of source location in humans. Furthermore, the ITD varies strongly inside the frequency range of ITD dominance, therefore the ITD cannot be considered constant across the bandwidth of single auditory filter.

1.1 Introduction

For a given listener, ITDs are modified by the spatial position of the sound source, *i.e.* its direction [Blauert, 1997] and its distance [Duda and Martens, 1998, Brungart and Rabinowitz, 1999]. In essence, this variation in the structure of the inputs is why ITD is an primary directional localization cue. For example, when a sound is presented with conflicting cues (e.g. level differences, ILDs), it is perceived at the position indicated by the ITD, provided that it has energy below 2kHz. Therefore, in this range ITD is said to *dominate* the perception of laterality [Wightman and Kistler, 1992].

Yet, the spatial position of the sound source it is not the only parameter affecting the ITD in a given species. Among human listeners, morphological parameters [Fels and Vorlander, 2009, Duda et al., 1999a, Algazi et al., 2001b, Treeby et al., 2007, Xie and Zhang, 2010, Jo et al., 2008, Nicol, 2010] such as head radius and shape, torso, ear positions, and even hair, have been shown to influence ITDs. As an example, the morphological parameter that influences ITDs the most is the head radius: bigger heads imposing bigger ITDs than smaller ones. For a given listener, it was shown that early reflections on the ground the subject stands on [Gourevitch and Brette, 2012] creates variations in ITD at different frequencies. Similarly, it is thought that early reflections on the torso is the cause of bigger absolute ITDs in humans [Kuhn, 1977]. Finally, for a given listener and a given spatial position, ITDs are frequency-dependent in the low-frequency range (¡1.5kHz) because of the diffraction of sound waves occurring in the vicinity of the head [Kuhn, 1977].

The variation of ITD across frequencies for a given position was first experimentally shown in [Abbagnaro et al., 1975] in humans and a sophisticated manikin, and soon after by [Kuhn, 1977] who also provides estimates for low- and high-frequency ITDs. In this study, ITDs have been measured in the horizontal plane for an artificial human head and correlated successfully with a theoretical approach based on a spherical head. For a rigid sphere, the fact that the inter-aural phase differences (IPDs) are not linear with the frequency (which leads to frequency-dependent ITDs) had been established at the early beginning of the XX$^{\text{th}}$ century [Stewart, 1914, Hartley and Fry, 1921]. Later, [Wightman and Kistler, 1989] showed experimentally the frequency-dependence of the ITDs for real human subjects. In their structural model for binaural sound synthesis, [Brown and Duda, 1998b] also rendered these ITD variations with the frequency. [Brungart and Rabinowitz, 1999] show theoretically and experimentally that for humans ITDs were frequency- but not much distance-dependent. In the animal realm, frequency-dependence of ITDs has been highlighted in many species, including cat [Roth et al., 1980b], the Rhesus monkey [Spezio et al., 2000b], the guinea pig [Sterbing et al., 2003a],

the Mongolian gerbil [Maki and Furukawa, 2005a], the rat [Koka et al., 2008a], and the Chinchilla [Koka et al., 2011a]. In the rabbit [Kim et al., 2010a], ITDs decrease with decreasing distance and with increasing frequency.

The precise coupling between the frequency- and direction-dependence of ITDs remains an open question that can be addressed both theoretically for a rigid sphere and experimentally on human acoustical recordings. Unlike the influence of morphological and spatial parameters on ITDs, the variation with frequency of ITD has received significantly less attention. One of the reasons is that measuring frequency-dependent ITDs is a tedious process. A common way to measure acoustical ITDs was the *pulse measurement* method [Abbagnaro et al., 1975]. A signal is presented in space and recorded at the ears, the ITD is obtained by computing the difference of the onset-times of each monaural recording Therefore, to compute ITDs, [Abbagnaro et al., 1975, Kuhn, 1977] had to perform one measurement per frequency point. In contrast, modern methods allows one to compute the complete acoustical filter in one recording [Farina, 2000, Hammershi et al., 1992]. For each position in space, a pair of Head Related Transfer Functions (HRTFs) (one for each ear) fully describes the acoustical transformation of sound by the head and body. Available computational power allows nowadays to simulate easily spherical-head HRTFs, and a great number of HRTFs [Blauert, 1997, Wightman and Kistler, 2005a] have been measured [Algazi et al., 2001b] on humans listeners for almost all directions. These data afford the possibility of studying the ITD across a wide population of subjects and conditions, while simulations of the head model provides insight into the physical processes at play [Feuillade, 2004].

Before we turn to the developments of this chapter, it is important to note that the definition of ITD taken here is the total time difference between the fine structures of the monaural signal at each frequency. This is the *total phase ITD*, which will be noted $\mathrm{ITD}_{\mathrm{p}}$, it is computed by dividing the unwrapped interaural phase by the frequency, as precisely put in Subsection 1.3.2. Much will be said as to the precise interpretation of this ITD in the next chapter.

First I introduce classical estimates of broadband ITDs (i.e. invariant on frequency), highlighting the acoustical hypotheses made to derive them. The spherical model is the simplest acoustical model that accounts for the diffraction of the sound wave around the head. This phenomenon is responsible for the variation of ITD in the low-frequencies. Given this theoretical basis for the variation of ITD across the spectrum, we turn to measuring human HRTF data in an anechoic setting. I expect that, consistent with previous observations, the ITD in humans is frequency-dependent. I generalize this result across populations of human subjects, an approach made possible by normalizing out the contribution of morphological parameters of each subject. I then quantify the

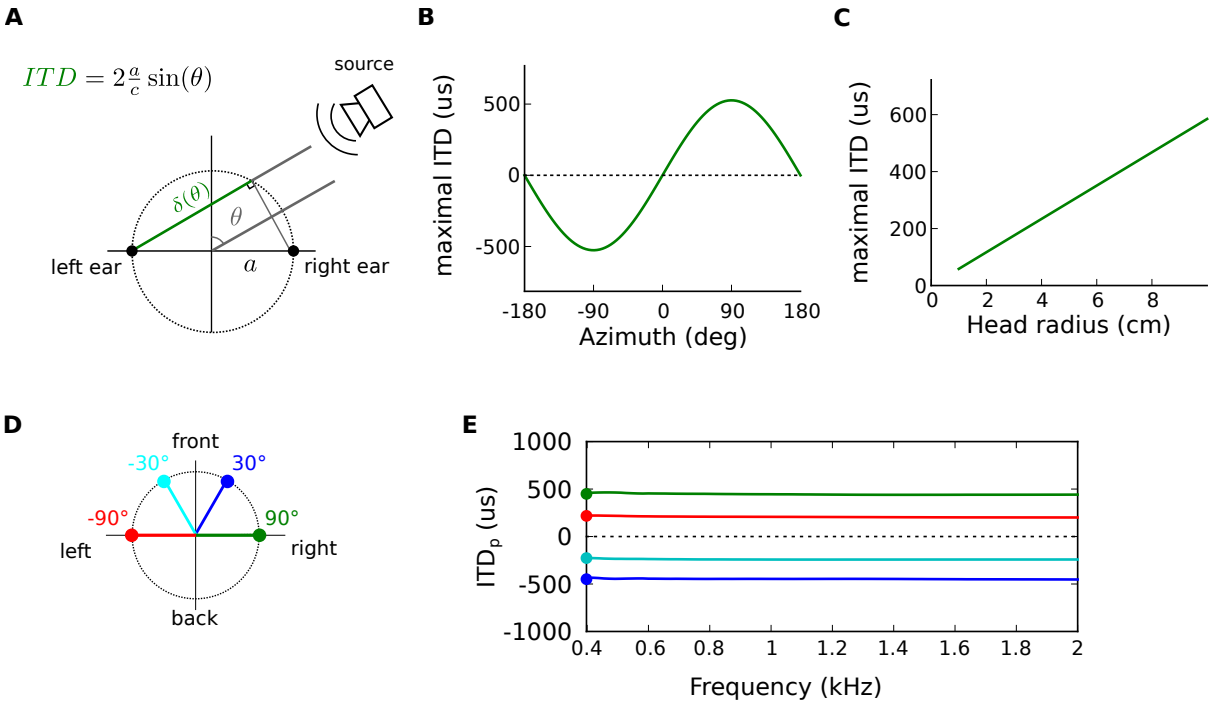

Figure 1.1: **A** Depiction of the situation: two microphones are positioned at a distance $2a$ of each other. A source is presented a azimuth θ, The additional path length $\delta(\theta)$ (green) is obtained using simple trigonometry. **B** ITD as a function of position as predicted by the two-mics model. **C** Influence of the head radius a on the maximal ITD (at 90°) in the two-mics model. The second row shows results of recordings made in anechoic settings with two mics positioned 15 cm apart. **D** Positions reported on **E** for the recordings in the two-mics setup. **E** Measurements of ITD_p as a function of frequency for a two-mics setup. Colored dots on the y-axis report the estimates according to Equation 1.1.

potential computational and perceptual impact of this phenomenon, and argue that the variation of ITD happens at two scales. Overall my results suggest that the frequency-dependence of ITD is a ubiquitous feature of binaural acoustics, that bears worthy information about the position of the sound source.

1.2 Models of ITDs

1.2.1 A geometrical model of ITD

A simple estimate of ITD as a function of position and head size can be inferred from a geometrical model. This model I term the *two-mics* model, and it is depicted in Fig. 1.1, **A**, for a source at θ azimuth (here $\theta < 0$). It is based on the idea that ITDs originate from the additional path-length the sound has to cover to reach the contralateral ear, $\delta(\theta)$ (green segment on **A**). Assume that the ears are two receivers

separated by a distance of $2a$ where a is the half of the interaural distance, i.e. the head radius of the subject (Fig. 1.1). It is clear that $\delta(\theta) = 2 \times a \sin(\theta)$, thus dividing by the celerity of sound in the air, the $ITD_{geom}(\theta)$ the geometrical ITD is obtained (θ in radians):

$$ITD_{geom}(\theta) = 2 \times \frac{a}{c} \sin(\theta) \tag{1.1}$$

This formula for ITD displays two interesting dependencies. First it is a sine function of the azimuth, as plotted in Fig. 1.1 (plain green line), which means that it can be uniquely mapped back to the azimuth of the sound source. Secondly, it includes a morphological parameter: the head radius a, and predicts that the maximal ITD (the ITD for $\theta = 90°$) is linearly dependent on this parameter, as shown on Fig. 1.1 (**C**, green line). Hence smaller animals experience smaller ITDs than bigger ones for the same positions, which is an intuitive result.

This geometrical approach for computing the ITD is acoustically justified when some assumptions are met. The ears are modeled as two omnidirectional point receivers in space separated by $2a$. Real microphones are not omnidirectional, let alone human ears. Furthermore, the environment is empty of other object than a point source. This source also is situated far enough from the receivers so that the emitted wave can be modeled as propagating as a plane wave. In this case only, the wave front is a straight line (as opposed to a circle centered on the source), propagating orthogonally towards the receiver. Provided that those assumptions are met, the argument and Equation 1.1 are physically valid.

Yet, these assumptions obliterate many parameters influencing the ITDs. Because the ears are point receivers, there are no frequency-dependent ILD cues: informational dips and notches in the mid to high frequencies. Here no influence of the azimuth or elevation of the source on the monaural responses of the ears is taken into account. Furthermore, the head is considered *acoustically transparent*: no acoustical energy is absorbed or reflected away by its surface. Also, because the wave is assumed to be planar, there is no influence of distance on the ITD. Finally, the environment is free of any other acoustical object. This kind of environment is usually known as *anechoic*, which can be achieved by covering walls with carefully shaped acoustic foam that absorbs all acoustic energy. Because they are neutral environments, ITDs are commonly measured in anechoic setups, which ensures that data can be compared across different setups. Yet, in natural environments the ITD depends on a multitude of other environmental parameters, for example ground reflections [Gourevitch and Brette, 2012].

Acoustical measurements confirm that under these constraints, the two-mics model (Equation 1.1) accurately predicts the ITD. I performed acoustical measurements with

two miniature microphones separated by $d = 15cm$ placed in an anechoic chamber Fig. 1.1. The setup was common with the other acoustical measurements performed for this project, that are described in Section 1.3. From those measurements, the *total phase* ITD, ITD_p, at each frequency and position was computed from HRTF data. The definition of ITD_p can be found in Subsection 1.3.2, and it was computed according to the method described there. **E** displays the ITD_p as a function of frequency for four positions depicted in **D**. We observe that the total phase ITD (for $\theta = 90°$, green line) is in very good adequation with the estimates provided by the two-mics model (colored dots on the y-axis, Equation 1.1) Namely, measured ITDs are independent on frequency for all positions.

1.2.2 Spherical head model

Assuming a spherical model of the human head, a more realistic model of acoustical ITDs is obtained. The head is assumed to be a rigid sphere, which reflects all incoming acoustical energy away, and the ears lie on a diameter of this sphere. Incoming and reflected waves interact, creating interferences and frequency-dependent effects. An example is the frequency-dependence of ILD, which is classically explained using insights from a spherical model. ILDs are observe to be quasi non-existing for low-frequency sound sources (say, below 600 Hz)

Spherical head models have also been extensively studied in the context of the ITD [Blauert, 1997, Duda and Martens, 1998, Brungart and Rabinowitz, 1999, Algazi et al., 2001b, Treeby et al., 2007]. It is also the simplest acoustical model that accounts for the frequency-dependence of ITD. It can be solved, using a development provided by lord Rayleigh [Rayleigh and Lodge, 1904], providing an efficient HRTF simulation tool. In what follows, we recall main results and intuitions about ITD gained from the spherical head model. A more complete development of the physics of model can be found in Appendix 1.B.

ITDs for a rigid sphere

Low-frequency approximations Low-frequency analytic developments can be obtained from the equations in Appendix 1.B. Keeping only the first two terms in the spherical harmonic development of the solution (Equation 1.B.3), the low-frequency approximation of the ITD [Kuhn, 1977, Duda and Martens, 1998] is obtained for a sound source positioned at θ azimuth (in radian) on the horizontal plane (cases with nonzero eleva-

tions are reported in Appendix 1.B):

$$\mathrm{ITD}_{\mathrm{LF}}(\theta) = 3\frac{a}{c} \times \sin(\theta) \tag{1.2}$$

This estimate has the same form as Equation 1.1, it is a sine function of the azimuth and a linear function of head radius.

High-frequency approximations High frequency approximations are harder to obtain because the convergence of the solution is slower in high frequencies (see Appendix 1.B) Equation 1.B.3). Therefore one cannot rely on the first terms of the infinite series as we did in the low-frequency case. A solution was provided by [Kuhn, 1977], formulating the equation in terms of *creeping waves*. This is the classical *Woodworth model* of the ITD [Kuhn, 1977, Blauert, 1997, Larcher and Jot, 1997]:

$$\mathrm{ITD}_{\mathrm{HF}}(\theta) = \frac{a}{c} \times [\theta + \sin(\theta)] \tag{1.3}$$

This limit is best understood geometrically, wherein the computation of the contralateral path is changed. The wave first approaches the sphere tangentially and covers the rest of the path on the surface of the sphere This is shown in Appendix 1.B, Fig. 1.14, **A**. The plain blue line shows $\delta'(\theta)$ the additional path the creeping wave travels to reach the contralateral ear.

In Equation 1.3, for positions close to the median plane ($\theta \ll 1$), we can approximate the sine term, $\sin(\theta) \approx \theta$. Therefore, in this limit the equation can be put in the same form as the low-frequency approximation of Equation 1.2:

$$\mathrm{ITD}_{\mathrm{HF}}(\theta) = 2\frac{a}{c} \times \sin(\theta) \text{ for } \theta \ll 1 \tag{1.4}$$

Which is incidentally, the same value of ITD as that predicted in the case of a a transparent head (two-mics model, Equation 1.1). Furthermore, the low-frequency ITD for a given position (near the median plane) is 1.5 greater than the high-frequency limit $\mathrm{ITD}_{\mathrm{HF}}$:

$$\frac{\mathrm{ITD}_{\mathrm{LF}}(\theta)}{\mathrm{ITD}_{\mathrm{HF}}(\theta)} \approx \frac{3}{2} \text{ for } \theta \ll 1 \tag{1.5}$$

For all positions, Equation 1.2 and Equation 1.3 provide ITDs that are larger or equal than ITDs provided by Equation 1.1 for the acoustically transparent head. Therefore the diffraction phenomenon overall increase ITDs, and this increase is more pronounced in the low-frequency range, over which ITD cues dominate lateralization (below 2 kHz).

Normalized time and frequency

Introducing a non-transparent head between the ears produces frequency-dependent responses at each ear. Before we introduce the results, it is necessary to explain the *normalized frequency scale* used to present them. The solution to the spherical model equation, when the diameter of the head is the unit of space, is independent on frequency. This means that a change in frequency translates into a change of the sphere radius. In essence, the response of the spherical model for a given frequency and head size (f, a) is the same as for $f/2, 2a$ and so forth. This motivates the introduction of a *normalized frequency* with which to describe the results.

For a given head radius a, the "real" frequency f is linked to the *normalized* frequency f_n via:

$$f = f_n \times \hat{f} \quad \text{with} \quad \hat{f} = \frac{c}{2\pi a} \tag{1.6}$$

Where c is the celerity of sound waves in the air ($c = 342 m.s^{-1}$), and thus $\hat{f}$ is frequency of a sound wave having a wavelength of exactly one sphere circumpherence.

The time dimension is also scaled with respect of the head size via the speed of sound. This means that in turn the ITD, which is a duration, will be reported normalized. The "real" ITD can be obtained via the relation:

$$\mathrm{ITD_n} \times \frac{a}{c} = \mathrm{ITD} \tag{1.7}$$

To conclude with, and ease the interpretation of the figures in the rest of this chapter, I provide the following relations to obtain the "real" time and frequencies. Assuming a radius of $a = 10.3$ cm, roughly corresponding to a human (see Subsection 1.3.2), the folowing relations hold:

$$\begin{aligned} \mathrm{ITD}_n = 1 \quad &\Longrightarrow \quad \mathrm{ITD} = \frac{a}{c} \approx 300\ \mu\mathrm{s} \\ f_n = 1 \quad &\Longrightarrow \quad d = \frac{c}{2\pi a} \approx 520\ \mathrm{Hz} \end{aligned}$$

Variation of ITD with frequency: diffraction effects

A spherical head model was computed in the steady-state regime (as described in Appendix 1.B), so as to assess the physical origin of the variability of ITD across frequencies. The results of $\mathrm{ITD_p}$ computation on this model (according to Appendix 1.B) are reported on Fig. 1.2, **B**, for a range of positions to the left of the subject ($\theta > 0$). ITD estimates from the equations in the previous section are reported and consistent with the $\mathrm{ITD_p}$computed in high and low frequencies. At all positions $\mathrm{ITD_p}$ is dependent on frequency, showing a decreasing trend between the two spherical estimates on **B**. The

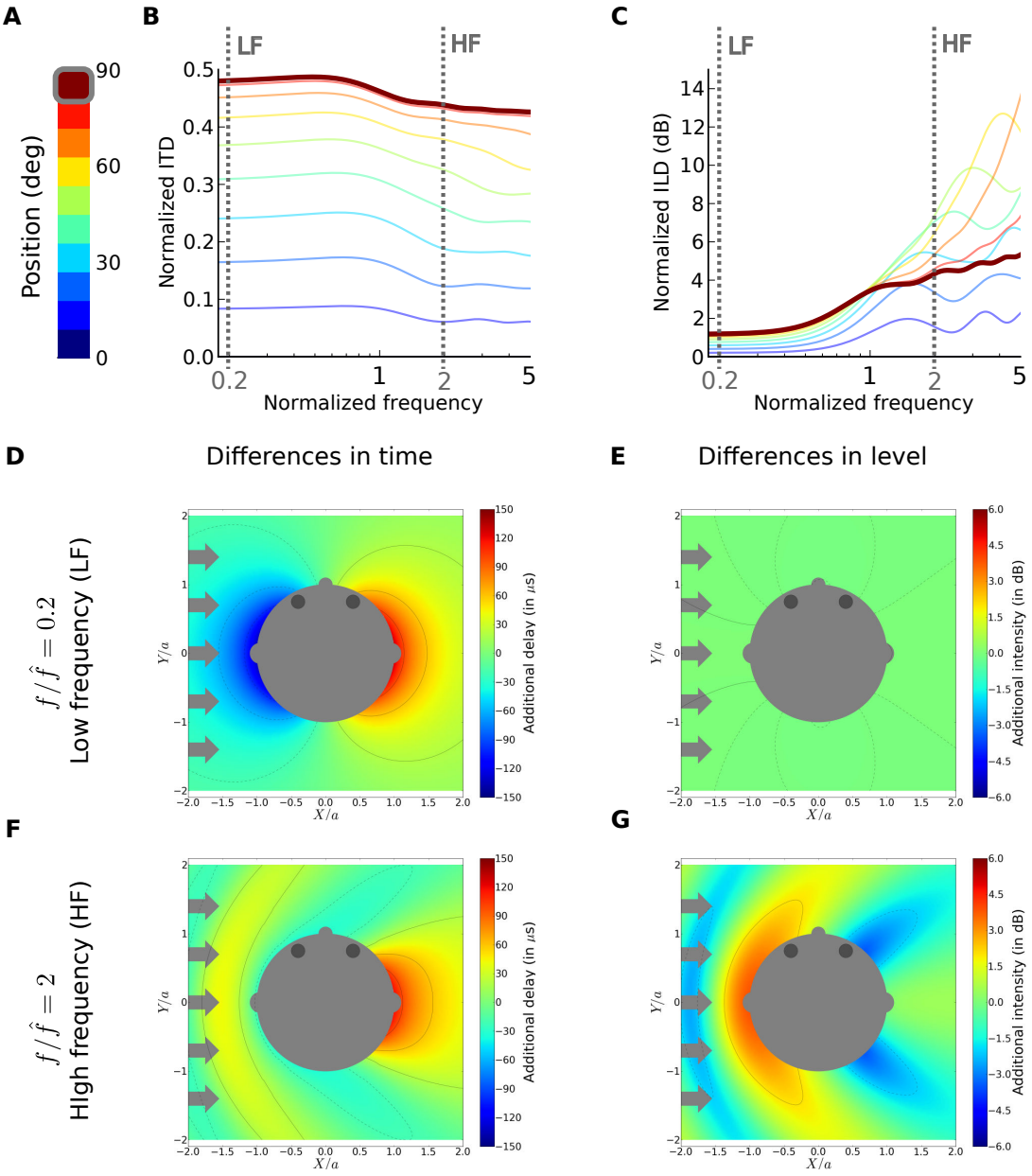

Figure 1.2: Acoustical origin of the binaural cues in a spherical head. The first row shows ITD_p, and ILD, computed for 10 positions between 0 and 90°(color code on **A**). **B**: normalized ITD_p as a function of f_n a different positions. The position $\theta = 90°$, simulated in the bottom four panel is thicker. **C** ILD as a function of normalized frequency at the same positions. Panels **D** through **G**: the left column shows the differences in time between the transparent and spherical head models (see text) and the rightmost panels show the differences in level. The first row (middle row) is the low frequency case with $f_n = 0.2$, and the last row is the high frequency case with $f_n = 2$. **D** additional phase, $f_n = 0.2$ **E** additional gain, $f_n = 0.2$ **F** additional phase, $f_n = 2$ **G** additional gain, $f_n = 2$

roughly sigmoidal shape of the curve is quite consistent across positions, as well as the frequency at which it transitions.

In addition to computing ITDs for a rigid sphere, I modeled the acoustical field near the head of the subject ([Feuillade, 2004] and Appendix 1.B. This helps to understand the acoustical phenomena at play, by comparing the simulation to predictions made in the absence of a head. We report the additional phase delay and additional gain that the wave undergoes at each point in space (and two frequencies), relative to the situation with a transparent head (the incidence field in Subsection 1.B). These differences are reported for two normalized frequencies 0.2 and 2 (bottom and middle rows), in time on the left panels and in level on the right panels. Perhaps this is better understood in gain, the plots **E** and **G** show the gain difference of the wave at this point in space when the sphere is present versus when it is absent. When it is positive, it indicates an increase in intensity due to the present of the head, as on **E**, at the left ear. Conversely, positive values of the additional phases reported on **D** and **F** stand for an advance in phase, and negative values a delay in phase. Note that advances in phase are possible only because the acoustical pressure field are computed in the *steady-state*.

On Fig. 1.2, **C** I show the ILD in decibels computed for a spherical model across frequencies and for various positions (**A**). As expected, the ILD is almost zero in low frequencies. This is because the head does not shadow the incoming wave. When simulated in the field near a spherical head, on Fig. 1.2, **E**, one can see that the difference with respect to the two-mics case is null. In high frequencies, on the contrary (**G**), the wave is shadowed and hence the level at the contralateral ear is greatly reduced. On Fig. 1.2, **G**, the field displays the so-called "bright spot" [Duda and Martens, 1998], the wave pressure is higher for positions completely opposite to the sound source. This is justified by the fact that the waves travelling around the head always arrive in phase at this point, creating constructive interferences.

As for ITDs, in high-frequencies (Fig. 1.2, **F**) the phase of the at the contralateral ear is lagging, which is consistent with any explanation based on acoustical rays (e.g. the two-mics model). At the ipsilateral ear, in high-frequencies (**F**), there is no noticeable difference, because there is no reflection on the surface of the sphere. When the frequency is lowered (**D**), diffraction comes into play: the incoming wave now is in advance of phase at the ipsilateral ear. At the contralateral ear, the situation appears similar in the two frequencies considered. Therefore, the ispilateral advance in phase creates the additional ITD in the low-frequencies in the spherical model. This is confirmed by the more detailed development of the spherical head model provided in Appendix 1.B.

Morphological parameters

In a spherical head model, the only morphological parameter of interest is the sphere radius a. It is the only space constant of the model, and it influences at which frequencies diffraction effects will start to play an important role. On Fig. 1.3 the ITD and ILD curves are shown for the same position but two different simulated head sizes, $a = 10cm$ for the green curve and $a = 5cm$ for the blue curve. The ITD is an increasing function of head size, because both the high frequency and the low frequency ITDs are higher for a bigger head (**A**). This relationship is linear, as shown on **C** where the low and high frequency estimates are recalled. Additionally, the size of the head changes the frequency at which the binaural cues transition from they low-frequency to high-frequency regimes. It is especially obvious from the ILD plots on **B**: the ILD curve for a bigger head is not merely a vertical stretch of that for smaller heads, it is rather a translated version of it along the frequency axis. Note that in our case, it is not a perfectly translated version of the ILD curve, because the source was modeled as coming from a source at a finite distance, in this case the ILDs in low-frequency increase, which is a cue to distance in for close sources [Duda and Martens, 1998].

The fact that binaural cues scale with frequency when the head size varies is not surprising from the physics of the model. An interesting measure of an ITD curve in this respect is the frequency of transition between high and low regimes of propagation. This transition starts to occur in the vicinity of $f_n = 1$, yet from Fig. 1.3, **A** the frequency of maximal ITD_p variation is in general closer to $f_n = 2$. Irregardless, the real frequency at which this transition occurs will be proportional to $\hat{f}$ (i.e. $f_n = 1$). The dependence of $\hat{f}$ on head size is plotted on **D** of Fig. 1.3. The frequency at which the ITD transitions between its low and high frequency regimes is proportional to $\hat{f}$ and thus, it is lower in animals with bigger heads. Finally, in all cases the transition frequency is well within the range in which the ITD is a dominant cue to sound source azimuth. For human listeners $\hat{f} \approx 600$ Hz, well below 1.5 kHz [Brown and Duda, 1998b, Wightman and Kistler, 1992], Fig. 1.3, **D**. It also holds for other animals: e.g. cats have heads that are roughly twice as small as a human's, their transition frequency is of about $1200Hz$. This is also in the range where cats use ITD cues for localization [Grothe et al., 2010].

I will now show the results of acoustical measurements on humans to show how these insights hold for acoustical data. Because human heads are reasonably approximated by a sphere, it is expected that the transition frequency between high and low plateaus in humans also variesd with head size. Therefore, when averaging across subjects, the head-size dependence of the ITD will add biases, and should be accounted for. I describe a method to do so and then turn to the presentation of the results.

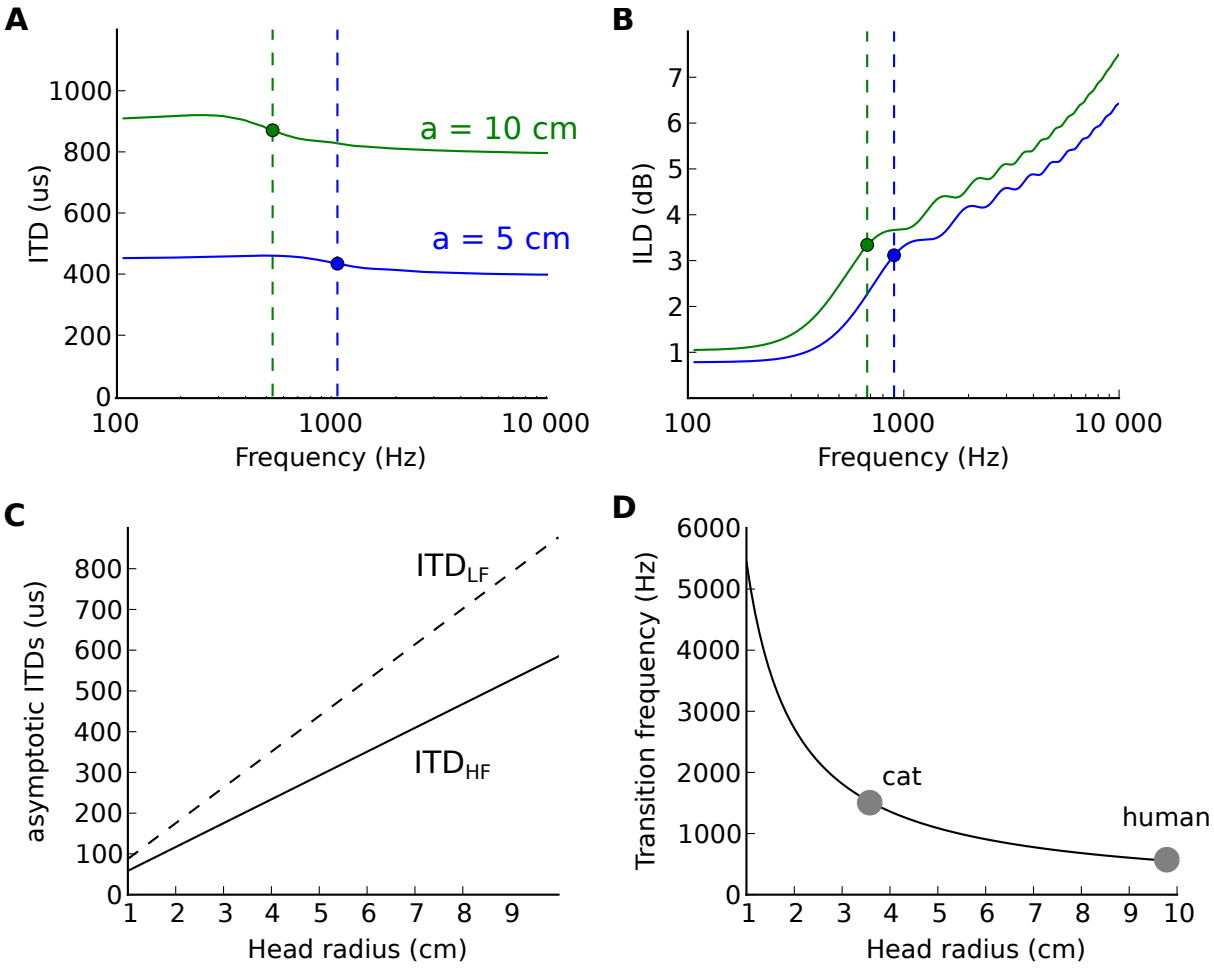

Figure 1.3: **A** Two ITD vs frequency curves simulated on a spherical head model for the position at 90°. The blue curve is for a model with head radius $a = 6cm$ and the green curve with head radius $a = 12cm$. The vertical bars indicate $f = \hat{f}$ for each model **B** The ILD vs frequency curves in a similar situation as on **C** Both the low and the high frequency estimates of ITDs in the spherical head model are a linear function of the head radius. **D** $\hat{f}$ is a hyperbolic function of the head size.

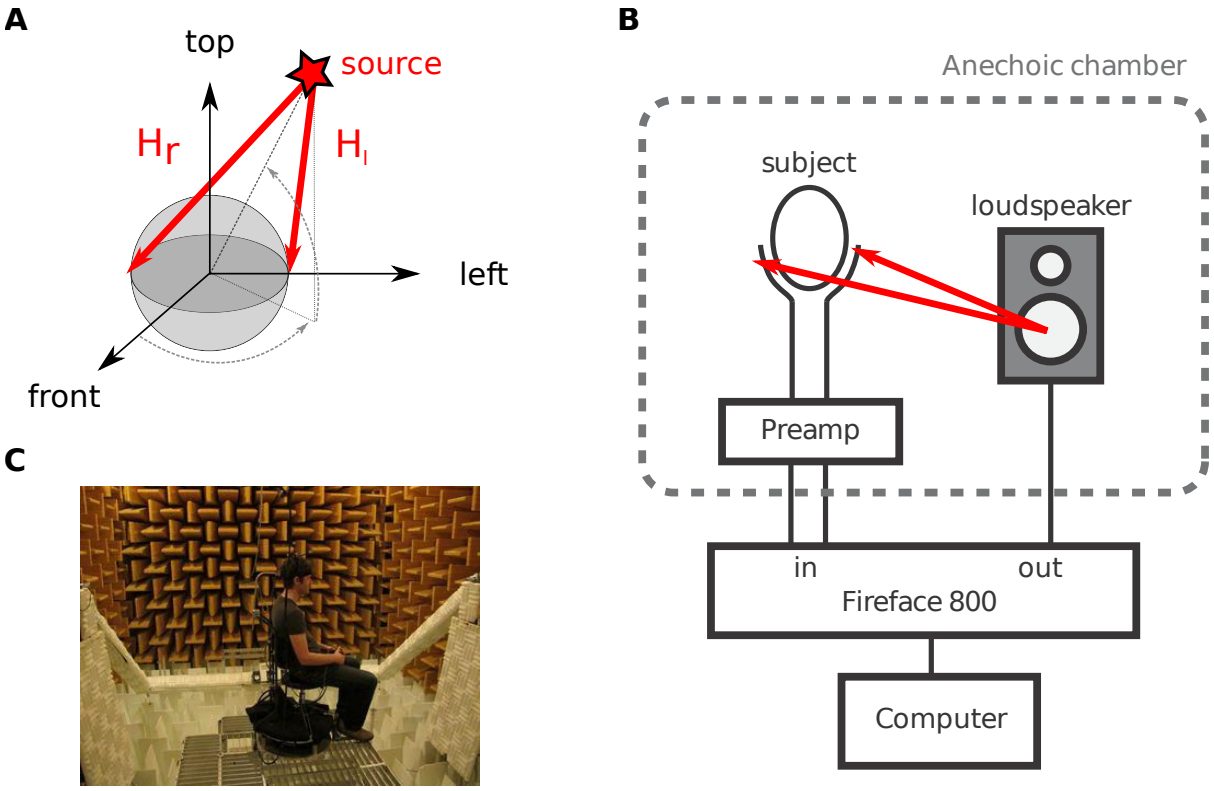

Figure 1.4: Measuring human HRTFs. **A** Depiction of the situation, the goal is to measure the two acoustical filters H^L and H^R. **B** Depiction of the signal path in the experimental setup
C Picture of a human subject in the anechoic chamber.

1.3 Methods

We first explain the experimental apparatus used to measure human HRTFs in Subsection 1.3.1, then we turn to the HRTF database analysis in Subsection 1.3.2.

1.3.1 Measuring human HRTFs

We need to measure two acoustical filters per position of the sound source, one for each ear (Fig. 1.4, **A**). The method used was largely based on the aparratus used to record the IRCAM LISTEN database [Database, 2002]. The signal path is depicted on Fig. 1.4, **B**. Microphones are placed in each of subject's ears and a probe sound is played through a speaker at a given position. The two filter responses (one for each ear) are then processed out of the probe sound and the sounds recorded at the ears, and constitute the HRTF at that position. The same process is repeated for every position of interest while keeping the subject seated still (Fig. 1.4 **C**). On this picture, the crane is in position for measuring a low elevation, while the subject has the speakers to his

left.

The method introduced by A. Farina [Farina, 2000] was used to measure the acoustical filters. It is based on exciting the system with and exponential sine-sweep, and deconvolving the excitation signal out of the recorded responses. More details is provided in the appendix Appendix 1.A, in what follows we provide description of the recording apparatus and then the useful definitions and notations.

Recording apparatus

Subjects were positioned in an anechoic room at IRCAM in Paris, depicted on **C** of Fig. 1.4. The room was designed to fit a concert piano, with wall-to-wall dimensions of 8.1 x 6.2 x 6.45 meters, which is plenty for a full HRTF recording setup. The walls are covered by 1.1 m thick glass wool wedges, which absorbs sound waves above 75 Hz, sufficient to ensure anechoicity.

Three speakers are arranged on a crane so that they face the subject from three positions separated by 5°degrees in azimuth (at 0 elevation). They were positioned so as to be at 1.95 meter from the subject. The subject was seated on a rotating platform, to measure horizontal positions the subject's chair is rotated. The crane supporting the speakers (white foam on Fig. 1.4, **C**) moved around a horizontal axis, and was rotated so that the speakers were at different elevations. It could reach from the lowest position -40°elevation to the apical position (90°elevation). The crane was carefully weighed to avoid using motors to stabilize it and hence produce noise.

The initial position of the subject relative to the system was set manually using lasers. Two lasers pointed from the crane supports so as to indicate the horizontal axis of rotation. The crane's horizontal rotation axis matched the intaraural axis when the subject was in position facing the speaker at 0°azimuth. A third laser pointing down indicated the center of rotation of the support.

Miniature, custom-made microphones were placed in a mold fitted to the ear. Recordings are thus in a blocked-meatus setting. The signal was first amplified by a specialized amplifier circuit (as in [Database, 2002]). This was done inside the anechoic room in order to get the signal strong enough to be carried all the way to the outside workspace Fig. 1.4, **B**.

The computer was connected to a RME Fireface 800 which received amplified signals from the microphones. Those signals were recorded by the computer using Max MSP and analyzed through various Matlab scripts. Max/MSP also controlled the movement of the turntable and crane between each recording. Recordings were performed at 24 bit resolution and 44.1 kHz sampling frequency for 16384 samples. They are compensated for the response of each of the three individual speakers, by measuring the "bare"

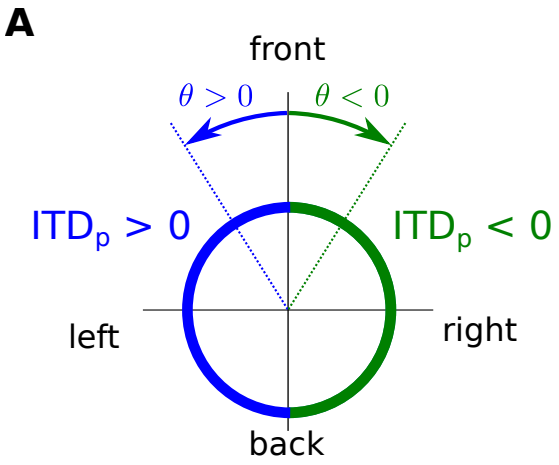

Figure 1.5: A depiction of the coordinates system used for sources in the horizontal plane. Positions are referred to with azimuths θ between -180°and 180°. By convention, positive azimuths produce positive ITDs (blue part of the circle), and conversely (green part of the circle.

response between the speakers and a microphone placed at the origin of the coordinates space.

1.3.2 HRTF database analysis

In this section I describe the analysis method used to exploit ITD_pdata across subjects and databases. I explain briefly the way the total phase $ITD_p$was computed on each pair of HRTF for a given position, then present the additional publicly available HRTF databases used. Finally I describe the method that was used to compensate the morphological parameters influencing the ITD_pdata.

Definition of the total phase ITD

Several definitions of the ITD exist (e.g. the interaural group delay, ...) and are discussed in depth in Chapter 2. Throughout this manuscript, ITD will corresponds to the total interaural difference in fine structure, that is the total phase delay ITD_p. Given a couple of HRTF files for one position, H^L and H^R, ITD_p as function of frequency is defined from the Interaural Phase Differences (IPD). Notice that, ITD_p and IPD are defined in such a way as to be positive when the azimuth of the sound source θ itself is positive, as shown on Fig. 1.5.

The IPD for a given position is the phase difference of the two monaural filters:

$$\mathrm{IPD}(\theta, f) = \frac{1}{2\pi} \left\langle \frac{H^R(\theta, f))}{H^L(\theta, f)} \right\rangle \tag{1.8}$$

Where the notation $\langle . \rangle$ is the phase operator that returns the phase in radians. Here, the IPDs are expressed in cycles, whithin the range IPD $\in]-0.5, 0.5]$. From this, the total phase ITD is defined as

$$\mathrm{ITD_p}(\theta, f) = \frac{U(\mathrm{IPD}(\theta, f))}{f} \tag{1.9}$$

Where $U(.)$ is the *unwrapping operator*. On a set of discrete measures of IPD, (IPD_k) arranged by ascending frequencies, it replaces any absolute jump greater than half a cycle, that is $\Delta_{\mathrm{IPD}} = |IPD_k - IPD_{k-1}| > 0.5$ by its complement $IPD_k = IPD_k - \mathrm{sgn}(\Delta_{\mathrm{IPD}})$, where sgn is the sign (+1 or -1) of Δ_{IPD}. This effectively *unwraps* the phase information that is returned by the Fourier transform.

Raw Impulse Responses (IRs) measured in the experimental setup are first trimmed to a smaller size. This is done by locating the maxima of each impulse response, and taking a fixed-length window around it. This allows for removal of parts of the IR that does not contain any meaningful information. The IPD is then computed out of the pair of HRTF in the frequency domain. The phase of the ratio of the two HRTFs at any given location is the IPD, which is converted to cycles and then unwrapped. The total phase ITD is finally attained by dividing by the frequency.

Estimates of low frequency ITD, high frequency ITD, transition frequency

For each position recorded in the HRTF dataset under study, the total phase ITD, $\mathrm{ITD_p}$ was computed at all frequencies between 400 Hz and 5 kHz. From these computed ITD curves, statistics are reported, to estimate the range of variation of the function between the low and high frequencies, and the band in which this transition happen.

The low- and high-frequency ITD estimates in the data, $\hat{\mathrm{ITD}}_{\mathrm{LF}}$and $\hat{\mathrm{ITD}}_{\mathrm{HF}}$for each position are computed as averages of $\mathrm{ITD_p}(f)$ over a given normalized frequency range. The $\hat{\mathrm{ITD}}_{\mathrm{LF}}$is the mean of the normalized averaged ITDs between the frequencies $f_n = 0.5$ and 0.6 (260 Hz and 320 Hz for the average head). The $\hat{\mathrm{ITD}}_{\mathrm{LF}}$is the mean of the normalized averaged ITDs between the frequencies $f_n = 7$ and 8 (3700 Hz and 4200 Hz for the average head). In the Results section I take the difference and the ratio of those two quantities to quantify the magnitude of the ITD variation.

Another interesting measure of the $\mathrm{ITD_p}(f)$ function for a given position is the *transition frequency*, that is the frequency at which the ITD transititions from the low to the high frequency regime. To estimate this quantity in the data, I define the transition frequency $\mathrm{f_{tran}}$as the frequency at which the $\mathrm{ITD_p}$ is equal to the mean of its high and low frequency values, $\hat{\mathrm{ITD}}_{\mathrm{LF}}$and $\hat{\mathrm{ITD}}_{\mathrm{HF}}$. The transition frequency is the

Database	F_S	N_s	Δ_θ	Δ_ϕ	N_{sub}	Room type
LISTEN (previous recordings)	44.1 kHz	8192	15°	$\simeq$ 15°	49	Anechoic
LISTEN (present project)	44.1 kHz	8192	5°	$\simeq$ 15°	35	Anechoic
CIPIC	44.1 kHz	200	$\simeq$ 10°	5.6°	36	Anechoic
ARI	48 kHz	2400	5°	10°	10	Semi-anechoic

Table 1.1: Overview of the different databases from which ITDs have been extracted. For each database, F_s: sampling frequency. N_s: length of the head-related impulse responses in sample. Δ_θ, Δ_ϕ : approximate spatial resolution in azimuth and elevation. N_{sub}: number of subjects from each database that have been used in the present study. The LISTEN database is made of the 49 subjects freely available on the IRCAM website and of 35 subjects that have been measured later using a setup with an increased spatial resolution in azimuth (for this specific project). 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°.

frequency at which an ITD closest to this intermediate ITD:

$$\mathrm{f_{tran}}(\theta) = \mathrm{argmin}_f |\mathrm{ITD_p}(f,\theta) - \mathrm{ITD_{intermediate}}|$$

Where the intermediate ITD is defined as a function of the HF and LF ITDs:

$$\mathrm{ITD}_{mid} = \frac{1}{2}\left[\hat{\mathrm{ITD}}_{\mathrm{LF}}(\theta) + \hat{\mathrm{ITD}}_{\mathrm{HF}}(\theta)\right] = \mathrm{ITD_p}(\mathrm{f_{tran}},\theta)$$

Databases under study

In addition to the recordings performed for the specifics of this study, the same analysis was conducted on publicly available HRTF databases, listed in Tab. 1.1. Those databases are the CIPIC [Algazi et al., 2001c], ARI [ARI, 2010], and LISTEN databases [Warusfel, 2002] (the setup of which was used to performs the recordings specific to this study). To compensate for different temporal sampling rate, the ARI impulse responses were downsampled to 44.1kHz, this samplerate was used throughout the analysis. Compensating for the different spatial samplings of each database was a somewhat more complicated issue. First, measures of interest were obtained separately on each database's spatial grid. Those measures were then spatially interpolated to a common grid, corresponding to the LISTEN database (present project, Tab. 1.1). This interpolation was performed using a nearest-neighbor approach, no special care for the circularity of both azimuth and elevation measures was taken in this process.

Morphology normalization

Thanks to insights from the spherical head model explained in Subsection 1.2.2, the shape of the ITD curves for one position should vary across the population of subjects in two ways. Extremal ITDs in HF and LF should scale with the head size, and the transition frequency with the inverse of the head size. Thus, to compare the direction- and frequency-dependent ITDs extracted among the different listeners it is necessary to normalize frequencies and ITDs with respect to listeners morphological parameters.

In the spherical head model (assuming ears lie on a diameter) the only morphological parameter that is necessary to compensate for morphological differences is the spherical head radius a. The real spherical head radius can be measured on the human directly, but it is not available in some of the databases under study here (Tab. 1.1). Indeed, human heads are not spherical but more or less ellipsoidal [Duda et al., 1999a, Jo et al., 2008], an *"equivalent spherical head radius"* is then needed for normalization purposes. This equivalent spherical radius is estimated from the acoustical data itself. In this project it was done by minimizing, in a least square sense and over the whole sphere, the difference between high-frequency ITD in the data $\hat{\mathrm{ITD}}_{\mathrm{HF}}$ from the Woodworth high frequency approximation (Equation 1.3). This procedure was inspired by the one proposed by [Algazi et al., 2001b].

When applying this procedure to the HRTFs of the 130 listeners of the databases of Tab. 1.1 a mean equivalent spherical head radius of 10.3 cm with a standard deviation of 0.6 cm are obtained. The obtained mean value is higher than the mean value of 8.7 cm reported by [Algazi et al., 2001b] among 25 listeners, and commonly used in the literature.

1.4 Results

1.4.1 ITD as a function of frequency in humans

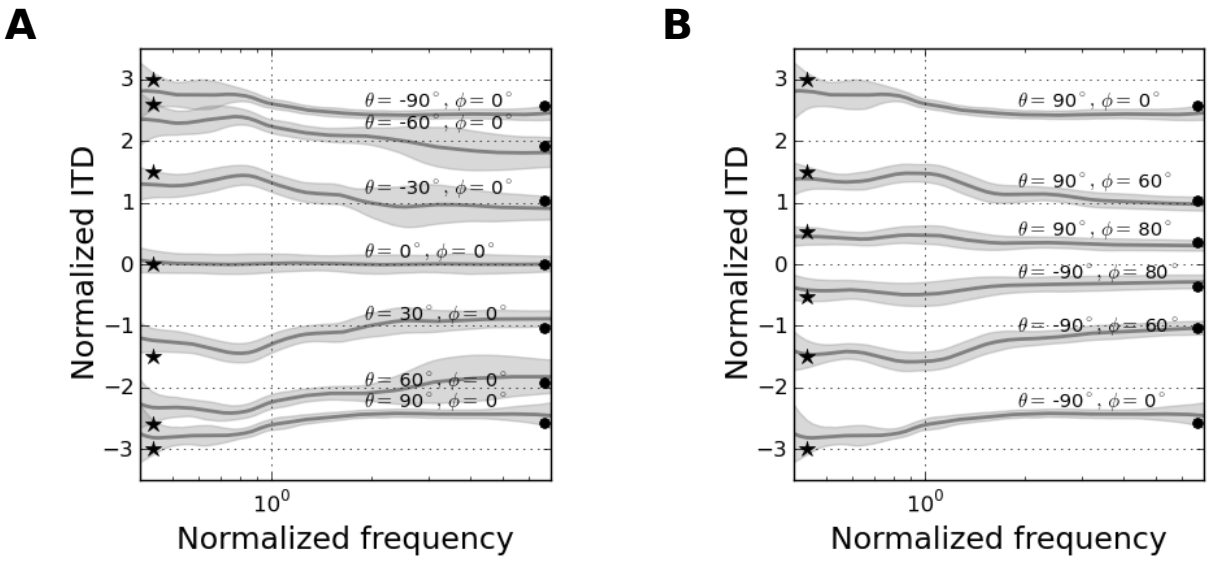

Figure 1.6: Means (solid line) and standard deviations (Grey area) of normalized ITDs. **A** ITD_{norm} in the horizontal plane (and frontal positions). **B** ITD_{norm} in the upper interaural plane (the vertical plane containing the interaural axis). All data is averaged over the 130 listeners of the different databases for the frontal horizontal plane and in the upper interaural plane. The $\star$ symbols stand for the low-frequency approximations ITD_{LF} calculated using Equation 1.2. The $\bullet$ symbols indicate the high-frequency approximations ITD_{HF} calculated using Equation 1.3.

Following morphological normalization, the frequency- and direction-dependence of the obtained ITDs can then be analyzed across the populations in normalized values. Means and standard deviations of the normalized ITDs averaged over the 130 listeners of the different databases for the frontal azimuthal plane and the upper inter-aural plane are presented in Fig. 1.6. The $\star$ and $\bullet$ symbols on the panels show the theoretical predictions provided by Equation 1.2 and Equation 1.3. Recall that the high frequency estimate was used for the morphology normalization, and is therefore necessarily a good fit to the data . The low-frequency approximation ITD_{LF} (calculated using the radii fitted on the *high-frequency* ITDs) Equation 1.2) is reported, and is a good fit to the low-frequency ITD regime. This suggests that the normalization mechanism, based on the spherical model, does not fail to account for the increase in ITD in low frequencies. For comparison, the exact same analysis was performed on HRTFs modeled from the spherical head model. The result of this analysis is reported in Appendix 1.B.5 (Fig. 1.15).

Standard deviations associated with the normalized ITDs averaged over the 130 listeners are very small (Fig. 1.6, shaded area), of the order of tenth of microseconds. This validates a posteriori the use of the *"equivalent head radius estimation"* of Subsection 1.3.2 and of the normalization procedure. Furthermore, the standard deviation at each frequency (here about 0.2 normalized ITD, 60 μs for the average subject) are smaller than the variation of the ITD across freuqency (1 normalized ITD, 300 μs for the average listener).

The variation of ITD observed across the spectrum for each position on Fig. 1.6, **A** is consistently bigger than would be attributable to noise. Indeed, the difference in high and low frequency ITDs relative to the standard deviations across subjects is large. Therefore, the variation of ITD observed in our data cannot be attributed to noise, and are robust to individual variations of head shape.

1.4.2 ITDs at low- and high-frequencies

The ratio and the difference in the low- and high- frequency ITDs ($\text{ITD}_{\text{LF}}$and ITD_{HF}) are reported for positions in the horizontal plane on **A** of Fig. 1.7.

The difference in high- and low-frequency ITDs in the data has to be understood relative to human perception. A measure of the ability of humans to discriminate sounds with different ITDs is the Just-Noticeable-Difference (JND) (see Chapter 9 for the psychoacoustics of ITD processing). This measure gives the smallest ITD that humans can discriminate in binaural sounds. The JND for ITD discrimination by human listeners is of about 20μs [Mossop and Culling, 1998, Akeroyd, 2006]. The difference in LF and HF ITD reported in Fig. 1.7 is in units of JND, that is 1 JND = 20 μs. Therefore, when the ITD difference reported is bigger than one, then it is detectable by humans. For most azimuth, the ITD difference is of several JNDs with a mean of 5.63 JND (standard deviation = 2.17, and 5.53 JND, std = 2.4 on the full sphere). In the bright spot, around 110°, the ITD difference is minimal Fig. 1.7. In fact, ITDs at those positions are roughly constant, even though the histogram on Fig. 1.7, **B** shows that the ITDs are bigger than 2. Overall, these results suggest that the ITD variation across frequencies is of sufficient magnitude over the whole sphere to be discriminated by human subjects.

On Fig. 1.7, **C**, the pattern of ITD differences is plotted for all azimuths and elevations. It can be noticed that the ITD difference is not centrally symmtric around 90°azimuth, as predicted in the spherical head model Fig. 1.15. Rather, the cones of confusion are distorted and centered around a position at 115°. This azimuth corresponds to the ear position, and thus the ITD is constant when the source is facing the

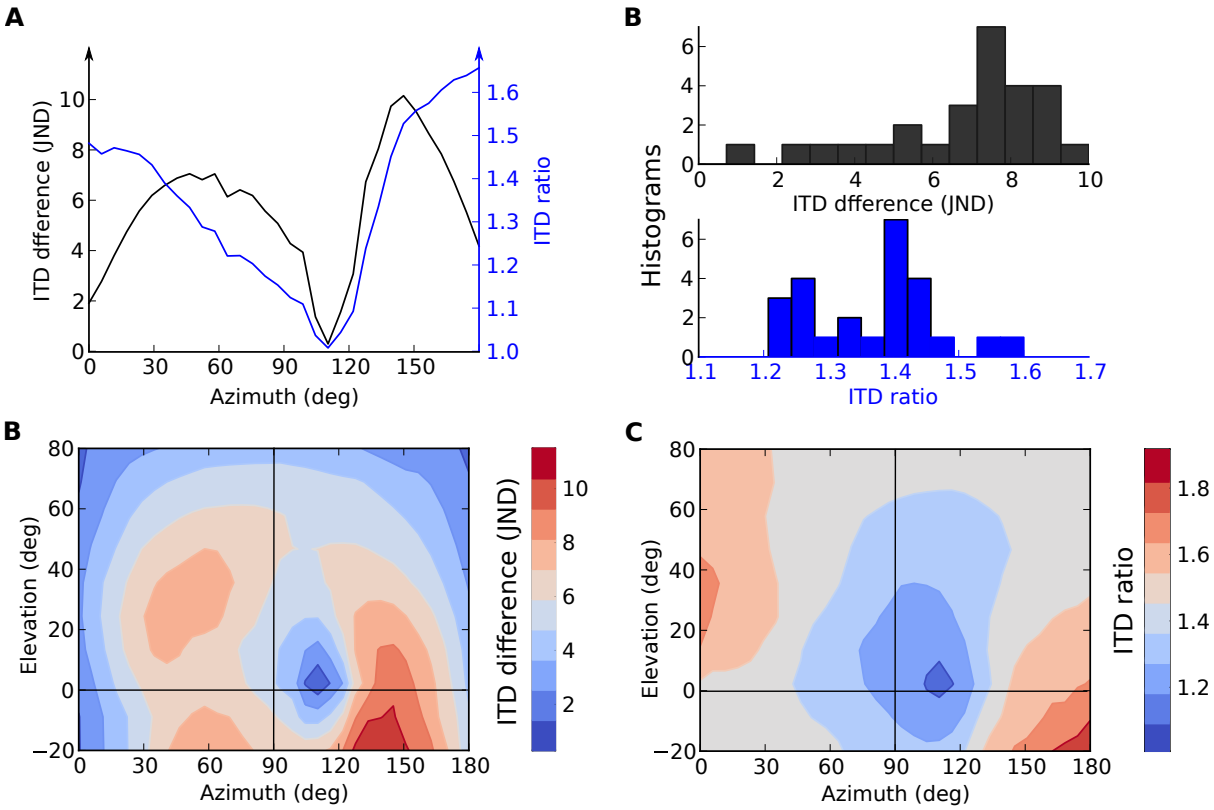

Figure 1.7: Mean ratios and JND-normalized difference between the low- and high-frequencies ITDs. **A** The ratio (blue curve) and JND-normalized difference (black curve) between high and low frequency ITDs are reported for all positions on the horizontal plane. Differences can amount to up to 10 JNDs for the intermediate positions (around 60°and 150°). **B** Histograms of the average ratios and differences over the full sphere (from -20°to 80°elevation and one side). **C** and **D** show the average ratios and differences as a function of azimuth and elevation. Notice that the expected symmetries are not quite respected (front/back and up/down).

ear. The relative brak of symmetry can also be attributed to the contributions of the shoulders and of the torso.

As opposed to the ITD difference, the ratio of low- to high-frequency ITDs provide a measure of ITD independent on the head size. The ratios are presented on Fig. 1.7, at elevation 0°on **A** and as a function of azimuth and elevation on **D**. This ratio is maximal for positions around 60°and 140°, where the low-frequency ITD is 60% bigger than the high-frequency ITD. On average, the ratio is 140% (std = 14 %), and minimal when sound source points at the ears (azimuth 115°). This is also reflected on the histogram of ratios reported in Fig. 1.7, **B**. The LF/HF ITD ratios, when observed across elevation and azimuth Fig. 1.7, **D** is quite close to the predictions of the same measure in a spherical head (Appendix 1.B.5), with the noticeable difference that the symmetry is displaced around the ear (100°azimuth). These data suggest that in all instances the low-frequency ITD is bigger than the high frequency ITD by 40 % on average (over the full sphere).

1.4.3 Transition between low- and high-frequencies

The frequency band in which the transition from low and high frequency regimes is especially important because in this zone the ITD varies strongly with frequency. Indeed, in the low and high frequency regimes, the ITD plateaus and therefore can be said to be constant ITD. In particular, if the transition frequency is well outside of the range where ITD is a dominant cue to sound source azimuth, then it cannot be argued that the variation of ITD is relevant to perception.

The *transition frequency* that is the frequency at which ITDs switch from their estimated low frequency value to their high frequency one, are examined here on the databases of human HRTFs. Normalized transition frequencies have been computed from the averaged normalized ITDs using Equation 1.3.2 (described in Subsection 1.4.2). They are reported on Fig. 1.8, **A**, as a function of azimuth and elevation, and **B** on the horizontal plane. Consistently with the spherical model, transitions between low- and high-frequency averaged normalized ITDs occur always for normalized frequencies larger than 1. Furthermore, for normalized transition frequency appears to be direction-dependent in a way that is qualitatively similar to the theoretical predictions (Fig. 1.15). Dormalized frequency increases as the source is moving away from the median plane, up to an azimuth angle $\simeq 70°$ where a maximum value $f_{\text{tran}} \simeq 2.8$ is reached Fig. 1.8. Normalized frequency then decreases and a minimum value is obtained for positions close to the ear position (115°). Overall, the transition frequency is well within the range in which humans are known to rely on ITD for azimuth sound

localization, consistently below 1.5 kHz Fig. 1.8, **B**.

1.5 Discussion

I have shown that for human listeners, the ITD cues are different for sounds with different frequency content at the same position. This tendency is broadly consistent with predictions based on a spherical head model. To some extent, this analogy is limited for some regions of space, by the fact that several minor breaks of symmetry and the positioning of the ears influence ITD. Also, the head is ellipsoidal, rather than spherical [Duda et al., 1999a]. Yet, the spherical model-inspired morphology normalization allowed me to show that the magnitude of variation of the spectrum is on average of several JNDs, and that the variation occurs below 1.5 kHz. I argue that the variation of ITD are relevant to human perception on two scales, For any given position, ITD in high-frequency and low-frequency sounds are different. For a given sound with a sufficiently broad spectrum, or within a cochlear channel, the ITD is in fact also significantly variable.

1.5.1 Global variations of ITD

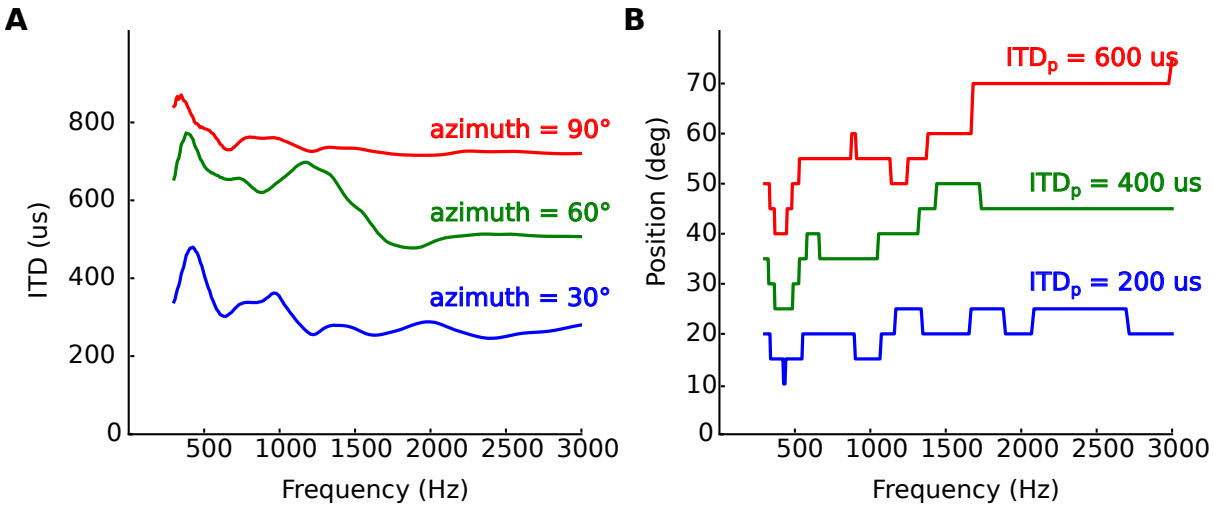

Figure 1.9: **A** ITDs as a function of frequency for three positions, one human subject (main author, new recordings). Positions plotted are 90°, 60°and 30°azimuth. **B** ITD as a function of frequency for three positions, same subject. Because recordings are spaced 5°apart, the curves are vertically quantized. **C** show HF and LF physiological ranges

The frequency-dependence of ITD means that a given ITD value signals different positions depending on the frequency content of the source signal. On **A** of Fig. 1.9 we

show example ITD curves for one human and three positions in the horizontal plane. If we were measuring an ITD of 700 μs, on a bandlimited signal of 500 Hz, we would conclude to a position of 60°. But if we were to measure it at 1500 Hz, then the ITD would signal the position 90°. Using multiple recordings on the azimuthal plane we can reconstruct the function that gives the signaled position as a function of frequency for a given ITD value. This is reported on **B** of Fig. 1.9

It has been argued in the literature that the frequency-dependence of the ITD was not meaningful to the lateralization of sounds, we review those studies in the next section. We argue that this cannot, logically, be the case because of the property highlighted in Fig. 1.9 The binaural system has to somehow take into account this variation of ITD across frequency in order to produce a consistent perception of azimuth. If it didn't, it would mean that in the high frequencies sounds at 90° would be perceived closer to the midline (this prediction is tested in humans in Chapter 10).

Because of the tonotopicity of the auditory and binaural system, the problem of taking into account the frequency dependence of ITDs is not hard to solve. In most animals, mammals and bird alike, there exists populations of neurons that are sensitive to the ITD presented to the animal. This is also done in a tonotopic way, that is arranged in frequency bands, and each cell only is excited by a small part of the spectrum, centered around its Center Frequency (CF). In sum, each neuron has a preferred frequency (the center of the neuron's passband), and an ITD selectivity curve, of which the Best Delay (BD) is the maximum.

Those populations of neurons have been for a long time assumed to represent collectively the cross-correlation of the inputs (see, e.g. [Colburn and Durlach, 1978]). A common problem is to estimate the position of the sound source from th activity of such a population of cells, the *binaural display* [Joris and Yin, 2007] Amongst those models, the *lateral-image* model [Stern et al., 1988] provides a mechanism in which the position of the sound source is read out using a the weighed activity of the population. The relevant weight component in their model to our problem is the *straightness of ITD*. The authors argue that the neurons that represent the same ITD across different frequency bands should be weighed together. The argument is that a stimulus evokes a consistent ITD over its spectrum, which may encompass many cochlear filters. This, as we show in this chapter is actually not the case, indeed, even in anechoic settings the ITD for a given position is not constant. In the framework of [Stern et al., 1988], the meaning of *straightness* can be modified to account for our findings: the weighing should be done according to the consistency of ITDs across frequency bands *with the ITD vs. frequency curve of a given position.* Therefore, taking into account the global frequency dependence of ITDs in the usual view of binaural hearing is just a

matter of changing a the “wiring” between successive layers of neurons.

1.5.2 Local variations of ITD

In the previous section we have shown that the ITD varied across frequencies, with on average two plateaus and a transition. We have shown as well that because of the tonotopy of the binaural system, this feature can be quite naturally incorporated in neural models of the perception of sound. It could be argue that in the low and high frequency regimes the ITD is roughly constant, at least over the width of a human cochlear filter, or a neuron's frequency tuning. But as is clear from Fig. 1.8, the transition between those two regimes always occurs well inside the range where ITD is a dominant cue to localization [Wightman and Kistler, 1992]. Hence it is likely that ITD is best understood as being variable even at the scale of a single neuron. To quantify this, the difference in ITDs over bands of frequency was measured on the human data. A sliding passband with $Q = 3$ was used, and the difference of the two ITDs at the edge of the band is plotted on Fig. 1.10, **A** as a function of the center frequency of the band. This graph illustrates for a few positions (black lines) the ITD extent, that is the amount to which ITD varies over a third octave band centered on this frequency. The right axis provides the ITD in time units. The variation of ITD over this small bandwidth is quite big, of the order of 30 μs, (**C** and **D**), and can reach a hundred microseconds around $f_n = 1$, that is 600 Hz in humans.

Therefore, in some frequency bands, the ITD varies significantly even within the bandwidth of a single auditory filter. This effect could be boosted by effects of the environment. A recent study has shown that the environment could have a great effect on the ITD, and specifically on the magnitude of its variation across frequencies [Gourevitch and Brette, 2012]. The fact that the ITD depends locally in frequency implies that each monaural filter has a non-constant total phase delay. This kind of delays is not trivial to understand, and introduce phase-distortion in the signal. This point is specifically adressed in the next chapter of the dissertation.

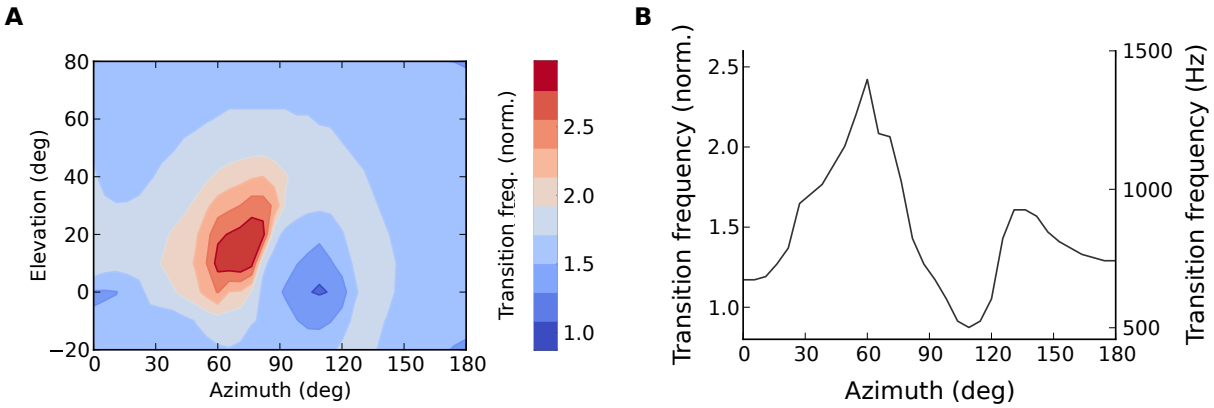

Figure 1.8: Transition frequency between the low- and high-frequency ITD regimes and local variations of ITD. For humans and all positions reported before, the transition frequency is color coded on **A**. Notice that similarly to before, the up/down front/back symmetries are lost. The transition frequency on horizontal plane positions is plotted as a function of azimuth on **B**, which emphasizes the lack of front/back symmetries.

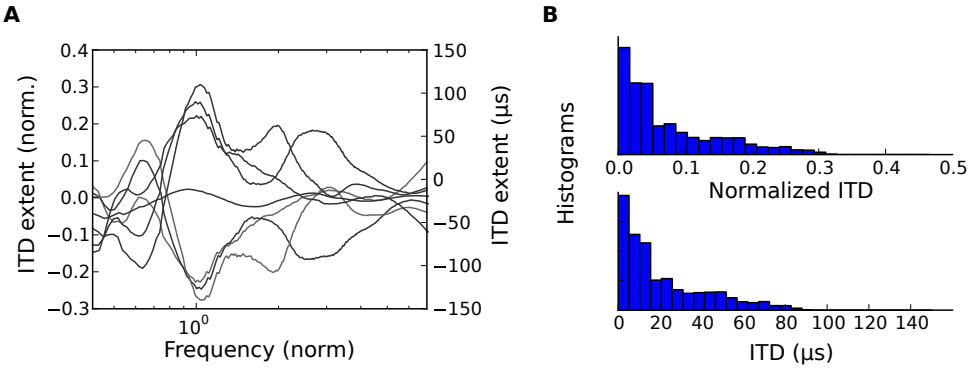

Figure 1.10: Variation of ITD in a cochlear bandwidth. The variation of ITD across a third-octave bandwidth is shown for a few positions on **A**. A discussed, the ITD varies broadly over the bandwidth of a single cochlear filter, especially around the normalized frequency equal to 1-2. **B** Histograms of the ITD extent as measured on third octave band, top panel reports the normalized ITD values and lower panel the ITD value in microseconds.

1.A Measuring filter responses: deconvolution

Many methods have been devised in the signal processing field to characterize the transformation imposed by a system. Especially in the audio field have those measures been critical, for example to ensure a proper response of audio recording or rendering systems, but also the acoustics of a room...

Such measurements generally consist in recording the response of the system to an excitation stimulus $s(t)$, the system can then be characterized by comparing the output $y(t)$ to the input. Consider the system on figure 1.11, we are interested in finding the system response h. In the case of linear, time invariant systems, this task only requires the measure of an impulse response (see, e.g. [Oppenheim et al., 1999]), that is the response of the system to a Dirac function. As an illustration, a method that is still widely used in the field of concert hall acoustics consists in recording a pistol shot. The pistol shot indeed has a wide band of energy concentrated at a time point, much like a Kronecker delta function. Another method (used in [Kuhn, 1977]) it to measure the response of the system to pure tones at various frequency points. Obviously this is a very time consuming endeavor.

But more sophisticated methods have been developed over time. The Maximum Length Sequence (MLS) method used a pseudo random, spectrally flat stimulus that has a delta-like autocorrelation. Cross-correlating the output $y(t)$ with the excitation signal $s(t)$ yields the response of the system. This method is tedious to use in practice, as the MLS excitation stimulus has to be quite long for the approximation of the auto-correlation is Kronecker delta. It also requires precise synchronization of the excitation stimulus and the response recording. Furthermore, if one wants to compute the non-linear part of the response of the system, one has to make another set of recordings, for example using Time-Delay Spectrometry. This method was used in the context of binaural recordings in many instances, see e.g. [Koka et al., 2008a, Tollin and Koka, 2009c] for recordings in animals.

The method used here was developed by A. Farina [Farina, 2000]. It relies on a logarithmic frequency sweep as an excitation signal. The excitation signal can then be deconvolved out of the response of the system. The probe stimulus used was an 8192 samples long upward logarithmic frequency sweep at 44.1 kHz sampling frequency (or 192 kHz in the case of animals). This signal was then deconvolved out of the recorded response so as to extract the impulse response of the filters.

This method has been shown to be robust to jitter between the excitation and recorded stimulus, and provides a fast measure of the impulse response of the linear part of the system. Also, because the energy is spread out over time this method allows

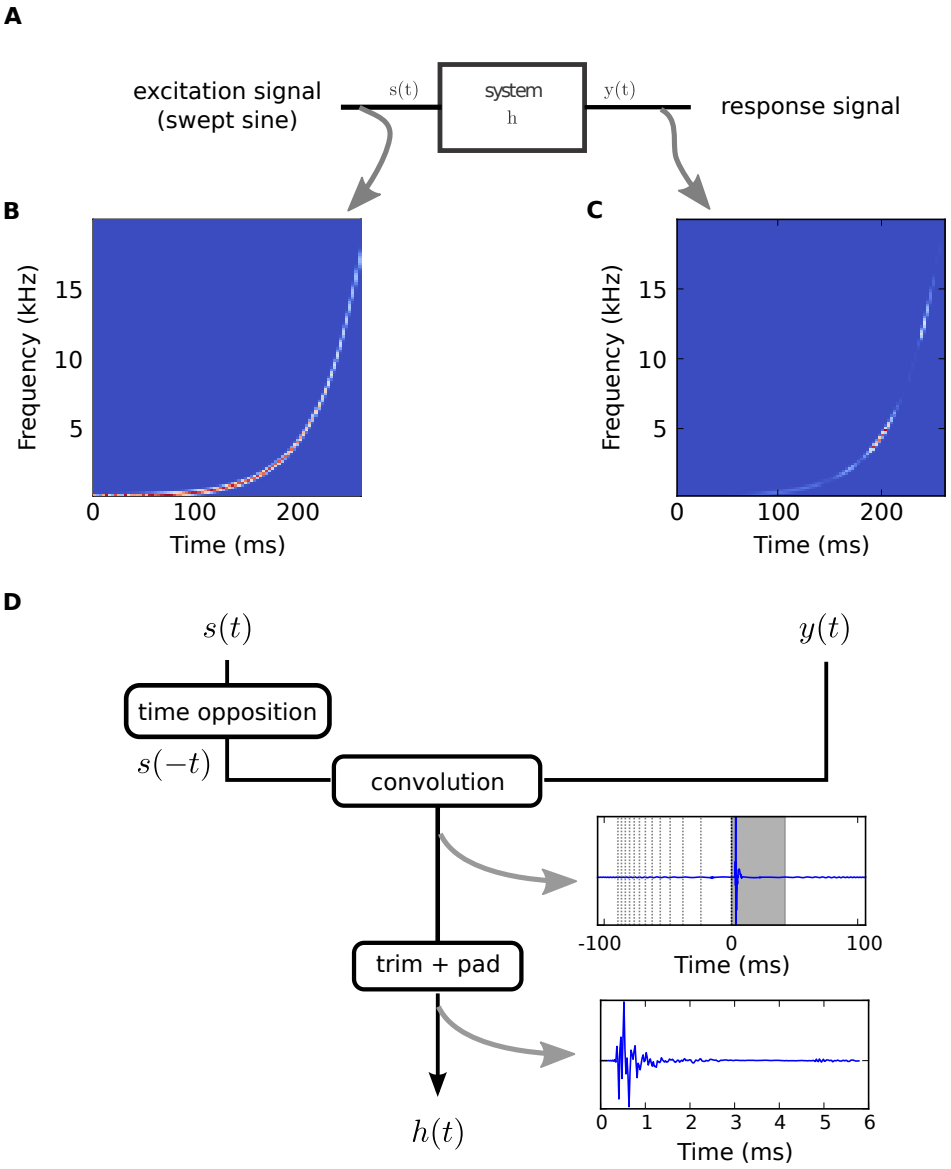

Figure 1.11: Depiction of the deconvolution process. **A** Shows the system under study. A signal s(t) is fed through the system with impulse response h. The output signal, $y(t)$, in our case the signal at the ear is recorded. **B** The excitation signal is an upward ramping sine sweep with exponential instantaneous frequency. **C** The recorded signal is almost the same signal as in **B** It is only sometimes visibly attenuated at some time and space points. **D** Schematics of the deconvolution process. Because the sweep's autocorrelation is a Dirac function the cross correlation of the excitation signal and the response signal is an impulse. This impulse response is then trimmed to retrieve the linear part of the response of the filter.

to output the sound through loudspeakers with minimal artifacts due to non linearities. It is even possible to obtain a sequence of impulse responses one for each harmonic distortion order. This information can be used to completely identify a subclass of non linear systems, as shown in [Rébillat et al., 2011]. But in our case, because the head is a passive acoustic object, we can safely consider that the system is linear. The nonlinearities observed have to arise from the rendering (loudspeakers) or recording (microphones, amplifier) apparatus, which were already assessed [Database, 2002] on the same recording setup.

1.B Spherical model

The problem is that of the scattering of a planar sound wave by a rigid sphere centered on the origin? In this section I provide the solution to this problem in the steady state for spherical coordinates and derive the estimates of the ITD shown in the main text. Also, I explain how the contribution of the spherical head to the acoustical field can be assessed by computing the diffracted and incident fields separately, as done in Fig. 1.2.

The problem of the scattering of the sound field by a sphere was first solved by Lord Rayleigh [Rayleigh and Lodge, 1904], in the beginning of the XX$^{\text{th}}$ century. The same formalism was then developed by many authors, [Wiener, 1947] provides the first recordings on crafted wooden balls that experimentally validates the model on rigid spheres. Then [Anderson, 1950] provided the modern formulation of this problem in the context of sound scattering of small animals in marine acoustics. Spherical-head models based on rigid spheres have been widely used in the literature in order to study acoustical localization cues. Among the first studies showing the variation of ITD across frequency for a given position, [Abbagnaro et al., 1975, Kuhn, 1977] both relate this phenomenon to the spherical head model and diffraction effects. The latter provides the well known ITD estimation in low frequencies Equation 1.2 (which is derived here). Interesting developments of the theory of the problem of the scattering of the sound waves by a sphere can be found in [Morse and Ingard, 1968, Bruneau, 2010].

1.B.1 Notes about incidence angle and spherical coordinates

In general, I will be working in standard spherical polar coordinates (Fig. 1.12). A point is identified by its distance to the origin, the radius r, and two angles between $-\pi$ and π. The azimuth θ and the elevation ϕ. In the main text, only the horizontal plane is considered, in which case $\phi = 0$ all the time. The left of the subject is at azimuth $\theta = \pi$ and the right at $\theta = -\pi$.

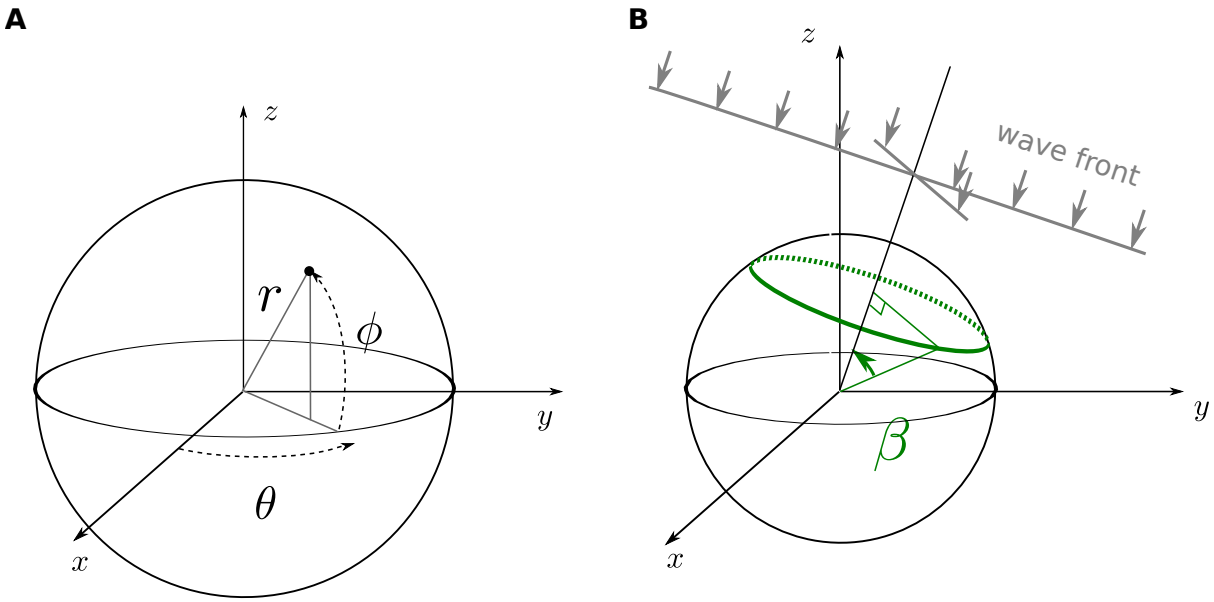

Figure 1.12: The spherical coordinates system used in the acoustics part. **A** The standard spherical coordinates used in this part. The subject faces along the x axis. **B** A planar wave is incident on a sphere centered on the origin. The points on the green circle all have the same pressure values because of the rotational symmetry of the problem. The only spatial variables of interest are the incidence angle β and the radius r.

The problem of the scattering of a planar wave by a rigid sphere is invariant under rotation around the axis of propagation of the sound wave. This means that in planes parallel to the wave front, all points on circles centered on the projection of the sphere have the same pressure (green circle, Fig. 1.12). This means that in essence, only the angle to any vector to the green circle on Fig. 1.12 with respect to the axis sphere-wavefront are relevant to the computation of the sound field. Hence one introduces the *incidence angle* β (Fig. 1.12, **B**), which is related to the azimuth and the elevation by the following formula:

$$\beta = \arcsin\left\{\cos(\phi)\sin(\theta)\right\} \tag{1.10}$$

1.B.2 Acoustical pressure field for a planar sound wave

In Cartesian coordinates

Let us assume that there is a harmonic source emitting a sound in the direction $x \to \infty$. Further, I assume that the wave is planar and propagates along the x-axis. Therefore, the wavefront is a plane orthogonal to the x-axis, and all points on the plane have the same pressure values. I compute the *incident pressure field*, that is the pressure field

in the absence of a sphere, due to a planar incoming wave. $p_i(x,t)$ expressed in complex notation, reads:

$$p_i(x,t) = p_0 \times e^{ikx} \times e^{i\omega t}$$

The *physical* pressure field $\hat{p}$ is obtained by taking the real part, $\hat{p} = \Re\{p\}$. The *wave number*, k, which has the unit of the inverse of space is introduced:

$$k = \omega/c = \frac{2\pi f}{c} \tag{1.11}$$

Where c is the speed of sound propagation in the atmosphere $c = 340 m.s^{-1}$, and $i^2 = -1$.

The above expression is simple because I am assuming that the sound source was originating from a particular coordinate axis. When describing situations in which the incident wavefront has varying directions, the solution to the wave equation has a more complex formulation.

In spherical coordinates

The problem is that of computing the acoustical pressure field from an arbitrary incidence direction defined by an incidence angle β and a modulus r (as in Fig. 1.12). The incident pressure field in spherical coordinates $p_i(r,\beta)$, can be decomposed on a basis of spherical harmonics (see, e.g. [Bruneau, 2010]):

$$p_i(r,\beta) = p_0 \sum_{n=0}^{\infty} i^n (2n+1) j_n(kr) P_n[\cos(\beta)] \tag{1.12}$$

Where $\mathrm{P}_n[.]$ are the Legendre polynomial of degree n, and $\mathrm{j}_n(.)$ the n^{th} order spherical Bessel function. The multiplicative term p_0 is the pressure amplitude of the source. This is the textbook expression for a planar wave in spherical coordinates, it assumes that the source is at infinity, but this can also be included in this analysis (see [Duda and Martens, 1998]).

1.B.3 Expression of the pressure field diffracted by a sphere

When a sphere scatters the sound, it can be modeled by assuming that the total pressure field p is the sum of an *incident field* p_i and a *diffracted field* p_d (also called scattered or reflected field in the literature). I assume that there is a diffracted field p_d that originates from the center of the sphere. This diffracted field must be a solution of the Helmholtz equation, verifying the Sommerfeld condition that the reflected field is null at infinity, and a boundary condition at the surface of the sphere. Notice that

all this study applies to steady state wave propagation. Such an outward-diverging (radiating) spherical wave originating from the origin can be expressed in spherical harmonics [Bruneau, 2010]:

$$p_d(r,\beta) = \sum_{n=0}^{\infty} a_n h_n(kr) P_n[\cos(\beta)] \tag{1.13}$$

The a_n are free complex coefficients. Where h_n is the spherical Hankel function of the second kind. The spherical Bessel function j_n that we met in the expression of the incident field is the real part of the spherical Hankel function $h_n = j_n - iy_n$. The imaginary part, y_n is sometimes called the Neumann function (e.g. in [Kuhn, 1977]).

Because the total pressure field must verify some boundary conditions on the surface of the sphere, the coefficients a_n can be identified. The boundary condition is that the normal velocity should be continuous at the surface of the sphere. This is linked to the specific acoustical impedance Z, defined as the ratio of sound pressure p and particle velocity v, $Z = p/v$ (measured in rayls). The boundary condition reads:

$$\frac{\partial}{\partial r} p + \frac{k\rho c}{Z} p = 0 \quad for \quad r = a$$

Where ρ is the density of air and c speed of sound in air.

One can express the total pressure field, $p = p_i + p_d$, and the normal velocity $\partial p/\partial r$, as a function of the a_n coefficients. Doing this and applying the boundary condition, the expression of the a_n coefficients is obtained:

$$a_n = -p_0 i^n (2n+1) \times \frac{\frac{k\rho c}{Z} j_n(ka) + i j_n'(ka)}{\frac{k\rho c}{Z} h_n(ka) + i h_n'(ka)}$$

Where the notation $'$ designates taking the derivative function with respect to its argument, as in $j_n'(x) = \partial j_n/\partial x$.

An interesting limiting case is that when the sphere is rigid. In this case the magnitude of the impedance Z goes to infinity and the coefficients become:

$$\lim_{Z\to\infty} a_n = -p_0 i^n (2n+1) \times \frac{j_n'(ka)}{h_n'(ka)}$$

In the end, the diffracted field reads:

$$p_d(r,\beta) = -p_0 \sum_{n=0}^{\infty} i^n (2n+1) \times \left[\frac{j_n'(ka)}{h_n'(ka)}\right] h_n(kr) P_n[\cos(\beta)]$$

And the total field is obtained as the sum of the diffracted and incident fields:

$$\begin{aligned} p(r,\beta) &= p_i(r,\beta) + p_d(r,\beta) \\ &= p_0 \sum_{n=0}^{\infty} i^n (2n+1) \left[j_n(jr) - \frac{j_n'(ka)}{h_n'(ka)} h_n(kr) \right] P_n[\cos(\beta)] \end{aligned}$$

Expressed as a function of time in the case of a harmonic source, the pressure at (r, β) reads:

$$p(r,\beta,t) = p_0 \sum_{n=0}^{\infty} i^n (2n+1) \left[j_n(kr) - \frac{j_n'(ka)}{h_n'(ka)} h_n(kr) \right] P_n[\cos(\beta)] e^{i\omega t}$$

From the above relationship, computing an HRTF is just a matter of taking $r = a$ in the above definition. Doing so, the expression of p is simplified. The transfer function from the source to a point on the sphere reads:

$$p(\beta,ka) = \sum_{n=0}^{\infty} i^n (2n+1) \left[j_n(ka) - \frac{j_n'(ka)}{h_n'(ka)} h_n(ka) \right] P_n[\cos(\beta)]$$

This expression can be simplified further given the following relationship between spherical Hankel and Bessel functions:

$$\forall n \forall x \quad h_n'(x) j_n(x) - j_n'(x) h_n(x) = \frac{i}{x^2}$$

Rearranging the terms in the expression of the total pressure field p yields:

$$p(\beta,ka) = p_0 \sum_{n=0}^{\infty} i^n (2n+1) \left[\frac{h_n'(ka) j_n(ka) - j_n'(ka) h_n(ka)}{h_n'(ka)} \right] P_n[\cos(\beta)]$$

Which leads to the common expression provided in [Kuhn, 1977] and [Duda and Martens, 1998]:

$$p(\beta,ka) = \frac{1}{(ka)^2} \sum_{n=0}^{\infty} i^{n+1} (2n+1) \left[\frac{1}{h_n'(ka)} \right] P_n[\cos(\beta)]$$

Notice that because the wave number has units of inverse distance, the term $ka = 2\pi f a / c$ is unit less. For a given a, it is proportional to frequency of the input wave f, hence the frequency-dependence in the above equation is "hidden" in the term ka.

Near field (plots in the main text)

In the main text, Fig. 1.2 (**D-G**) shows the computation of the additional delay and gain in amplitude between the two-mics case and the case where a spherical head is present.

In our current framework, this is simple. The incident field, p_i, is the solution of the wave equation *in the absence of a sphere*, and hence is the description of the two-mics model of the head (see main text). To quantify the difference in the two cases, we are interested in computing the ratio of those two complex pressure fields, which is a transfer function H defined for each point outside the sphere:

$$H = \frac{p(r, \beta)}{p_i(r, \beta)}$$

The phase of this transfer function is the additional delay the wave undergoes in the presence of the sphere as opposed to the case without a sphere. The amplitude is the difference in amplitude between those two cases.

Those are reported as a function of position on a Cartesian grid on Fig. 1.2, **D-G**. Because the incident field in this case is more easily expressed in Cartesian coordinates, the incident field was computed according to Equation 1.B.2. The coordinates were then changed into polar coordinates and the diffracted field was computed according to Equation 1.13. All computations of Hankel, Bessel functions and Legendre polynomials, was done using the development of [Duda and Martens, 1998].

Estimates of the ITD

1.B.4 Low-frequency approximation

In the low frequencies, one can approximate the ITD using a development by [Kuhn, 1977]. Only the first two terms of the series in the expression of the total pressure field Equation 1.B.3 have to be computed, because the convergence of the series is quite fast for $ka \ll 1$ [Duda and Martens, 1998]:

$$\begin{aligned} p(r, \beta) &\approx \\ &\approx 1 - i\frac{3}{2}ka\cos(\beta) \end{aligned}$$

Now, here β is the incidence angle.

We want to compare the pressure field at the left and right ears. The situation is depicted on Fig. 1.13. The left ear has an incidence angle β_l, and the right ear an

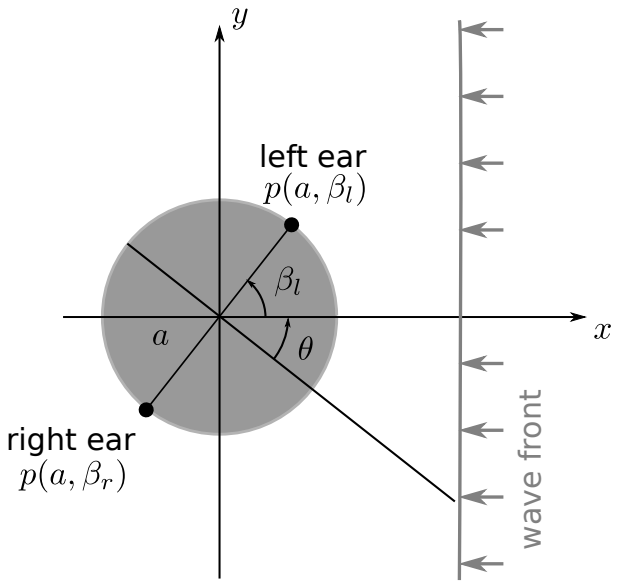

Figure 1.13: Depiction of the situation when computing the low frequency ITD approximation.

incidence angle $\beta_r = \beta_l + \pi$. Hence, the pressures at the left and right ears read:

$$p(ka, \beta_l) = 1 - i\frac{3}{2}ka\cos(\beta_l)$$
$$p(ka, \beta_r) = 1 - i\frac{3}{2}ka\cos(\beta_l + \pi) = 1 + i\frac{3}{2}ka\cos(\beta_l)$$

We can compute the IPD by computing the phases of each pressure field:

$$\angle(p(ka, \beta_l)) = -\angle(p(ka, \beta_r))$$
$$\angle(p(ka, \beta_r)) = \tan^{-1}(\frac{3}{2}ka\cos(\beta_l))$$

The IPD (here measured in radians) reads:

$$\begin{aligned}\mathrm{IPD}(ka, \beta_l) &= \angle(p(ka, \beta_l + \pi)) - \angle(p(ka, \beta_l)) \\ &= 2\tan^{-1}(\frac{3}{2}ka\cos(\beta_l))\end{aligned}$$

Assuming that the argument of the $\tan^{-1}$ function is sufficiently small with respect to $\pi/2$, then the $\tan^{-1}$ function is approximately linear, therefore:

$$\mathrm{IPD}(ka, \beta_l) \approx 3ka\cos(\beta_l) \quad \text{for} \quad \frac{3}{2}ka\cos(\beta_l)) \ll \pi/2 \tag{1.14}$$

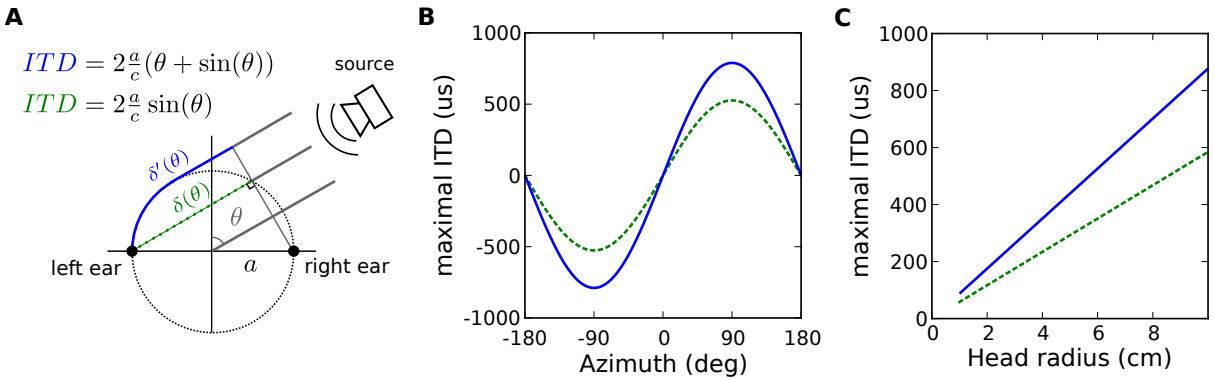

Figure 1.14: Similar plots as Fig. 1.1 with the creeping wave formulation **A** The blue path $\delta'(\theta)$ represents the additional path that high frequency waves travel along the sphere to reach the contralateral ear. The dotted green line recalls $\delta(\theta)$ from the two-mics development. **B** ITD as a function of position as predicted in the creeping waves formulation (blue line) and two-mics situation (dotted green line). **C** Influence of the head radius a on the maximal ITD in the creeping wave formulation (at 90°).

In turn, the phase ITD is the IPD in radians divided by the pulsation:

$$\begin{aligned}\mathrm{ITD}(ka, \beta_l) &= \frac{\mathrm{IPD}(ka, \beta_l)}{\omega} \\ &= 3\cos(\beta_l)\frac{ka}{2\pi f} \\ &= 3\cos(\beta_l)\frac{(2\pi f)a}{(2\pi f)c} \\ &= 3\cos(\beta_l)\frac{a}{c}\end{aligned}$$

It is more common to express this formula using the spherical coordinates system of azimuth and elevation. This can easily be done by replacing the incidence angle β_l by the incidence angle computed from θ and ϕ (Equation 1.10). In the case where $\phi = 0$, the incidence angle directly relates to the azimuth (as seen on Fig. 1.13), and $\cos(\beta_l) = \sin(\theta)$. The expression of the low-frequency ITD is simpler:

$$\mathrm{ITD}(\theta) = 3\sin(\theta)\frac{a}{c} \tag{1.15}$$

Which is the expression obtained by Kuhn [Kuhn, 1977, Kuhn, 1983].

High-frequency approximation: creeping waves

The infinite series Equation 1.B.3 converges more slowly at high frequencies. Thus, one cannot rely on the first few terms of the sum. An alternative solution to the problem, formulated in *creeping waves* provides estimates of the ITD in high frequencies [Kuhn,

1977]. In this formalism, which is only briefly described here, the limit of the shadow boundary on the sphere emits an infinite series of dispersive modes in both directions. Under this hypothesis, the pressure at the ear facing away from the source is expressed as an attenuation wave traveling around the sphere to the ear.

Therefore, the computation of ITD in this limit can be achieved by geometrical considerations, quite close to the case of the two-mics method. The additional path that the wave has to cover is defined as $\delta'(\theta)$, and corresponds to the blue line in Fig. 1.14. Intuitively, the wave approaches the sphere tangentially, and when this point is reached, travels around it. Similarly to the two-mics situation the ITD is given by the difference in propagation time outside of the sphere and on the sphere:

$$\mathrm{ITD}_{\mathrm{HF}}(\theta) \approx \frac{a}{c}(\sin(\beta) + \beta)$$

Where β is the incidence angle, defined as usual (Equation 1.10). This equation is also presented in the main text Equation 1.3.

The dependence of this high-frequency ITD estimator on both the azimuth and the head radius is very similar to that of the geometric ITD of the two-mics model. Both are reported on **B** and **C** of Fig. 1.14 as a function of azimuth and head radius (respectively). Notice that, as it is obvious from Equation 1.3, the ITD in this formulation is always bigger or equal to that predicted by the two-mics model. That is, even in high frequencies a simple geometrical model always underestimates the ITDs imposed by a spherical head.

1.B.5 Analysis of ITDs on the spherical head model

We simulated spherical head HRTFs for the same positions as was done for the human data. We then used those impulse response and computed the same metrics as in the main text. The results are reported on Fig. 1.15 for the upper-left-frontal quadrant. **A** shows the low to high frequency ITD ratio, **B** the low to high ITD difference and finally **C** the transition frequency as a function of position.

The solution of the spherical head model only depends on the incidence angle, as shown in this section. Hence the variation of ITD in a spherical model also displays symmetries, namely, the ITD in fact only depend on the incidence angle of the wave onto the sphere. This means that the incidence angle is an even function of the elevation, and an odd function of the azimuth (in Equation 1.10).

This justifies the apparition of so-called *cones of confusion*, those a cones around the ears where the binaural cues are equal. This is obvious from the ITD statistics (ratio, difference, transition frequency) for the spherical model (as reported on Fig. 1.15),

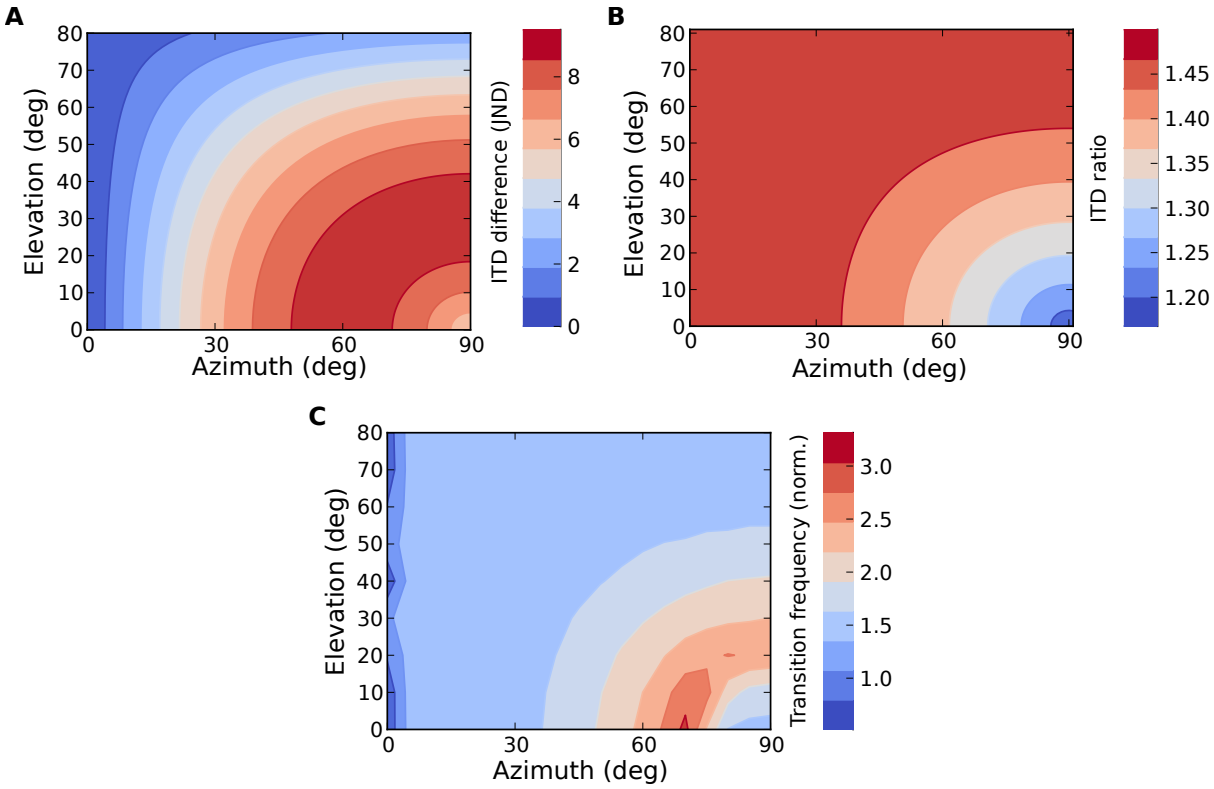

Figure 1.15: The results of the same analysis as for the human databases on the spherical model. **A** corresponds to **A** of Fig. 1.7 **B** corresponds to **B** of Fig. 1.7 **C** corresponds to Fig. 1.8

where the statistics are equal on circles concentric on the origin of the graph. The data can intuitively extrapolated by respecting the equality of the statistics on cones of confusion.

Chapter 2

Envelope and fine-structure ITDs

Contents

In the first chapter, I showed that ITD varied with frequency, even within the bandwidth of a single auditory filter. If it is easy to understand what the effect of a

pure delay is on a given sound signal, the interpretation of non-constant phase delays is more complicated. In this chapter I use signal-processing to explain how one can interpret the effect of frequency-dependent ITDs.

I show that the variability of ITD implies that the monaural filters have have non-constant phase delays. Therefore, mononaural filters have a non-linear phase response. These filters produce *phase-distortion* artifacts in audio. I propose to study this kind of distortion by approximating the non-linear phase response of monaural filters locally (in frequency) by an affine function. The slope of this function, the group ITD $\mathrm{ITD}_{\mathrm{g}}$, is the delay of the envelopes of the two monaural signals. The intercept of this function we define as the *Interaural Iiffraction Index* IDI. The main intuition of this chapter is that the IDI represents an additional phase difference between the fine structures of the monaural signals. I show how this is true by applying such filters amplitude-modulated (AM) signals and passing them through a model of the auditory periphery.

Finally, I explain that those two cues are not independent for one given position, i.e. because they covary negatively. Yet, I argue that they can be extracted from the cross-correlation of the two monaural signals, and therefore convey separate information about the position of the sound source.

2.1 Introduction

2.1.1 Non-linear phase and phase-distortion

Assuming that ITD is constant across the frequency spectrum is convenient for multiple reasons. As we have seen, it is a simplification of the acoustics of the subject, and provides an understanding of the existence of ITDs in graspable terms: acoustical rays, difference in path-length. More relevant to the present part of the dissertation, it also affords a simple understanding of the effect of the ITD on source signals. Indeed (assuming there are no ILDs), one of the two ears receives *the same signal* as the other one, simply delayed in time by the ITD. In other words, there exists a time delay τ such that one signal is the delayed version of the other: $x(t) = y(t - \tau)$. In this context the task of the binaural system is simple: find the delay such that the signals at the two ears are *equal*.

As we have seen in the previous chapter, the total phase ITD depends on frequency. This is in fact because *each* monaural HRTF filters have varying phase delays (data not shown). Equivalently, monaural HRTF filters have non-linear phase spectra. Such non-linear monaural HRTF phase spectra are present in the spherical head model, as well as the structural model for humans HRTF by [Brown and Duda, 1998b]. Moreover, they are most likely the source of the frequency dependent ITDs observed in many studies (see Chapter 3 for a review).

Filters with non-linear spectra introduce phase-distortion in their input signals. Phase-distortion is not the common form of distortion, as characterized by distortion product, or level-dependent effects. Yet, many filters have non-linear phase responses, e.g. the Butterworth filter's phase response is a hyperbolic tangent (tanh) function of frequency (e.g. [Oppenheim et al., 1999]). It is only approximately linear for the high and low frequencies, when the hyperbolic tangeant is approximately constant. Otherwise, in the transition band of a Butterworth filter, the signal is phase-distorted. Note that this is not usually a problem because those frequencies are attenuated by the gain of the filter. To quantify this, the *group delay* is usually defined, it is the derivative of the phase response with respect to the pulsation (2π times frequency). The quantity of phase-distortion imposed by the filter is then measured by the variations of the group delay.

The first organ of the auditory system, the cochlea, transduces acoustical signals into neural signals. The cochlea is usually described as a frequency analyzer, because each point on the basilar membrane of the cochlea is excited by a small band of frequencies. As a result of this, binaural neurons that are responsible for the processing of ITD (e.g, in the cat medial superior olive [Yin and Kuwada, 1983]) only respond to acoustic

signals in a restricted range of frequency. Consequently, it is safe to restrict ourselves to the study of ITDs *locally* around a frequency, representing the center frequency of a neuron's pass-band. I argue that ITD_paround a frequency is best understood as an *affine-phase* delay. Those kind of filters have phase responses that are "affine", and are consequently not strictly speaking *linear*, namely because their phase response does not vanish at zero frequency. These considerations lead to the definition of *two local ITD cues*: the Interaural Diffraction ndex (IDI) and the Interaural Group Delay (ITD_g). Then, I expose a simple mathematical framework that allows us to understand the phase-distortion imposed by frequency dependent delays on band-pass filtered signals. Namely, I show that the group ITD is the total ITD of the envelope of the monaural signals, and that the IDI is a measure of the additional difference in phase between the fine structure of the monaural signals

I finally provide a way to estimate those quantities from HRTF data, and expose the results for one human subject recorded as part of the project of Chapter 1. Finally, I discuss how those cues relate to the position of the sound source, and how they could be extracted independently from the cross correlation of the monaural signals.

2.1.2 Illustration of phase-distortion effects

As an illustration of this phase-distortion, and the analysis I am about to make of it, I provide a simple mathematical example for a stimulus consisting in the sum of two pure tones. This section is meant to be an intuitive approach to the analysis that is conducted in the rest of the chapter.

Let us consider a signal consisting in the sum of two infinitely long pure tones at different pulsations ω_0 and ω_1:

$$s(t) = \sin(\omega_0 t) + \sin(\omega_1 t)$$

Now, if this signal is passed through a filter with a variable total phase delay, with unit amplitude response. This means that each frequency component is delayed by two different phase delays, $\tau(\omega_0)$ and $\tau(\omega_1)$. This delaying is equivalent to dephasing each individual sine component by $\phi(\omega_0) = \tau(\omega_0)\omega_0$ and $\phi(\omega_1) = \tau(\omega_1)\omega_1$:

$$\begin{aligned} y(t) &= \sin(\omega_0(t-\tau_0)) + \sin(\omega_1(t-\tau_1)) \\ &= \sin(\omega_0 t - \phi_0) + \sin(\omega_1 t - \phi_1) \end{aligned}$$

Let us try to find a delay value τ such that the output signal is a delayed version of s. We try to solve the equation $s(t-\tau) = y(t)$ for τ. To solve the equation one has to

identify the sin terms at different pulsations, which leads to:

$$\begin{aligned}\phi_0 &= \tau\omega_0[2\pi] \\ \phi_1 &= \tau\omega_1[2\pi]\end{aligned} \quad \Longleftrightarrow \quad \frac{\phi_0}{\omega_0} = \frac{\phi_1}{\omega_1} = \tau$$

In essence, the output signal is similar to the input provided that the phase response is proportional to the pulsation (hence frequency), here with factor τ. This means that the total phase delays are constant, equal to τ. When this condition is not met, i.e. the total phase delay is not constant, the output signal is phase-distorted.

Now, let us examine the effect of phase-distortion. I consider a case where the phase is an affine function of the frequency:

$$\phi(\omega) = -\tau\omega + d$$

In usual terms, τ is the group delay of the filter, while d in radians is the intercept of the phase response. When d is nonzero, then the phase response is non-linear: it is affine (i.e. it does not vanish at zero). Thus, it is expected that phase-distortion should occur. First, because the signal is the sum of two sinusoidal functions, it is possible to write it as a product of sine and cosine functions:

$$\begin{aligned}s(t) &= \sin(\omega_0 t) + \sin(\omega_1) \\ &= 2\sin\left(\frac{1}{2}(\omega_0 + \omega_1)t\right)\cos\left(\frac{1}{2}(\omega_0 - \omega_1)t\right) = 2\cos(\omega_e t)\sin(\omega_c t)\end{aligned}$$

When the two pulsations are close enough, that is $|\omega_0 - \omega_1|$ is sufficiently small, then it is common to say that the signal $s(t)$ is a pure tone carrier signal, modulated by a low-frequency envelope. Thus, $\omega_c > \omega_e$ is the pulsation of the *carrier signal* and ω_e the pulsation of the *envelope*. If I now compute the result of filtering with an affine phase the signal $s(t)$ I obtain $y(t)$:

$$y(t) = \sin(\omega_0 t + \phi(\omega_0)) + \sin(\omega_1 + \phi(\omega_1))$$

Writing the signal in terms of envelope and fine structure I obtain:

$$y(t) = 2\cos(\omega_e(t - \tau))\sin(\omega_c(t - \tau) + d)$$

In the representation using envelope and fine structure it is clear that the envelope of the signal is delayed only by τ the slope of the phase response of the filter. On the other hand, the fine structure of the signal is delayed by a τ, and then undergoes an additional

phase shift equal to d. Hence I showed that an affine phase delay has differential effects on the fine structure and envelope of this rudimentary signal. Phase-distortion effects can be understood as the existence of such differential delays on envelope and fine structure of the signal. The intercept of the phase response of the filter quantifies its non-linearity (i.e. when it is zero, the filter is linear). I will now use this kind of approximation to describe IPDs locally, to asses the effect of non-linear IPD in terms of interaural envelope and fine structure delays.

2.2 Study of frequency-dependent interaural delays

In this section, the main text only contains the mathematical development through an example model of human frequency-dependent ITDs. The Appendix 2.A will be quoted often, as it contains more details and demonstrations.

2.2.1 Frequency-dependent ITDs: Working example

A human ITD vs. frequency curve is modeled, for simplicity as a sigmoidal curve, with a hyperbolic tangeant (tanh) shape. It roughly corresponds to the situation of a human being of head radius $a = 8cm$. This model is inspired by the human data, but is rather arbitrary in its parameters.

The low and high frequency limits of the ITD are given by the Kuhn and Woodworth formulas (Chapter 1, Equation 1.2 and Equation 1.3) for a source at 90°azimuth. The values here are $\mathrm{ITD}_{LF} \approx 700\mu s$, $\mathrm{ITD}_{HF} \approx 470\mu s$. Furthermore, the transition frequency of the sigmoidal curve is given by the frequency $\hat{f}$ defined in Chapter 1: $\hat{f} = c/2\pi a \approx 680Hz$. Hence the ITD function under study here is given by:

$$\mathrm{ITD}_{\mathrm{p}}(f) = ITD_{LF} - \frac{1 + \tanh(3\frac{f-\hat{f}}{\hat{f}})}{2}(ITD_{LF} - ITD_{HF})$$

It must be noted that the value of the $\mathrm{ITD}_{\mathrm{p}}(0)$ is not exactly ITD_{LF}. The value of the function is plotted on Fig. 2.1, as well as the two low- and high- frequency approximations of the ITD.

I now consider the IPD associated with this $\mathrm{ITD}_{\mathrm{p}}(f)$. It is defined as the unwrapped IPD measured in cycles therefore:

$$\mathrm{IPD}(f) = \mathrm{ITD}_{\mathrm{p}}(f) \times f$$

Because $\mathrm{ITD}_{\mathrm{p}}$ here is a positive function, and has no short-scale variations, the IPD is

a monotonously increasing function of the frequency as shown on Panel B of Fig. 2.1. Furthermore, because the ITD admits two limits near 0 and $+\infty$, the IPD function admits two linear asymptotic limits. They are reported on Panel B of Fig. 2.1 as well (two black lines). Each corresponds to the IPD function if the ITD were a constant function of frequency, with value ITD_{LF} (plain line) and ITD_{HF} (dashed line). Note that the function $\mathrm{ITD_p}(f)$ is not exactly equal to ITD_{LF} at zero frequency, it converges to it as $f \to -\infty$. I will assume that $\hat{f}$ is sufficiently big so that I can make this approximation. Furthermore, the IPD function in any case admits an asymptotic limit at zero, which is what I show here.

2.2.2 Local approximation of $\mathrm{ITD_p}$

As I argued in the introduction, it is safe to study the variation of $\mathrm{ITD_p}$ locally around a frequency, because of the frequency selectivity imposed on the binaural system by the cochlea. If the pass-band of a binaural neuron is narrow enough, the IPD function on this range can be approximated by its tangeant. This is a first-order Taylor series expansion of the function around this point. Doing this, two quantities are defined locally around any f_0 frequency point in the spectrum Fig. 2.1 **D**. The group ITD, $\mathrm{ITD_g}(f_0)$ is the slope of the function at that point, which corresponds to the usual definition of group ITD (as in, e.g. [Roth et al., 1980b]):

$$\mathrm{ITD_g}(f_0) = \left.\frac{d\mathrm{IPD}(f)}{df}\right|_{f_0} \tag{2.1}$$

It is plotted across the whole spectrum on Panel E of Fig. 2.1. The $\mathrm{ITD_g}$ for one position seems to match the $\mathrm{ITD_p}$ when the latter is constant, and deviate from it when it varies. This follows from the definition of $\mathrm{ITD_p}$, plugged in the Equation 2.1:

$$\mathrm{ITD_g}(f) = \frac{d}{df}\left[\mathrm{ITD_p}(f)f\right] = \left[\frac{d}{df}\mathrm{ITD_p}(f)\right] f + \mathrm{ITD_p}(f)$$

Hence, when $f > 0$, the group ITD is greater than $\mathrm{ITD_p}$ when $\mathrm{ITD_p}$ increases and conversely is smaller when it decreases. On **E** of Fig. 2.1 the group delay is always below the $\mathrm{ITD_p}$ because it is strictly decreasing.

I am interested in quantifying the difference between the group and phase ITDs. The Interaural Diffraction Index, IDI is the intercept of the tangeant of the IPD function

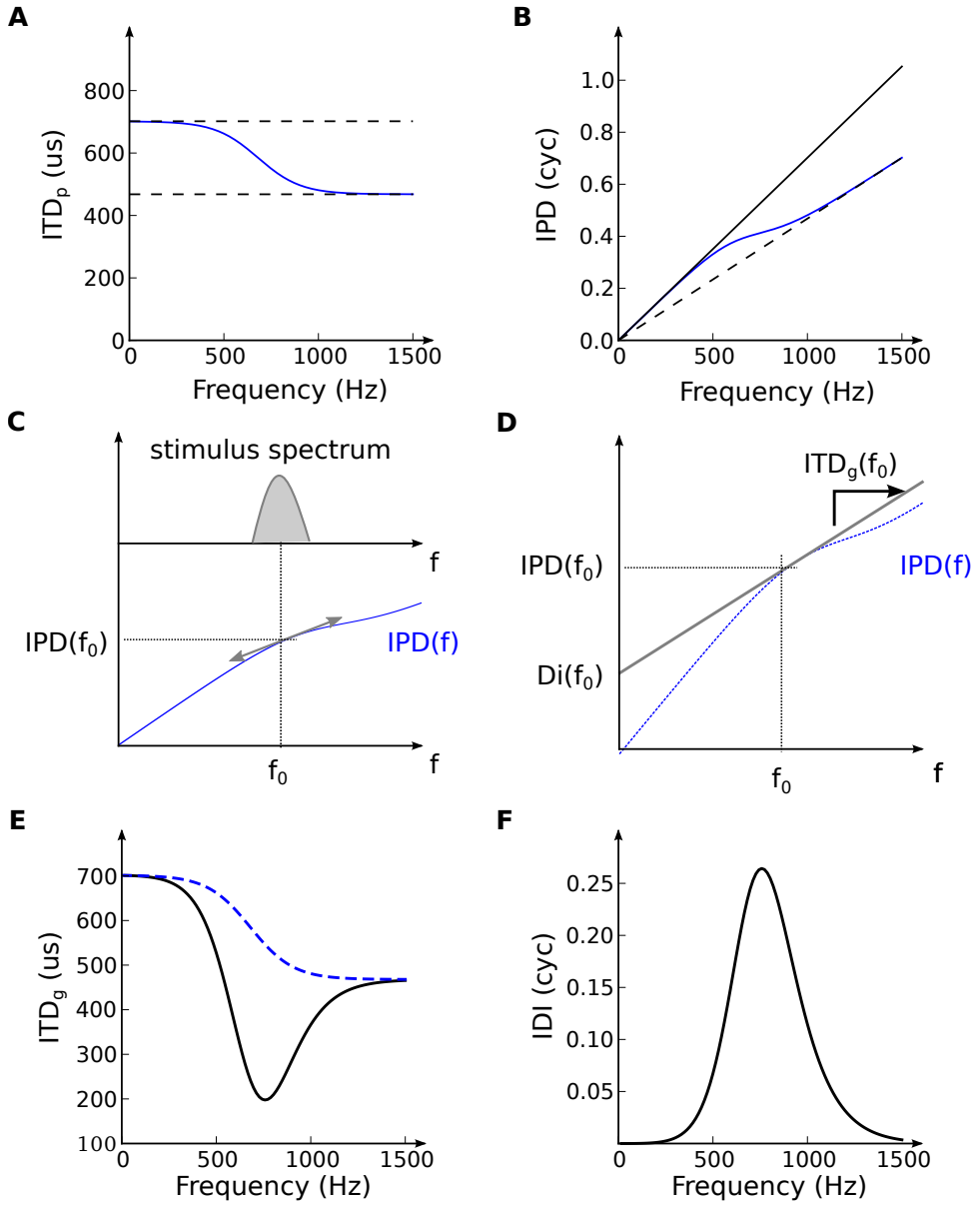

Figure 2.1: **A** The example of ITD function of a modeled human (blue curve). Low and high frequency limits of the ITD function are plotted, in plain and dashed lines respectively. **B** IPD patterns corresponding to the model ITD function (blue curve). Dashed and plain black lines correspond to the phase response of pure delay filters with delay equal to the low and high frequency approximations of the example ITD. **C** Depiction of the situation. If the stimulus spectrum is narrow around f_0 (upper panel, gray area) then it is sensible to approximate the IPD pattern (blue curve) by its tangeant function around this point. **D** The slope of the tangeant is homogeneous to a time and is usually called the group ITD, ITD$_g$. The intercept of the tangeant is homogeneous to a phase, and is termed the diffraction index IDI. **E** The derivative function of the IPD is the group delay, plotted here (black plain curve) as a function of frequency. The dashed blue curve is the total phase ITD as in **A**. **F** The IDI for the example model as a function of frequency, plotted here (black plain curve).

at the frequency point f.

$$\begin{aligned} \mathrm{IDI}(f) &= (\mathrm{IPD}(f) - \mathrm{ITD_g}(f)f) \\ &= (\mathrm{ITD_p}(f) - \mathrm{ITD_g}(f))f \end{aligned}$$

Because IDI is a phase quantity, it is presented *wrapped* in cycle units within the interval $]-0.5, 0.5]$. IDI is plotted as a function of frequency on **F** of Fig. 2.1. We observe that it deviates from zero when the variation of $\mathrm{ITD_p}$ is the greatest, or when the difference between group and phase ITDs is greatest. This occurs, in our model, for frequencies higher than the unit normalized frequency here $\hat{f} = 680Hz$. A non-zero value of the IDI is thus indicative of the transition of the $\mathrm{ITD_p}$ from its two plateaus. As I showed in Chapter 1, is the result of the diffraction of the sound wave around the head For this reason, we have called this quantity the *Interaural Diffraction Index*.

Interestingly, in signal-processing the quantity of phase-distortion introduced by a filter is measured by the variations of group delay. The IDI is a measure of these variations, since when the $\mathrm{ITD_g}$ is constant over a sufficiently big range, it is equal to the $\mathrm{ITD_p}$. Therefore the IDI quantifies the degree of phase-distortion in narrow band signals due to diffraction.

2.2.3 Fine structure and envelope ITDs

Let us now turn to the signal processing interpretation of IDI and $\mathrm{ITD_p}$. For simplicity, we construct one filter whose phase response is an IPD modeled as an affine function with parameters $\mathrm{ITD_g}$ and IDI. Such a filter would correspond to the Interaural Transfer Function associated with an HRTF pair $ITF = \mathrm{H}^R/\mathrm{H}^L$. Assuming that s^L is the signal at the left ear, applying the ITF to s^L yields s_r. Because it is a common stimulus for auditory electrophysiology (see, e.g. [Joris et al., 2004]), I take the example of an amplitude modulated pure tone. This *original* signal can be thought of as the signal at the left ear.

Results shown in the Appendix 2.A prove that such a filter imposes separate delays on the envelope and the fine-structure of the signal. These interaural delays can be recapitulated as follows:

- The whole signal (envelope + fine structure) is *delayed* by the value of $\mathrm{ITD_g}$
- The fine structure of the delayed signal is *phase shifted*, by the value of IDI

An example of this is shown on Fig. 2.2. The original signal (**A**) is a 100 Hz carrier modulated by a 40 Hz envelope. It is also ramped up and down with a raised cosine

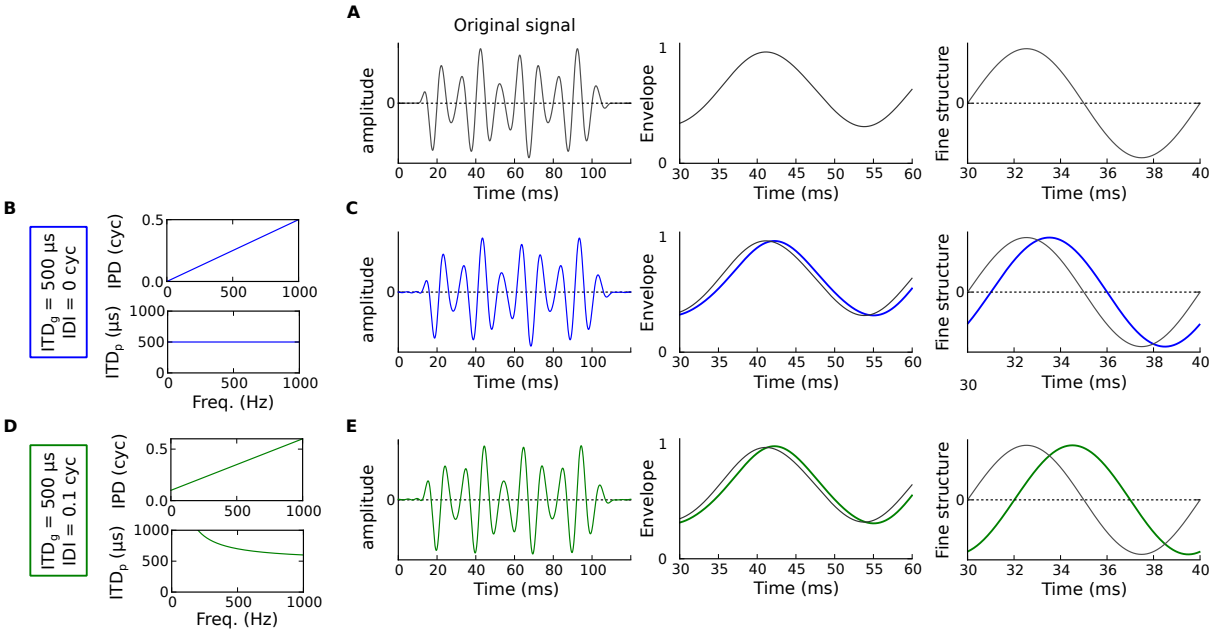

Figure 2.2: Schematic effect of affine phase filters. An original amplitude-modulated signal **A**, has a 100 Hz carrier and is modulated by a 10 Hz envelope. First panel is the signal, next two are inserts of the envelope and fine structure respectively. This signal is passed through an affine phase filter with $\mathrm{ITD_g} = 500\mu$s and $DI = 0$, that is a linear filter. The IPDand $\mathrm{ITD_p}$of this filter are reported as a function of frequency on **B**. Graphs on the line of **C** are similar to **A**, they show the signal, its envelope and fine structure. **B** Left plot: the signal, as delayed by a pure delay of group delay $\mathrm{ITD_g} = 500us$. Both the fine structure and the envelope of the signal are delayed by the same amount. Right plot: notice that the envelope and fine-structure are still in the same relative phase. **D** and **E** show the same as **B** and **D** except IDI = 0.1 cycles. In both filter case, the envelope is delayed by the same amount on the middle row panels. Yet, in the presence of a non zero IDI (bottom-most panels) the fine structure undergoes an additional delay.

ramp $r(x) = cos^2(x)$. On **B**, I show the model ITD_p and IPD for and affine filter with zero IDI and a constant $ITD_g = 500\mu s$. As seen on **C**, both the envelope (middle and rightmost row) and fine structure are delayed by the same amount (timescales are not the same). The same analysis is applied on **D** and **E**, this time with a filter with IDI = 0.1 cycles and the same ITD_g. As reflected on **E** (middle row), the envelope is delayed by the same amount as with IDI = 0 cycles.

Yet, the fine structure of the signal is not the same as in the previous case, showing that introduction of a non zero IDI imposes different interaural delays on the envelope and fine structure of the signal. In fact, as shown in Appendix 2.A, it undergoes an additional phase shift. Going back to the total phase ITD (**B** and **D**, bottom insert), when a non zero IDI introduced, the phase response is affine strictly (top inserts). This translates in a hyperbolic shaped delay, because each pure tone undergoes a constant *phase shift*. Therefore, the effect of IDI when measured in time, is bigger for lower frequencies. For example, a half-period phase-shift represents $500\mu s$ at $1000Hz$ but $5ms$ at $100Hz$.

2.3 Application to human HRTFs

As shown in the previous chapter (Chapter 1), human ITDs are frequency-dependent for a given position in space. It is now possible to quantify the cues that derive from this property, namely the interaural envelope delay as measured by ITD_g and the additional interaural phase shift measured by IDI. This analysis was performed on human subjects using acoustical data measured as described in the previous chapter. The method used to compute ITD_g and IDI is explained in Appendix 2.B.

One human subject's HRTFs were measured, and the ITD_p computed as described in Chapter 1. This is reported for four positions in the horizontal plane (Fig. 2.3, **A**) on **B** of Fig. 2.3. ITD_g and IDI are reported for the same positions and frequencies on **C** and **D**. The ITD_g displays a clear dependency on position: for all frequencies, the ITD_g is an increasing function of position. As a function of frequency, ITD_g oscillates around the ITD_p curve, which is consistent with our theoretical developments. The group ITD ITD_g is greater than the phase ITD ITD_p when ITD_p is increasing, and vice-versa.

The variation of IDIwith position is less trivial (**D**), because at some frequencies, it seem to be a decreasing function of eccentricity. As a function of frequency, IDIoscillates around zero, an is in general quite low. This is because IDI is proportional to the difference between ITD_p and ITD_g. Notice that relatively low IDI may represent quite long delays when in the time domain. For example for a tone at 500 Hz, the IDI for position 30°is of 0.15, which represents a delay of $0.15/500 = 300\mu s$. Finally, IDI has a

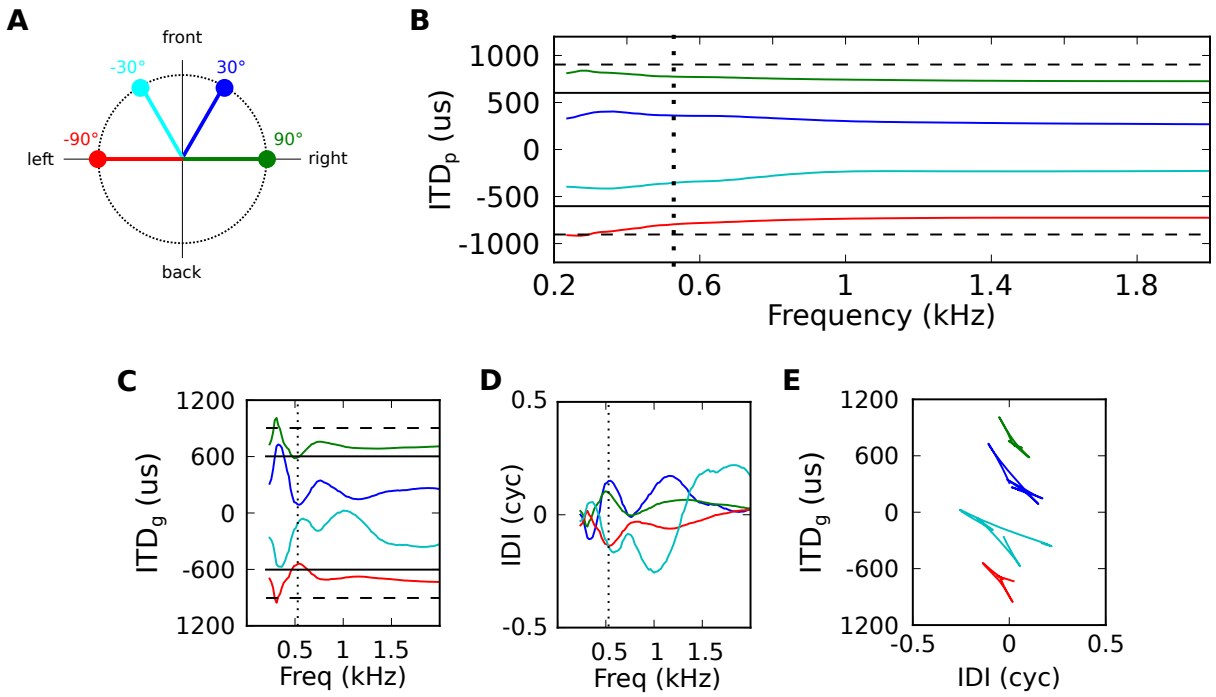

Figure 2.3: ITD_g and IDI for one human subject. **A** A picture of the coordinates presented on this plot. **B** ITD_pas measured for a human subject (main author) **C** ITD_gas a function of frequency for the same coordinates. **D** IDIas a function of frequency for the same coordinates. **E** The group ITD as a function of IDI at all frequencies on the ITD_g, IDIplane.

significant peak around 600 Hz, consistent with observations that this was the frequency at which ITD_p varies the most, and the developments of the previous sections. Because in this range, the effect of non zero IDIcan cause long additional delays on the fine structure of the signals (one period of a tone at 100 Hz one millisecond).

On **E** the ITD_g is plotted against IDI for the same positions (**A**) and frequencies as in the previous panels. This visualization allows us to see that for any given position, ITD_g and IDI measurements across the spectrum are negatively correlated. This can be explained intuitively, as reported in Appendix 2.C, because of the very definition of those quantities. As a result of this, for each position the (ITD_g, IDI) curve clusters along a negative trend. This is observed for all positions, even though it is reduced as the position is closer to the midline. The relationship is not completely trivial (see Appendix 2.C, and both cues do not lie on a segment (**E** of Fig. 2.3).

In conclusion, ITD_g and IDI are cues to the location of the sound source present in human recordings. IDI translate into delays of sufficient magnitude to be This is despite the fact that they covary to some degree for a given position.

2.4 Discussion: ITD_g, IDI and cross-correlation

As acoustical data suggests, when the sound source emits a sound in space, the envelope and fine structure of signals at the "output" of the cochlea have different envelope and fine structure delays. At each ear, those delays are different, creating two independent frequency-dependent ITD cues, the IDI and ITD_g. Those cues can be recovered by measuring independently ITD_p and ITD_g. The total phase ITD ITD_pcan be computed from the maximum of the cross-correlation of the monaural inputs: it is the position of the peak of this function. A lesser known result is that the group ITD is the position of the peak of the cross correlation *of the envelopes* of the signal. This is shown, for example in [Marple Jr, 1999]. As a consequence of this, all three timing cues can be extracted from a representation of the cross correlation of the binaural signal. In this section, we model the cochlea as a set of Gammatone filters, and extract those cues from the cross correlation of each frequency band.

A depiction of the model is shown on **A** of Fig. 2.4. A 100 ms long white noise input is filtered by HRTF filters, simulating the effect of head and torso. The HRTF filter is taken at $\theta = 70°$ in the horizontal plane. Each monaural pathway consists in a bank of Gammatone filters, with bandwidths on an ERB scale and simulated using the Brian.Hears package [Fontaine et al., 2011]. In each frequency band, the cross-correlation of the signals on both sides is computed (Fig. 2.4. An example of such cross correlation, for the filter with center frequency $CF = 600Hz$, is shown on **B** (blue

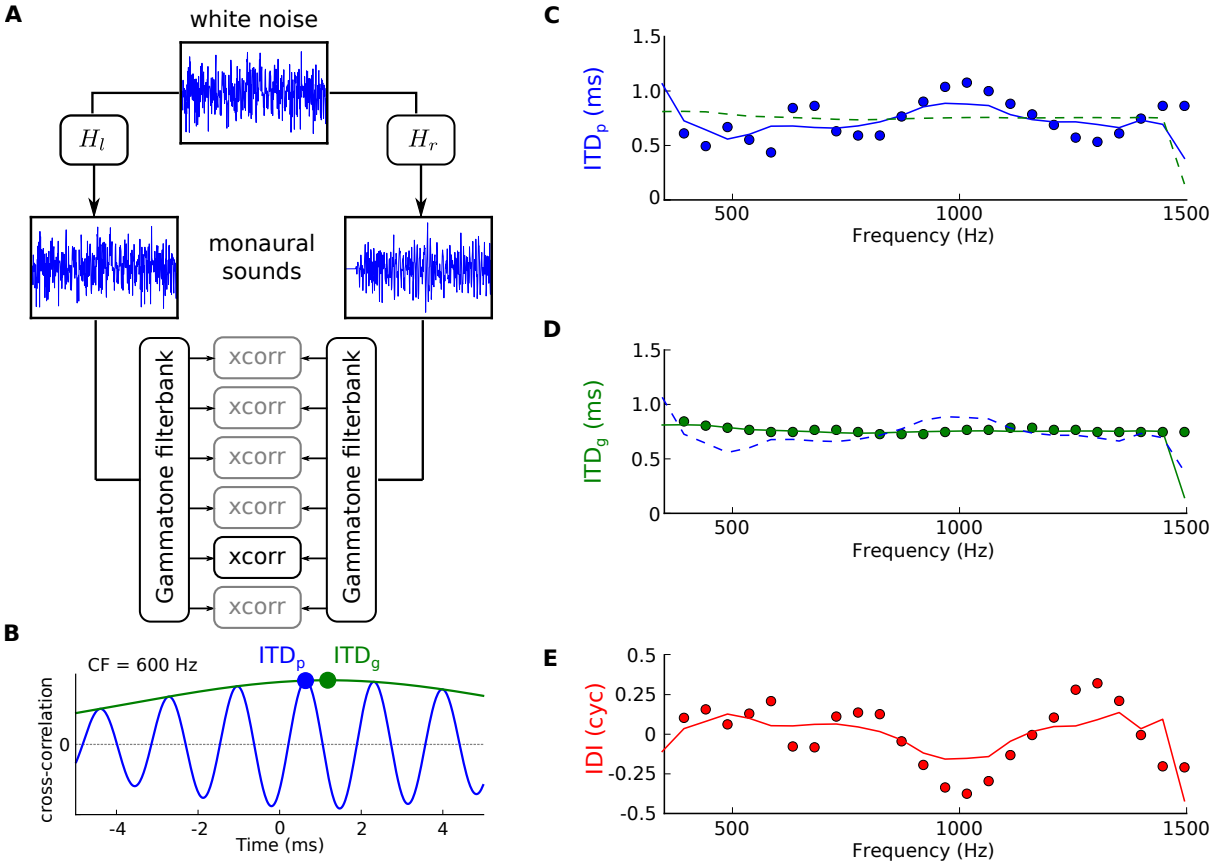

Figure 2.4: Recovering timing cues from the cross correlation of the monaural signals. **A** Schematics drawing of the situation. A white noise input is filtered through two monaural HRTF filters. Each is then band-passed filtered by a bank of Gammatone filters. Channels on both sides are then cross-correlated. **B** The results of the cross-correlation are presented here, the cross-correlation (blue) and its envelope (green). Each maximum represents the envelope or fine structure delays. The IDI is the difference between those two delays, divided by the frequency of the carrier. **D** The total phase delay predicted (blue line) and measured from the cross-correlated signals (blue circles). Green dashed curve is the predicted group delay (as in **E**). **E** The total phase delay predicted (blue line) and measured from the cross-correlated signals (green circles). Blue dashed curve is the predicted total phase delay (as in **D**). coordinates. **F** The predicted D (red line) and measured from the cross-correlated signals (red circles).

curve), alongside the envelope of the cross-correlation function (green curve). The maximum of the cross-correlation provides a measure of ITD_p at that frequency. This is reported for one simulation on **C**, blue dots. The peak of the envelope curve is the delay between the envelopes of the signals, and it is reported on **D** (green dots). This total envelope ITD should correspond to the the group ITD_g [Oppenheim et al., 1999]. From both the envelope and fine structure ITDs, the IDIis computed by dividing the difference by the center frequency of the filter. This data is reported on **E**, as red dots for the same simulation.

As we have seen in the previous sections, those delays can be predicted from the phase response of the monaural filters. The binaural measures IDIand ITD_gwere also computed from the HRTF data and are reported on Fig. 2.4 (plain lines on **B**-**D**), alonside the simulation results. The prediction of the envelope ITD on **D** provides a good fit to the cross-correlation data. ITD_p(and thus IDI) is not as good of a predictor to the position of the peak of the cross-correlation (**C**). This validates the measure and interpretation of the interaural delays that have been defined and studied in this chapter.

From a full representation of the cross correlation of the two signals it is thus possible to extract both ITD_g and IDI. This makes it plausible that the binaural system processes those two quantities separately. Since those quantities cluster nicely even across frequencies (Fig. 2.3), a single ITD_g IDI measure on the whole spectrum is enough to uniquely identify a position in space. This is not the case with the phase ITD, because a given ITD_p signals different positions depending on the frequency content of the signal.

2.A Mathematical developments

The rest of this section is dedicated to understanding the effects of filters with non-linear phase response. The phase response of a filter is an arbitrary differentiable function ϕ of frequency of the frequency f. In the main text, the case where ϕ is the IPD of the ratio of two monaural HRTF filters was shown, here I provide a more complete study in the general framework of arbitrary linear filters.

2.A.1 First-order phase approximation

Let us consider then the first order approximation of the phase delay around a given frequency f_0. Using a Taylor series expansion and rearranging one obtains:

$$\begin{aligned}\phi(f) &= \phi(f_0) + \tfrac{\mathrm{d}\phi}{\mathrm{d}f}(f_0)(f - f_0) + O(f - f_0) \\ &= [\phi(f_0) - \tfrac{\mathrm{d}\phi}{\mathrm{d}f}(f_0)f_0] - [\tfrac{\mathrm{d}\phi}{\mathrm{d}f}(f_0)]f + O(f - f_0)\end{aligned}$$

Where we can identify two quantities of interest, the first one is the well-known group delay [Oppenheim et al., 1999] termed τ_g:

$$\tau_g(f) = \frac{1}{2\pi}\frac{\mathrm{d}\phi}{\mathrm{d}f}(f)$$

The remainder (the zeroth order component) I term the diffraction index, because in the case of the ITD (in a spherical model), this quantity is non zero where diffraction occurs (i.e., here, when τ_p varies). It is a phase quantity measured in radians between $-\pi$ and π:

$$\mathrm{D}(f) = \phi(f_0) - \frac{\mathrm{d}\phi}{\mathrm{d}f}(f_0)f_0$$

Geometrically, those quantities are respectively the intercept and the slope of the tangeant of the phase response curve around f_0, as in Fig. 2.1.

$$\phi(f) = \mathrm{D}(f_0) + \tau_g(f0)f$$

2.A.2 Affine phase filters

Let us now consider a perfect band pass filter with an affine phase. That is, the transfer function of the filter reads:

$$\mathrm{g}(f) = \mathbb{1}_{PF}(f)e^{\phi(f)} \ where \ \phi_{\mathrm{g}}(f) = 2\pi\tau_g f + \phi_0 \ and \ PF = \{f_c - BW/2, fc + BW/2\}$$

For now, let us consider that the filter is an all pass filter, i.e. $PF = \mathbb{R}$. The effect of such a filter on a pure tone can be understood easily by considering Fourier's convolution theorem

$$\begin{aligned} g(t) * \sin(\omega_c t) &\underset{\mathcal{F}}{\Longleftrightarrow} G(\omega)(j\pi[\delta(\omega+\omega_c) - \delta(\omega-\omega_c)])] \\ &\underset{\mathcal{F}}{\Longleftrightarrow} e^{j\tau_g\omega + j\phi_0}(j\pi[\delta(\omega+\omega_c) - \delta(\omega-\omega_c)])] \\ \sin(\omega_c(t-\tau_g)+\phi_0) &\underset{\mathcal{F}}{\Longleftrightarrow} e^{j\tau_g\omega f}(j\pi e^{j\phi_0}[\delta(\omega+\omega_c) - \delta(\omega-\omega_c)])] \end{aligned}$$

We can see that the pure tone undergoes two separate delays, a group delay of value τ_g and an additional phase shift of value ϕ_0. Hence the total phase delay depends on the signal frequency and can be computed. Consider that:

$$\begin{aligned} g(t) * sin(\omega_c t) &= sin(\omega_c(t-\tau_g)+\phi_0) \\ &= sin(\omega_c(t-(\tau_g+\phi_0/\omega_c)) \end{aligned}$$

The total phase delay finally reads:

$$\tau_p = \tau_g + \frac{\phi_0}{\omega_c} = \tau_g + \frac{\phi_0}{2\pi f_c}$$

2.A.3 Effect on AM stimuli

A more useful thing to do is to consider the effect of such a filtering on an amplitude modulated stimulus. Let $s(t)$ a signal as defined with its temporal envelope $a(t)$ and carrier frequency f_c

$$s(t) = a(t)\sin(2\pi f_c t) \underset{\mathcal{F}}{\Longleftrightarrow} G(\omega)\,[A(\omega) * (j\pi[\delta(\omega+\omega_c) - \delta(\omega-\omega_c)])]$$

The result of passing this stimulus through an affine phase filter is straightforwardly computed by the following sequences of identitites:

$$
\begin{aligned}
(g * s)(t) &\overset{\mathcal{F}}{\Longleftrightarrow} G(\omega)\left[A(\omega) * (j\pi[\delta(\omega+\omega_c) - \delta(\omega-\omega_c)])\right] \\
&\overset{\mathcal{F}}{\Longleftrightarrow} e^{j(\phi_0+\tau_g\omega)}\left[A(\omega) * (j\pi[\delta(\omega+\omega_c) - \delta(\omega-\omega_c)])\right] \\
\delta(t-\tau_g) * (a(t)\sin(\omega_c(t)+\phi_0)) &\overset{\mathcal{F}}{\Longleftrightarrow} e^{j(\tau_g\omega)}\left[A(\omega) * (j\pi e^{j\phi_0}[\delta(\omega+\omega_c) - \delta(\omega-\omega_c)])\right] \\
a(t-\tau_g)\sin(\omega_c(t-\tau_g)+\phi_0) &\overset{\mathcal{F}}{\Longleftrightarrow}
\end{aligned}
$$

Hence it is observed that, as expected, the group delay acts on both the envelope and the fine structure of the stimulus. The whole signal is delayed by this same amount, and in the case of linear phase filters it is the only delay happening. In the case of affine phase delays, the fine structure undergoes an additional phase shift, corresponding to the D of the filter.

2.B Estimating IDI and ITD_g

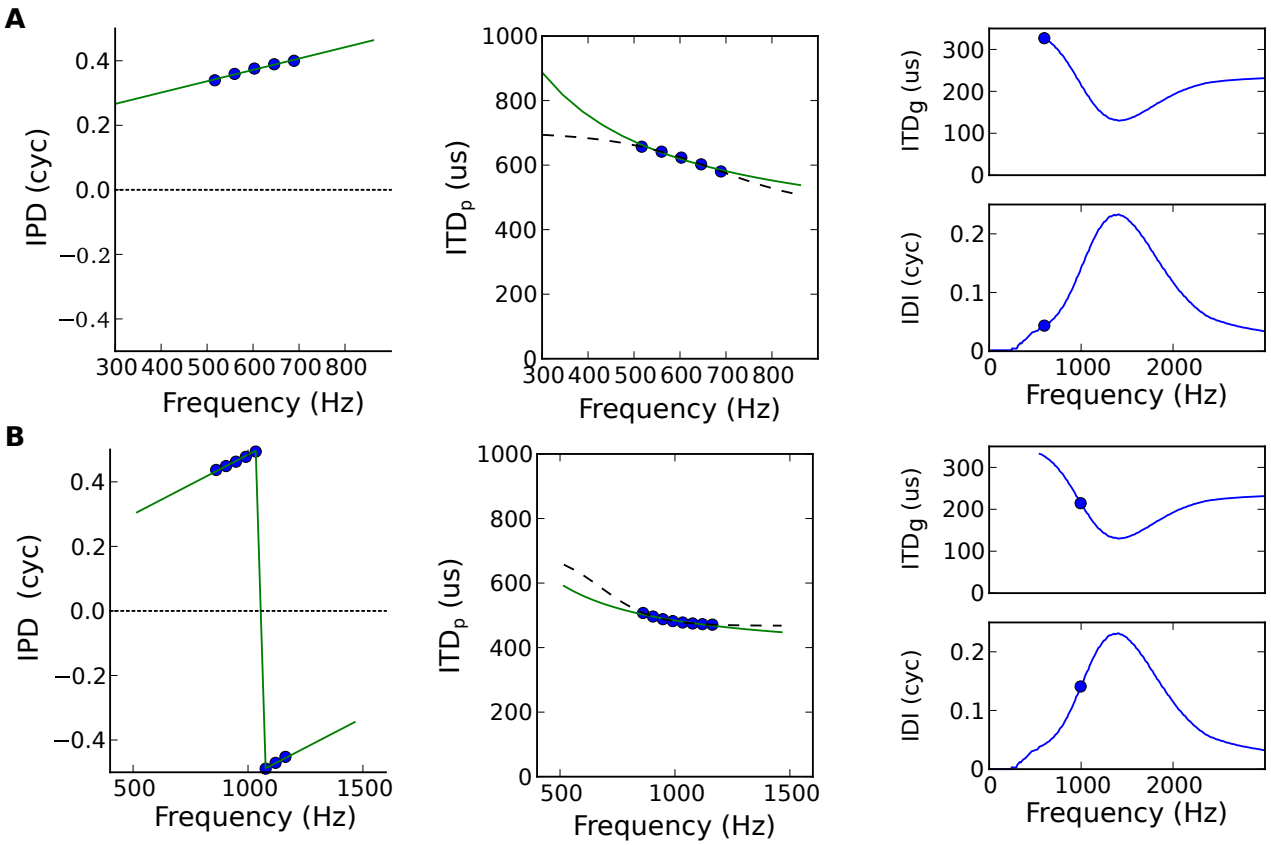

Figure 2.5: Circular linear regression of ITD curves. Each panel illustrates the computation of ITD_g and IDI in our example modem, at two different center frequencies (first and second row).. In phase space (leftmost panels), the IPD (blue circles) is regressed against frequency with an affine function (green line). This is equivalent to fitting the phase ITD ITD_p (middle panel) against frequency with an hyperbolic curve. Black dashed curve represents the full ITD_p response. Rightmost panels: The computed ITD_g and IDI curves for the model ITD curve. The circles depict the point at which the fit is taken in the panel. **A** Center frequency of 600 Hz, Q = 2 **B** Center frequency of 1000 Hz, Q = 2

2.B.1 Circular-linear regression of IPD

The estimation of IDI and ITD_g from the phase spectrum of the HRTFs is not necessarily trivial. This arises from multiple reasons: on discretized sets of measurements, the unwrapping operator is not perfect. For example, in the case of very large delays, it might be that the unwrapping underestimates the actual phase shift response. Secondly, in noisy data, the unwrapping process is unstable, and errors in unwrapping accumulate as one goes up in frequency. Therefore I perform circular-linear fits of the

IPD data against frequence to estimate $\mathrm{ITD_g}$ and IDI from the wrapped phase response of the HRTFs. This is not completely failsafe, as the least square problem typically admits many local minima.

On Fig. 2.5, the computation of $\mathrm{ITD_g}$ and IDI is illustrated at two different frequency points. On a bandwidth with a constant Q value (here $\mathrm{Q} = 4$), the IPD points (blue circles) are linearly regressed against frequency. In **B** an example where frequency wrapping arises is shown, because the algorithm works with circular distances this is no real issue. On the middle panel of each row in Fig. 2.5 the equivalent delay fit is shown, alongside the original $\mathrm{ITD_p}$ curve (black dotted line). The rightmost panel shows the result of this computation on the whole spectrum. The blue dot on each line shows the point where the regression was taken in the leftmost panel.

2.B.2 Fitting procedure

Assume that we are to compute $\mathrm{ITD_g}$ and IDI for a given center frequency CF and bandwidth BW. The passband is defined as:

$$PB = \{f || f - \mathrm{CF}| \leq \mathrm{BW}/2\}$$

The next step is to compute a circular-linear regression of the $\mathrm{IPD}(f)$ function for the frequency points in PB. The IPD data is inherently circular: it is a phase value measured in cycles in the interval $[-0.5, 0.5[$. This means that in order to perform a linear fit, one has to be careful with phase jumps, i.e. when in the passband two consecutive phase values are separated by at least π. One way to alleviate that issue is to *unwrap* the IPD function, that is replace every π-or-more jump by its 2π complement. The problem with this method is that it is quite unstable, and the threshold (biggest allowed consecutive jump) is generally arbitrarily set, which adds a free parameter to the analysis. Hence I use a circular linear regression, this kind of regression allows one to regress a circular value (here, IPD) against a linear value (here, frequency). It can be formulated as a non linear least square problem as follows:

$$\mathrm{IDI}, \mathrm{ITD_g} = \mathit{argmin}_{\mathrm{IDI},\mathrm{ITD_g}} \sum_{f \in PB} \|\mathrm{IPD}(f) - (\mathrm{IDI} + \mathrm{ITD_g} f)\|^2$$

Where the norm used here is a circular norm over phases measured in cycles, that is:

$$\|x\|^2 = \frac{1 - \cos(2\pi x)}{2}$$

This norm is more easily understood geometrically. Assume that we have two angles θ and ϕ measured in radians and the two associated complex numbers $z_\theta = e^{i\theta}$ (and $z_\phi = e^{i\phi}$) with unit modulus. Those complex numbers can be interpreted as vectors that form a θ (and respectively ϕ) angle with the first plane axis.

A distance between the two angles can be inferred from the dot product of those vectors. If θ and ϕ are equal, then the two vectors are colinear and in turn the dot product is maximal (and equal to one because the vectors are of unit amplitude). If, on the opposite, θ and ϕ are in antiphase $\theta = \phi + \pi$ then the two vectors are opposed and the resultant dot product is -1. The dot product of two complex vector reads:

$$\begin{aligned} z_\theta \cdot z_\phi &= ||z_\theta|| ||z_\phi|| \cos(\theta - \phi) \\ &= \cos(\theta - \phi) \end{aligned}$$

Because by construction $||z_\theta|| = ||z_\phi|| = 1$. Furthermore, we want that the defined norm be always positive, and between 0 and 1. Because the cos function varies between -1 and 1, we do an affine transformation and let $||x||^2 = (1 - \cos(2\pi x))/2$, which will always be in the $[0, 1]$ range, minimal ($= 0$) when the angles are equal, and maximal ($=1$) when they are in antiphase.

2.C Covaration of IDI and ITD_g

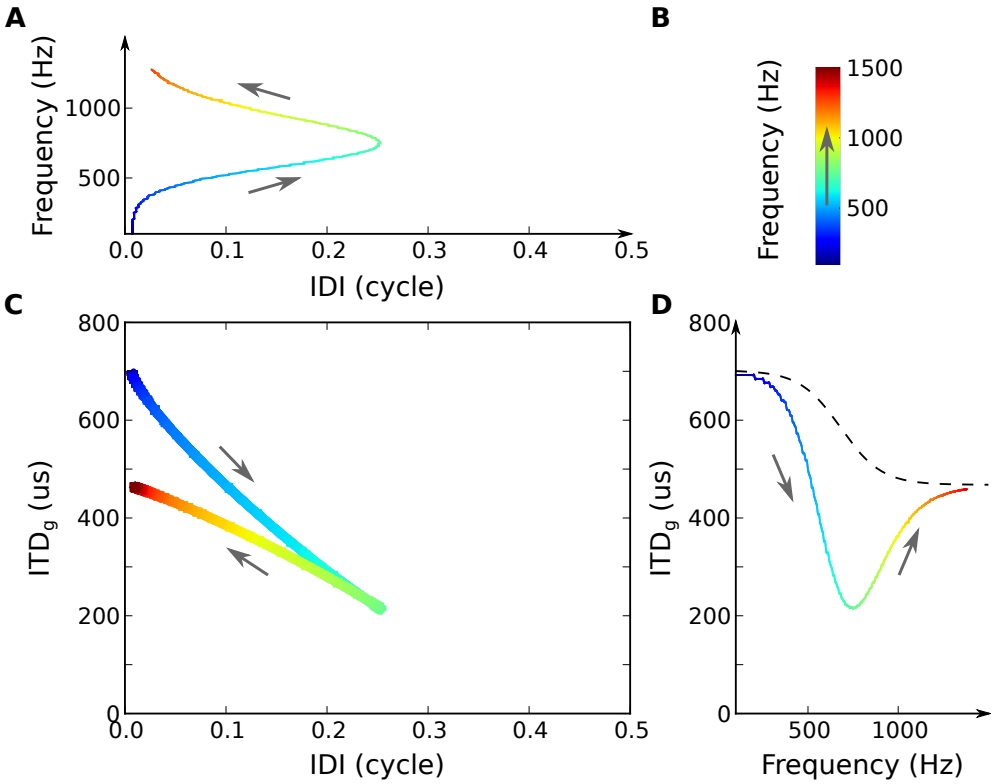

Figure 2.6: Depiction of the inverse covariation of ITD_g and IDI for a given position. Each point color corresponds to a frequency, as in the colorbar on **B**, and the grey arrows point along the direction of increasing frequencies on each line **C** Plot of the ITD_g against IDI across all frequencies for one position. The color code is the frequency, as reported in **B** and consistent throughout panels. **A** and **C** recall the individual curves as a function of frequency. At all frequency points, the curve (IDI, ITD_g) has a slope parallel to the second diagonal. It has one special point where it disconstinuously changes value. Note that in LF and HF, the curve goes back to the line $IDI = 0$ at a point that corresponds to the ITD_plimit.

An interesting point to study is the covariation of IDI and IDI for a given position across frequencies. That is I study the curve $(IDI(f), ITD_g(f))$ for all frequencies. When the frequency is very low, then the IPD curve is almost linear with slope ITD_{lf}, hence the IDI tends to zero while the ITD_g converges to ITD_{hf}. Similarly, in HF, the curve in the IDI, ITD_g-plane goes back to the line $IDI = 0$ to reach it at the ITD_{hf} point. This is illustrated on figure Fig. 2.6. The derivative vector, giving the direction of the curve

in the $\mathrm{IDI}, \mathrm{ITD_g}$-plane reads:

$$\begin{aligned}(\frac{d}{df}IDI, \frac{d}{df}\mathrm{ITD_g}) &= (\frac{d}{df}\mathrm{IPD}(f) - (\frac{d^2}{df^2}\mathrm{IPD}(f)f + \frac{d}{df}\mathrm{IPD}(f)), \frac{d^2}{df^2}\mathrm{IPD}(f)) \\ &= (-\frac{d^2}{df^2}\mathrm{IPD}(f)f, \frac{d^2}{df^2}\mathrm{IPD}(f))\end{aligned}$$

When f is positive, this vector will always be such that it is pointing to the lower-right or upper-left direction on the $(\mathrm{IDI}, \mathrm{ITD_g})$ plane shown on Fig. 2.6, **C**. Therefore, for a given position there is a negative correlation between $\mathrm{ITD_g}$ and IDI measurements at different frequencies. Importantly, this means that $\mathrm{ITD_g}$and IDIare not completely independent cues to location, because of their covariation. This is related to the fact that the estimators for slope and intercept in a linear regression covary (see the appendices to Chapter 7).

Chapter 3

Variation of ITD in animals

Contents

The variation of ITD across frequencies for a given position has been observed in many animals, including the cat [Roth et al., 1980b]. Yet, across-species study of acoustical cues, and specifically the variation of ITD across frequencies exist to date. The study of cues available to laboratory animals is just as important as in humans, if not more, because those laboratory animals are good models for electrophysiology. Furthermore, the range and the nature of possible ITDs can help modelers understand the way ITDs are coded by binaural neurons.

In this chapter, a thorough study of the binaural timing cues available to a wide range of species is presented. I measured HRTFs on stuffed animals, and applyied

the same approach as in the first chapter of this part to assess the ITD variation with frequency. I show that the ITD varies in mammals, in a way consistent with humans case. Finally, I refine the notion of physiological range, and provide ecological distributions of ITD for all measured species.

3.1 Introduction

3.1.1 Existing animal HRTF studies

Numerous studies have recorded animal HRTFs in relation with all aspects of binaural hearing. Because of the animal's individual shape, the monaural acoustical filters depend on the position of the sound source, which constitutes an interesting binaural cue. It is often observed that high-frequency spectral dips, "notches" are present in the monaural HRTFs. The frequency of those notches depend on the elevation of the sound source, therefore it is believed to be a cue to elevation. In the rhesus monkey [Spezio et al., 2000b], those spectral notches are also present at frequencies higher than 9 kHz. They are primarily due to the pinnae, which is of a small enough size to interact with waves at those frequencies. In the cat [Koka et al., 2008a], and the chinchilla [Koka et al., 2011a], removal of the pinnae results in elimination of spectral monaural notches. Similar conclusions are drawn in the rabbit [Kim et al., 2010a], where the effect of distance on the binaural cues are also discussed.

Another interesting aspect is the evolution of ILD for a given position across the spectrum. In most species, ILD is a monotonous function of frequency for a given position, influenced by the presence of the pinnae, and its orientation. In the barn owl realm, studies focus on the importance of the inter-aural canal in the binaural cues or on the facial ruff of the animal [Campenhausen and Wagner, 2006]. HRTFs are also sometimes studied in the context of the development of the localization system [Tollin and Koka, 2009c], both for monaural and binaural cues.

In most studies that recorded animal HRTFs, the focus is generally on one of the two aforementioned lines of study: monaural cues and ILDs. Yet, measures of the ITD in animals are well documented. The frequency-dependence of ITDs has been first demonstrated in the cat [Roth et al., 1980b], by measuring ITDs on pure tone stimuli. The authors report that the maximal ITDs are in general found to be greater in low than high frequencies, consistent with the spherical model. Yet, the ITDs found using the classical estimates of ITD and a measure of the animal's head radius are *consistently smaller* than measured acoustical ITDs. This indicates that the spherical model *underestimates* the magnitude of the ITD. All studies of ITD in other mammalian species are consistent with the spherical model story, with a slightly bigger head radius. As reviewed on Tab. 3.1.1, the same conclusions can be drawn from data in guinea pigs [Sterbing et al., 2003a], rats [Koka et al., 2008a], chinchillas [Koka et al., 2011a], rabbits [Kim et al., 2010a], macaque monkey [Spezio et al., 2000b].

Other interesting data in the rabbit point that ITDs decrease with increasing distance [Kim et al., 2010a] (also with increasing frequency), but also that the rather

Animal	ITD_{LF}, μs	ITD_{HF}, μs	Head radius	Reference
Barn owl	$300\mu s$	300 μs	3.5 cm	[Campenhausen and Wagner, 2006]
Barn owl				[Keller et al., 1998]
Cat	$350\mu s$ (300 Hz)	280 μs (7 kHz)	2.8 cm	[Roth et al., 1980b]
Cat	$390\mu s$ ($<$ 3.5 kHz)		30 mm	[Tollin and Koka, 2009c]+
Chinchilla	$340\mu s$ (250 Hz)	220 μs (4000 Hz)	17.5 mm	[Koka et al., 2011a]
Guinea pig	$330\mu s$ (400 Hz)	250 μs (1500 Hz)		[Sterbing et al., 2003a]
Mongolian gerbil	$130\mu s$ (1 kHz)	100 μs (10 kHz)	17 mm	[Maki and Furukawa, 2005a]
Rabbit	$350\mu s$	200 μs	28 mm	[Kim et al., 2010a]
Rat	$158\mu s$	$100 \pm 25\mu s$	18 mm	[Koka et al., 2008a]
Rhesus monkey	$500\mu s$ (1000 Hz)	400 (10 kHz)	52 mm	[Spezio et al., 2000b]
Humans				Chapter 1

Table 3.1: Summary of the available ITD recordings in animals species in the literature. In parenthesis the approximate Frequency at which the ITD is estimated. Methods used vary, yet the trend is clear

voluminous pinnae of the rabbit also enhances ITDs. A similar effect is seen in the cat by [Tollin and Koka, 2009c].

3.1.2 Motivation of the study

Recordings of animal HRTFs also also useful for the interpretation of binaural electrophysiological experiments. In many situations, small mammals are used as models for testing the neural basis of sound localization. Yet, usual laboratory setups generally are not anechoic and do not have sufficient space for an array of speakers. This severely limits the possibility of doing *free-field* experiments in which the stimuli are directly presented by a speaker in the lab space. For this reason, most of the electrophysiological experiments use artificially induced stimuli presented over headphones. In this case, recording HRTFs on an animal can be used to render spatialized binaural stimuli, hence emulating free-field experiments. One can record from the animal electrophysiologically using stimuli that have exactly the acoustical binaural cues. This technique is called Virtual Acoustic Space (VAS, e.g. in [Mrsic-Flogel et al., 2005, Sterbing et al., 2003a]).

Another important aspect of study of binaural cues is the discussion over the range of ITDs that an animal is likely to "hear" in its environment: the *physiological range* of ITDs. As an example, this argument has had a prominent position in the recent debate opposing tenants of the *peak code* and *slope code* hypotheses. Neurons in the mammalian MSO are sensitivite to the ITD presented in the binaural stimuli. Actually, their rate-ITD curve, measuring the rate of the neuron as a function of the ITD in the stimulus, displays peaks for multiple ITDs. For tenants of the peak code hypothesis,

the meaningful part of the rate-ITD curve is the peaks. When the ITD presented in the stimulus is equal to a neuron's BD, it fires maximally. Hence it is expected to be the most active neuron of the population, signaling that the ITD in the input is its BD. This peak code view has been presented as the textbook view of the binaural system. Yet, it has been oftentimes found since then that cells sensitive to low frequencies had BDs well oustide of the *physiological range* of the animal studied (e.g. [McAlpine et al., 2001]). Cells with such large BDs suggested that not only the peak of the rate-ITD curve was relevant to the system: a cell would then never effectively witness its preferred ITD. This lead to the proposal that the relevant measure is the slope of the rate-ITD curve around the midline (zero ITD), which leads to a population code of the azimuth.

Hence, we seek to provide measures of the *physiological range*, in many common laboratory species. Those are a valuable asset in the perspective of the study of the neural coding of sound azimuth through ITD. The purpose of this chapter is to conduct a study of the variation of ITD in many animal species, and using the same tools as those provided in Chapter 2. We measure HRTFs on naturalized animal models, as presented on Fig. 3.1. The species measured are the rat, cat, chinchilla, barn owl, macaque monkey, guinea pig and rabbit.

3.2 Methods

The HRTF measurement setup used to measure the animals was the same as that previously described for humans in Chapter 1. Here we recall only the parts of the measurement setup that are specific to the animal recordings.

3.2.1 Recording stuffed animal

When measuring HRTFs, the subject has to sit as still as possible for the complete period of the recordings. In the human recordings of Chapter 1, this was about one hour and a half. While human subjects can be made to sit still through promises of great retribution, this is merely possible with animals. Recording of HRTFs with awake animals has only been achieved in the rabbit [Kim et al., 2010a], while all the other animal recordings rely on anaesthetized animals (in Tab. 3.1.1). Keeping an ecological stance with an anaesthetized animal can prove to be a tedious process, which clutters the space where the animal is recorded with reflective objects.

Let aside the goal of displaying exotic species to the general public in museums, taxidermy is a serious craft that was initially developed in the context of biological

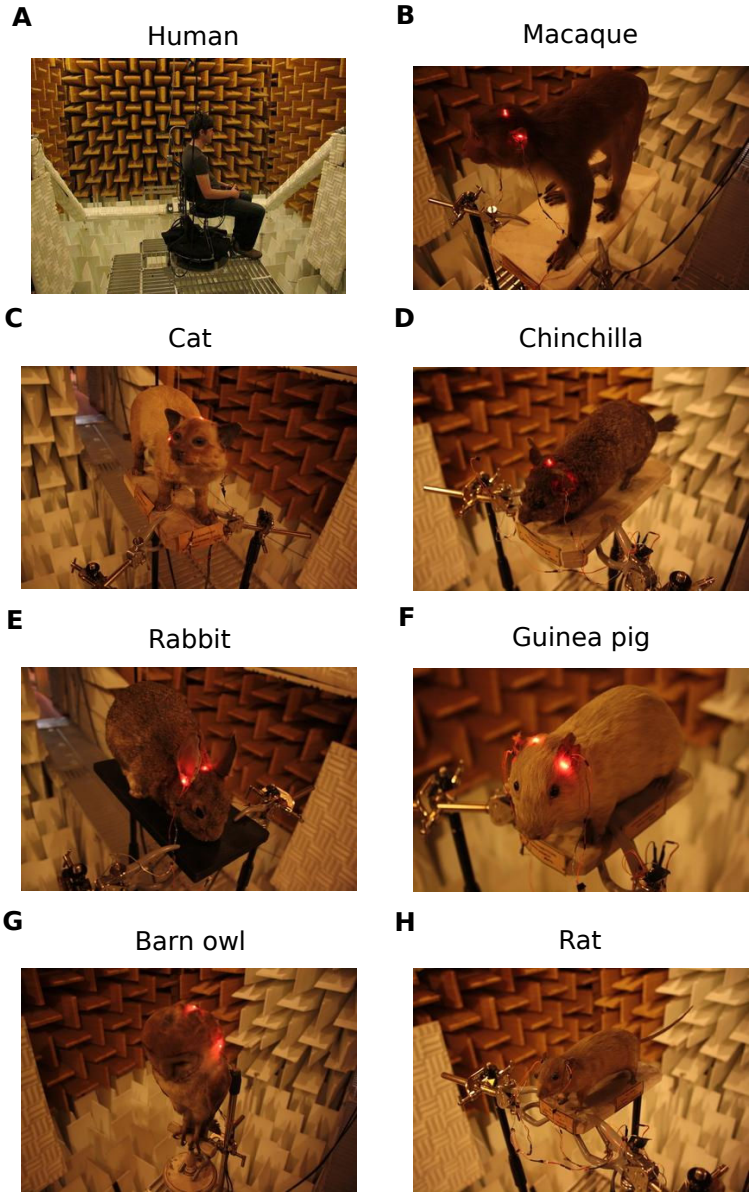

Figure 3.1: Pictures of all the recorded animals in the anechoic chamber. Miniature microphones are placed in the animal's ears. The red dots are the lasers that helped correct positioning of the animal with respect to the crane and speakers. Animals recorded: **A** Human **B** Macaque **C** Cat **D** Chinchilla **E** Rabbit **F** Guinea pig **G** Barn Owl **H** Rat

sciences. Zoologists, still today use naturalized animal specimens to make observations. For this purpose, great care is always taken in positioning the animal in a *natural stance*: the biological validity of the stance always constrains the "artistic" interest of the piece. In France, there is a "Meilleur Ouvrier de France" (MOF, Best French Craftsman) diploma in taxidermy. The MOF is a competition held in France every four years, recognizing outstanding achievement in various crafts domains. As an example, MOF holders in cooking usually have multiple Michelin starred restaurants.

If the ecological validity of the stance of the models is justified, the models are not generally in a neutral stance. It is seldom found an animal with a completely straight, front-facing position, but the poses are nontheless ecological. The animals were picked according to two main criteria. The "neutrality" of the pose: the animal has to be standing with its head facing as straight as possible. The overall quality: the animal had to be sufficiently well conserved, especially with the ears. Cartilages are hard to preserve and it happens that animals have broken ears. The chosen animals all had reasonably conserved pavillons and correct positions, as can be judged from figure Fig. 3.1

All stuffed animals depicted on Fig. 3.1 were borrowed at the *Museum d'Histoire Naturelle de Paris*, which has a collection of about 250 000 taxidermized mammals and birds. Most of the animals were "mounted" (or stuffed) in the course of the 19th century, when the popularity of taxidermy was at its peak . The technique used, based on arsenic soap, ensured longevity, and most of the animals are well preserved. But it prevented us from touching the animals bare hands.

3.2.2 Animal setup

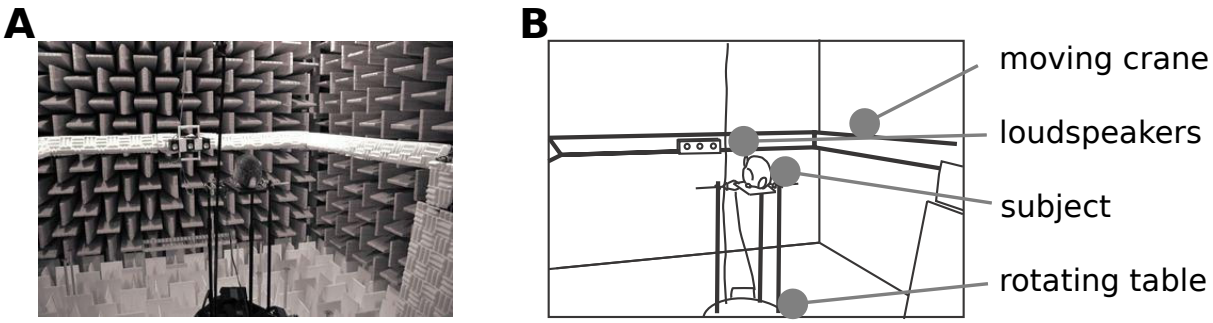

Figure 3.2: Pictures of all the recorded animals in the anechoic chamber. **A** A picture of the setup with the rabbit positioned for recording. **B** Drawing of **A** reporting the main components of the experimental setup.

The figure Fig. 3.2 shows a depiction of the recording setup with the rabbit in position facing the three speakers. The setup used is very close to that presented in Chapter 1

for humans, constrained by the animal model. The animals could not be separated from their wooden plate base. This plate varied in shape and size greatly across the animals Fig. 3.1. Those plate may introduce some reflections, especially for animals in which the plate is big (e.g. our macaque model), which means that low-elevation recordings should be taken with care. Animals were always placed so that the plate faced the 0°direction at the beginning of the experiment.

Animal	Equivalent head radius (cm)	Head pointing azimuth (°)
Human (main author)	10.3 cm	0°
Cat	4.34 cm	-55°
Chinchilla	4.04 cm	-20°
Guinea pig	2.6 cm	20°
Macaque	5.09 cm	-30°
Barn Owl	3.54 cm	5°
Rabbit	3.99 cm	-10°
Rat	1.86 cm	10°

Table 3.2: Equivalent head radii of the measured animal species, and angles used to compensate for the head pointing direction.

Because not all models we used were facing straight (relative to their body or supporting plate), the HRTF recorded when the setup is in position for 0°might not correspond to the position where the speaker is in the midline of the animal's head Fig. 3.2. This is corrected for in all the recordings, by re-centering the data on the position that yields the most *similar* left and right responses on the horizontal plane. Here, we computed the root-mean-square of the $\mathrm{ITD_p}$ over a broad spectrum and taking the minimum over positions. Those compensation angles are reported in Tab. 3.2.

Quantity	Notation	Value
Sampling frequency	f_s	192 kHz
Frequency spacing	Δ_f	11.17 Hz
Number of samples	N	16384
IR duration	T_{IR}	85 ms
Sample duration	T_s	5.2 μs
Recorded positions	N_{coords}	651
Approx. azimuth spacing	$\Delta\theta$	5°
Approx. elevation spacing	$\Delta\phi$	10°

Table 3.3: Summary of the recordings's characteristics

On Tab. 3.2.2 the main characteristics of the recordings that were taken on the animals are recalled. Note that because most of the animals have smaller heads, thus

smaller ITDs, we chose to measure at a greater sampling frequency than for the human recordings of Chapter 1. Furthermore, because the animals are completely still, this also makes the data very reliable in high frequencies.

3.3 Results

3.3.1 ITD across frequencies for animals

The total phase ITD, ITD_p was computed for all recorded animals, for frequencies between 500 Hz and 3 kHz, and is reported on Fig. 3.3. Reported are the two extreme contralateral and ipsilateral positions ($\theta = \pm 90^\circ$ and $\phi = 0^\circ$), and two positions closer to the midline, $\theta = \pm 30^\circ$and $\phi = 0^\circ$. We observe that the highest low frequency and high frequency ITDs in all animals seem to match the predictions of the Kuhn formula for high frequencies Equation 1.3, where the head radius was estimated from the response itself Tab. 3.2. The extent of the LF to HF variation of ITD_p is also qualitatively similar to that predicted by the estimations. This indicates that the qualitative predictions of the spherical model are in general met in those animals. Note that in our method the head radius has been estimated form the HF ITDs, so the only important prediction is the LF ITD.

The dashed line indicates the transition frequency estimated for the animal, that is the frequency at which the spherical model predicts the greatest variation of ITD Chapter 1. We observe that for all animals this transition frequency is always in the range in which the ITD is a dominant cue to localization.

In most animals, it seems like the ITD variation observed is not the same for symmetrical positions around the head (e.g. $\pm 90^\circ$). This emerges from the fact that the animals are in general not in symmetrical positions Fig. 3.1.

Note: it happens that the median sagittal plane, wherein the cues should be close to null because the head is quasi symmetric, had non zero cues.

3.3.2 ITD_g and IDI in animals

We have observed in Fig. 3.3 that the ITD varies with frequency across the spectrum in all animals. As was already done using cat data, we compute the group ITD, ITD_g [Roth et al., 1980b], and the IDI according to Chapter 2. Because of the results of the previous chapters, we expect the group ITDs to be different from the total phase ITD. Hence, it is interesting to measure separately ITD_g and ITD_p from the phase response of the HRIRs. This is done according to the method developed in Chapter 2. The results of

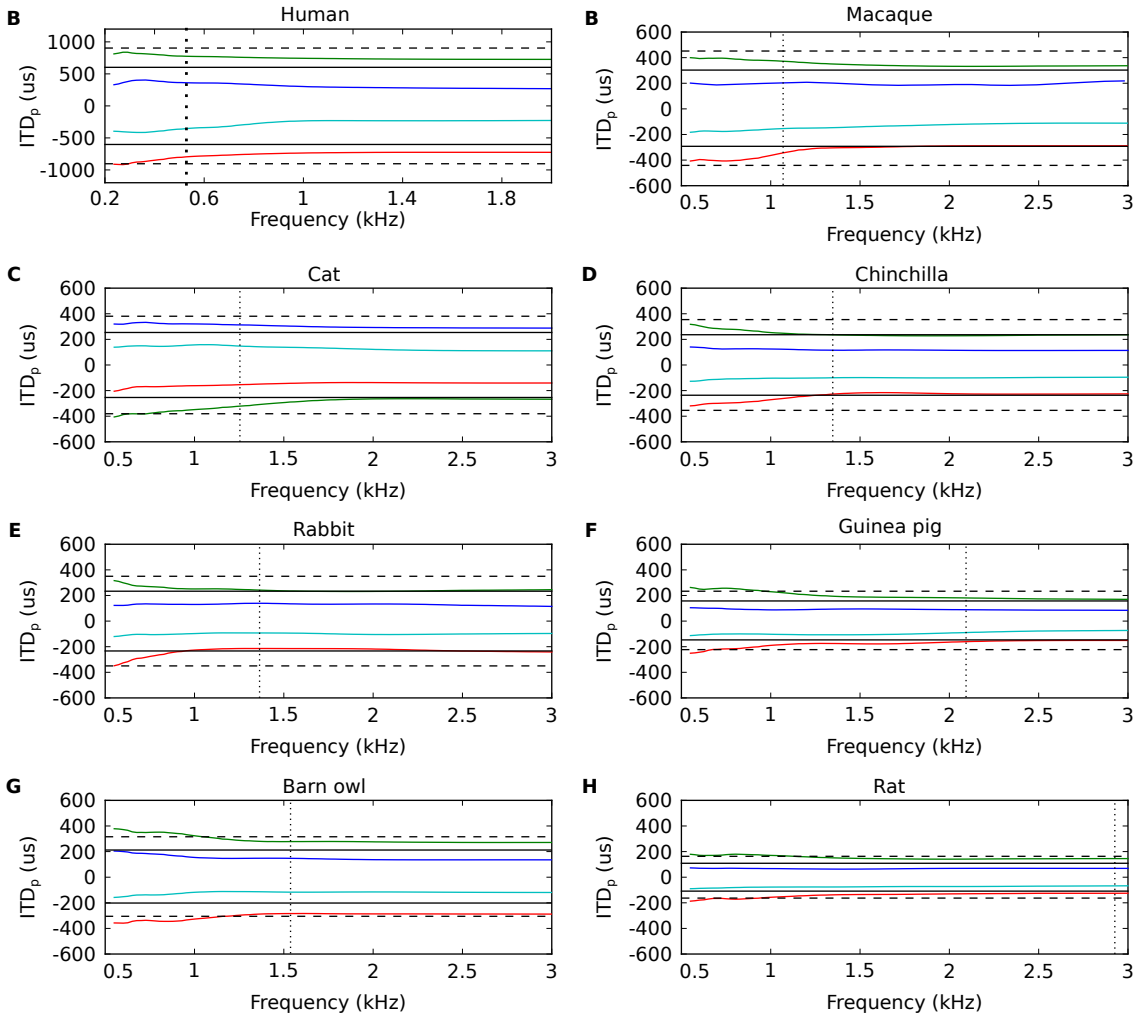

Figure 3.3: ITD_p as a function of frequency in all animal species, plotted in the relevant frequency ranges for each animal. ITDs are reported for five positions on the azimuthal plane. **A** Human (main author) **B** Macaque **C** Cat **D** Chinchilla **E** Rabbit **F** Guinea pig **G** Barn Owl **H** Rat

this analysis are plotted on Fig. 3.4. As expected, the IDI is generally nonzero at all positions.

The ITD_g functions for two positions are in general well separated. This means that at any given frequency point, the ITD_g seems to be an increasing function of deviation from the midline. This property does not seem to hold easily when considering the IDI. Indeed, the different curves reported on Fig. 3.4 cross and overlap many times, and at some frequency points they are indistinguishable. This could be seen as an argument to discard IDI altogether, because a clear dependency (say, cosine) between the measure and the azimuth cannot be found. When considering the plots in the (IDI, ITD_g) plane (Fig. 3.4, the picture becomes clearer. Patches of points aligned in a downward trend indicate the position of the sound source. In this plane there seems to be a clear dependency on the azimuth of the sound source, and positions are well segregated.

3.4 Discussion

3.4.1 Physiological ranges of ITD in different species

Here focus on the physiological range of ITD.. We adopt a proper definition of the physiological range, which corresponds to the absolute probability of observing an ITD given the frequency. Because we have evenly spaced positions in the HRTF data, this is actually assuming that the sound sources are distributed uniformly in terms of solid angle. Rather counterintuitively, this assumptions in turn leads to the fact that ITDs are not distributed uniformly for a given frequency. As an example, the ITD of a two-mics system (as in Chapter 1) is a sine function of the azimuths. If one measures the distribution of ITDs when sources are uniformly distributed in azimuth, then it will have peaks at almost-extreme ITDs. This is because the sine function dampens around $\pi/2$, which are the extremal value 90°and -90°. Those ITDs will be overrepresented, as opposed to the more central ITDs.

Physiological ranges of ITD are reported as a single value in different papers. On Fig. 3.5, we observe the ITD physiological ranges as 2D plots. In all species, the physiological range spans a greater range in low frequencies than high frequencies, consistent with the spherical model hypothesis. Interestingly, in small mammals, the low frequency ITD and high frequency ITDs are in general bigger than the ones provided in the literature (e.g. on a spherical model, Chapter 1), even though the difference might be accounted for by interindividual differences.

In general the arguments based on the physiological range were only interested in

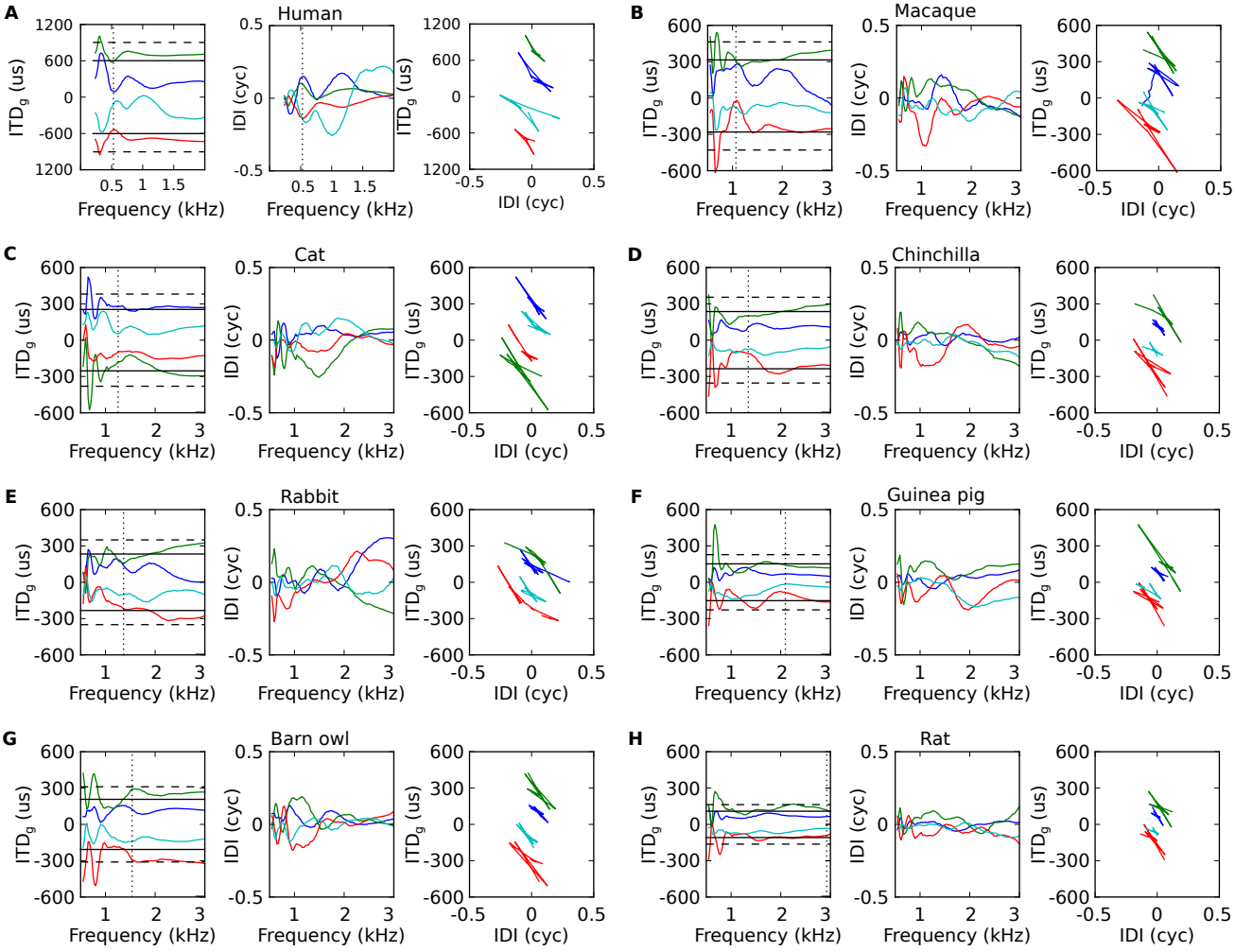

Figure 3.4: For each animal, two leftmost panels are $\mathrm{ITD_p}$ and IDI as a function of frequency. The third panel corresponds to the curve $(\mathrm{IDI}(f), \mathrm{ITD_p}(f))$ for all positive frequencies. All ITDs are reported for five positions on the azimuthal plane, the panel of each animal is the same as in Fig. 3.3.

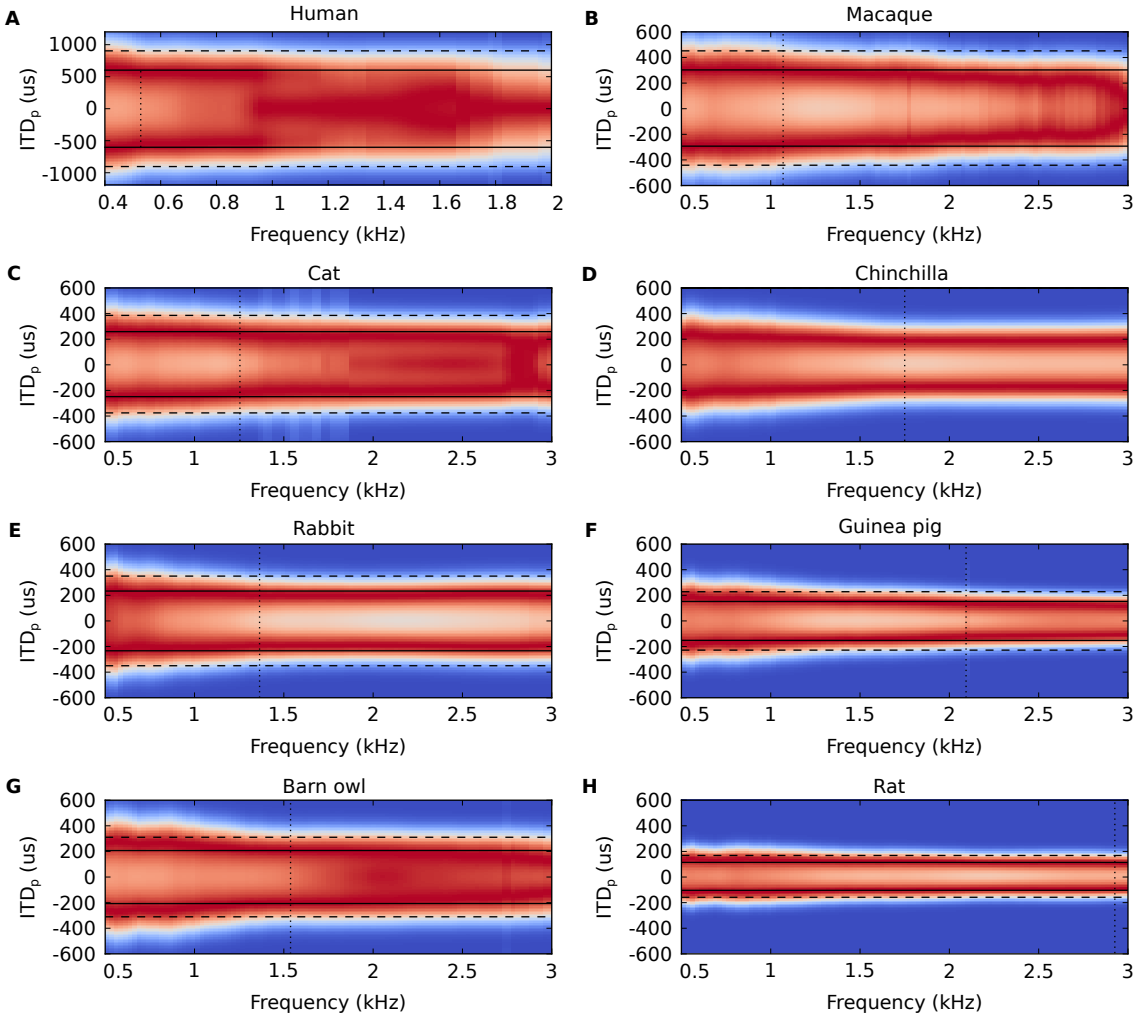

Figure 3.5: Physiolgoical ranges of $ITD_p$for all the stuffed animals presented here. Each panel is a color plot of the distribution of ITD_pon the azimuthal plane, as a function of frequency and ITD_p. Only the positions near the horizontal plane (within 20°of 0°elevation) are reported. Color code: blue is improbable, red is highly probable and white is intermediate.

the *maximal possible ITD*, which is a single figure. From our data, we observe that the physiological range of ITDs is not a uniform distribution, and that it also varies with the frequency in all measures. Hence, a description of the possible acoustical cues in an animal species cannot be obtained with a single figure. Indeed, for low-frequencies the physiological range is generally bigger in than in higher frequencies (Fig. 3.5

From these observations, we can draw a few conclusions, namely on the intrinsic limitation of the ability to localize sound sources. If we assume that the subject has a constant ability to discriminate *positions*, that is it is just as good in telling apart 0 and 5°than 80 and 85 °. ITDs of extremal positions (say, 80 and 85 °) are closer together than ITDs around the midline. Thusly, under this hypothesis, the ability to discriminate *ITD* whould be better for bigger ITDs. This is *not* what is obseved, first the ability to discriminate positions in free field is better for sounds around the midline, and second, it is generally observed that JNDs in ITD (i.e. an index of the precision in ITD discrimination) are bigger for bigger ITDs. Therefore, the data that we have gathered suggests that the ability to discriminate the position of a sound source is rather limited by discrimination abilites in ITD, and not the contrary. A constant ITD discrimination ability is consistent with the data indicating that humans are worse at telling appart sound positions for excentric positions

3..2 ITD_g and IDI physiological ranges in each species

We computed the physiological range for all species, in a manner similar as for Fig. 3.5

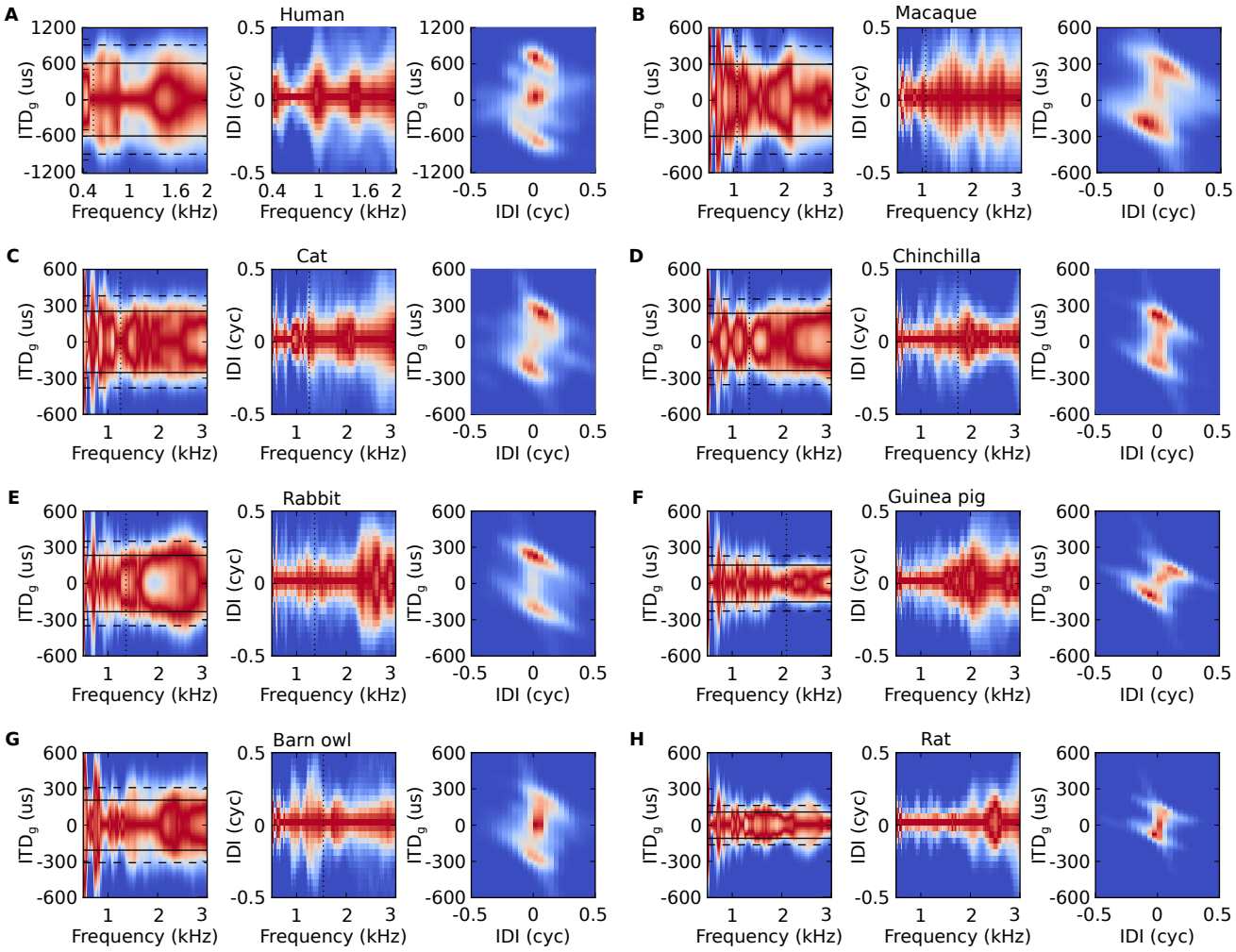

Figure 3.6: Physiological ranges of ITD_g for all the measured species. Data same as in Fig. 3.5

Chapter 4

Reconstructing HRTF from pictures: case of the cat

Contents

In this section we present a quick and efficient method to measure low frequency HRTFs in animals using pictures and numerical simulation. We apply this method to generate HRTFs for a cat model. Finally, we validate the appraoch by comparing its results with acoustical recordings performed on a stuffed cat.

This paper, done in collaboration with Marc Rébillat was submitted to JASA.

4.1 Abstract

Reliable animal head-related transfer function (HRTF) estimation procedures are needed for several practical applications.

For example, t oinvestigate the neuronal mechanisms of sound localization using virtual acoustic spaces, or to have a quantitative description of the different localization cues available to a given animal species.

In this context, we present and validate a method to estimate an animal's HRTF from photographs by taking into account as much morphological detail as possible. The first step of the method consists in building a 3D-model of the animal from pictures taken with a standard camera. The HRTF are then estimated by means of a rapid boundary element method implementation. The method is validated on a stuffed cat by comparing binaural and monaural localization cues extracted from estimated and measured HRTF. The main advantage of this method is that it can be used for any animal and necessitates no time-consuming acoustical measurements and no expensive equipment.

4.2 Introduction

Animals can localize a sound source in space by analyzing the sound signals arriving at their two ears. The acoustical transformation occurring between a point source and a receiving ear is usually termed *"Head Related Transfer Function"* (HRTF) [Wightman and Kistler, 2005b]. HRTF convey all the acoustical cues available for the animal to localize a sound source in space: *binaural* cues, such as interaural-time differences (ITDs) and interaural level differences (ILDs), and *monaural* cues, such as spectral cues [Blauert, 1997]. These acoustical cues however greatly depend on the morphology of the animal under study [Xu and Middlebrooks, 2000, Schnupp et al., 2003, Tollin and Koka, 2009d, Tollin and Koka, 2009a, Jones et al., 2011]. Reliably estimating the HRTF of a given animal is a challenging task that is necessary for various applications.

For example, to investigate the neuronal mechanisms of sound localization with physiological and behavioral experiments, HRTF can be used to generate controlled binaural stimuli presented through headphones using a virtual auditory environment [Jacobson et al., 2001]. Ideally, the use of HRTF that are not specific to the animal under study should be avoided as it alters the virtual-spatial sensitivities of neurons in the central nervous system [Mrsic-Flogel et al., 2001, Schnupp et al., 2001, Sterbing et al., 2003b]. Reliable HRTF are also needed to quantitatively estimate the localization cues available for different animal species. For this purpose, animal HRTF

are commonly measured experimentally by placing miniature microphones at the ear canal entrances and by measuring their responses to controlled sounds played through a movable loudspeaker [Mehrgardt and Mellert, 1977, Roth et al., 1980a, Xu and Middlebrooks, 2000, Spezio et al., 2000a, Maki and Furukawa, 2005b, Koka et al., 2008b, Koka et al., 2011b, Tollin and Koka, 2009d, Tollin and Koka, 2009a, Kim et al., 2010b, Jones et al., 2011]. Such procedure nevertheless require dedicated hardware and facilities as well as a substantial amount of time. In particular, sound localization in the horizontal plane relies mainly on low frequency ITDs, but such measurements are often limited in the low frequency range by the quality of the measurement setup (which is mostly degrader by loudspeaker response) and of the measurement room (anechoic or not) being used. Alternative HRTF estimation methods are thus needed to address those issues and to obtain reliable localization cues in the low frequency range.

To avoid HRTF measurements for each animal, HRTF of a given animal can be estimated using HRTF measured on another animal of the same species on the basis of some morphological parameters. For cats [Xu and Middlebrooks, 2000], ferrets [Schnupp et al., 2003], and gerbils [Maki and Furukawa, 2005c], HRTF amplitudes can be quite accurately predicted by a scaling operation on the log-frequency axis and by a rotation of the HRTF around the source coordinate sphere. The optimal scaling factor (OSF) and the optimal coordinate rotation (OCR) can furthermore be predicted on the basis of morphological parameters. However, this approach requires that for each species under study one reference measurement and the relations between morphological parameters and the OSF and OCR are available.

Simplified geometrical models such as spherical models [Brown and Duda, 1998a, Duda et al., 1999b, Algazi et al., 2001b], head and torso models [Algazi et al., 2002, Pernaux, 2003, Zotkin et al., 2003], or dummy head models [Dellepiane et al., 2008] have also been proposed to estimate human listener's HRTF. The morphological parameters needed to feed the geometrical model can be easily estimated, for example from photographs [Pernaux, 2003, Zotkin et al., 2003, Dellepiane et al., 2008]. However, the resulting estimated HRTF are only approximate and valid in the lower frequency range, where fine morphological details not taken into account in the simplified geometrical model have negligible influence. This approach is furthermore limited to humans as it relies on dummy models which are generally not available for animals.

Finally, animal HRTF can be estimated using a full 3D-model of the animal's body obtained for example from a scanning laser [Quaranta, 2003, Muller, 2004, Grace et al., 2008, De Mey et al., 2008]. As all the fine morphological details are taken into account in the full 3D-model, HRTF estimated using this method are valid over a larger frequency range than when estimated using simplified geometrical models. However, this

approach relies on expensive equipment that is not always available in practice.

In this paper, we present and validate a general method to estimate HRTF from raw photographs by taking into account as much morphological detail of the animal as possible. The first step of the method consists in building a 3D-model of the animal from photographs taken with a standard camera [Lafarge et al., 2010]. Based on this 3D-model, the HRTF are then computed by means of a rapid boundary element method (BEM) implementation [Otani and Ise, 2006]. The main advantages of this method are that it is generic, fast, and necessitates no direct acoustical measurements as well as no expensive equipment.

The proposed method is described in Sec. 4.3 and used to estimate the HRTF of a stuffed cat, which are also experimentally measured and compared to a simple spherical model. The method is then validated in Sec. 4.4 by comparing the binaural and monaural localization cues provided by the estimated, spherical, and measured HRTF. The limitations and new possibilities offered by the method are finally discussed in Sec. 4.5.

4.3 Methods

The HRTF estimation pipeline presented here is based on two steps and is illustrated on the stuffed cat (a female *felis domestica*) shown in Fig. 4.1(a). The HRTF of the stuffed cat are also measured on the actual stuffed cat and compared to a simple spherical head model.

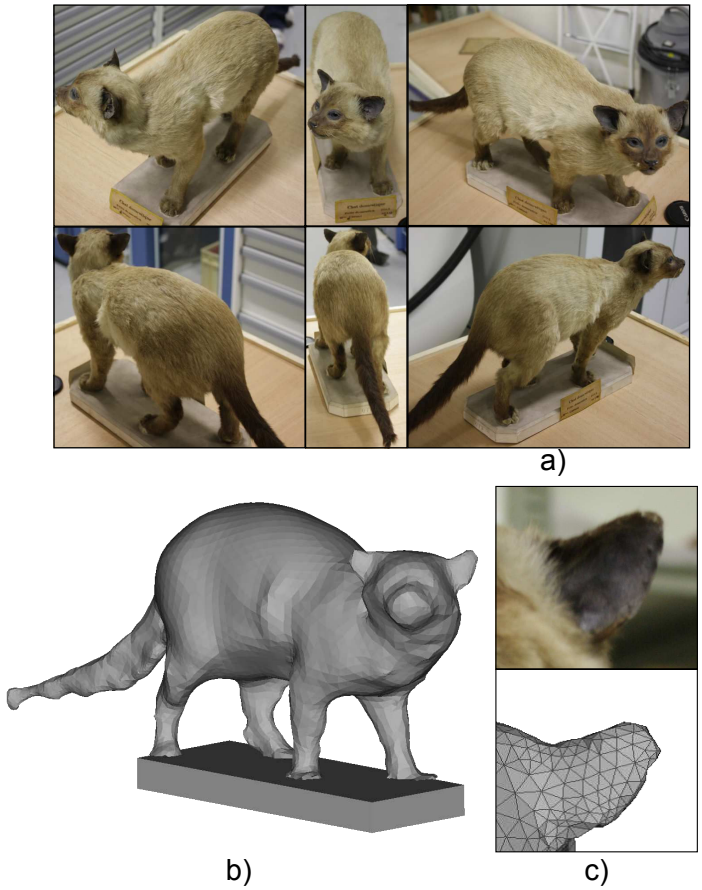

Figure 4.1: Stuffed cat used to validate the HRTF computation procedure. (Color online) (a) Examples of photographs used for the 3D-model estimation. (b) Overview of the estimated 3D-model of the cat. (c) Detail of the left ear of the cat: real cat (top) and 3D-model (bottom).

4.3.1 HRTF estimation procedure

3D head and body model reconstruction

In order to build a 3D-model of a given object, some input data carrying geometrical information regarding the object to be modeled is needed. A classical way to acquire such data consists in using laser scanning equipments [Bernardini and Rushmeier, 2002]. However, such equipment is particularly expensive and hard to use in practice. An alternative method to build 3D-models consists in using a set of raw photographs of the object as input data [Seitz et al., 2006, Strecha et al., 2008]. This kind of method requires only a standard camera and is able to provide 3D-models that are qualitatively in good agreement with the ones obtained through laser scanners [Seitz et al., 2006, Strecha et al., 2008].

For its practical ease of use, a method based on photogrammetry [Leberl and Thur-

good, 2004, Mayer, 2008] and using raw photographs as input data has been retained here [Lafarge et al., 2010]. This method models an arbitrary 3D-object as a combination of meshes and of geometrical primitives. On the basis of raw photographs, a Jump-Diffusion process [Grenander and Miller, 1994] is designed to sample these two types of elements simultaneously. The 3D-models reconstructed by this method have been shown to be qualitatively comparable to models acquired through the use of laser scanners [Lafarge et al., 2010]. Only the camera focal distance is required for the method to output a scaled 3D-model.

We applied this 3D-model estimation procedure to the cat shown in Fig. 4.1(a). A 3D-model of the stuffed animal was estimated based on 68 photographs taken with a standard 10 Mpixels camera under normal lighting conditions, see Fig. 4.1(b). This model is made up of 11526 triangles with sides of $\simeq$ 1 cm, which is larger than the resolution of scanning lasers ($\simeq$ 1 mm).

As is shown in Fig. 4.1(c) the fine geometrical details, such as the exact concave shape of the cat pinnae, are not well captured by this method. It should also be noted that the cat fur has been modeled here as a continuous surface. However, all the major body part shapes, *i.e.* the head, body, legs and tail are accurately modeled. The 3D-model obtained here is thus more precise than a simple geometrical model (spherical model or head and torso model) but not as precise as would be obtained using a precise laser scanner.

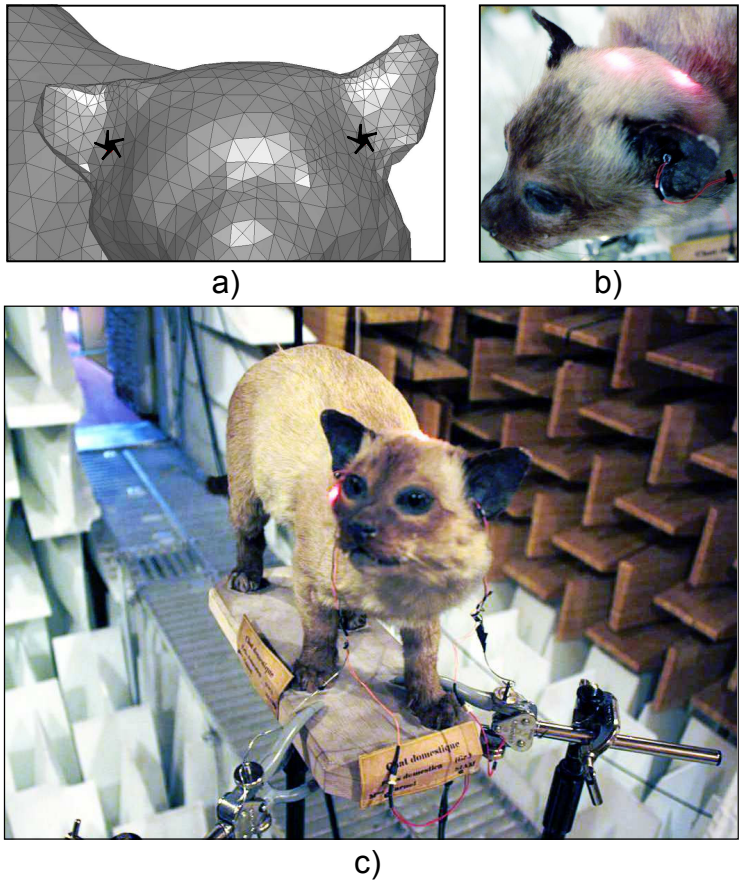

Figure 4.2: Overview of the measurement and BEM procedures. (Color online) (a) Red dots indicate the position of the acoustical sources in the BEM calculations. HRTF measurement setup for the cat: (b) zoom on the positioning laser and ear canal microphone and (c) overview.

The HRTF are then computed from this 3D-model using the fast HRTF calculation algorithm proposed by [Otani and Ise, 2006] based on the boundary element method (BEM). In this BEM formulation, the reciprocity theorem is incorporated into the computational process in order to shorten the processing time. Moreover, all the factors independent of the source position are precomputed in advance. Using this algorithm, the HRTF for any source position are obtained in a few seconds with a standard PC, once the pre-computation process has been achieved on the estimated 3D-model.

It is generally assumed that BEM performs well up to the frequency for which there are 6 to 10 triangular mesh elements per wavelength [Katz, 1998, Kahana, 2000]. Given that the nodes are spaced by $\simeq$ 1 cm (see Sec. 4.3.1) and according to this criterion, the upper frequency limit below which the HRTF are accurately computed is expected to be between 3 and 5 kHz.

Regarding boundary conditions, the acoustical properties of the cat fur are not known. It is however known that absorption coefficients are larger for a haired animal than for a hairless one [Ackerman et al., 1957, Katz, 2000]. More specifically, some studies focused on the effects of the hair and clothes on human HRTF: [Burkhard and Sachs, 1975] found that the ear entrance sound pressure was relatively insensitive to the head impedance. [Kuhn, 1977] has shown that ITD and ILD differences between a bare and a clothed torso are relatively small below 2 kHz. [Katz, 2001, Treeby et al., 2007] studied the effect of hair on HRTF and concluded that they do not play an important role below 3 kHz. [Katz, 2001] furthermore emphasized the fact that modeling hair or fur as a normal reactive impedance is only a rough approximation of the physical reality. Thus, since 1) we are interested mainly in the frequency range below 3 kHz due to the relatively coarse meshing of our model, 2) no data about the impedance of the cat fur were available, and 3) the BEM implementation of impedance still is not perfect, we chose to apply rigid boundary conditions over the whole cat surface.

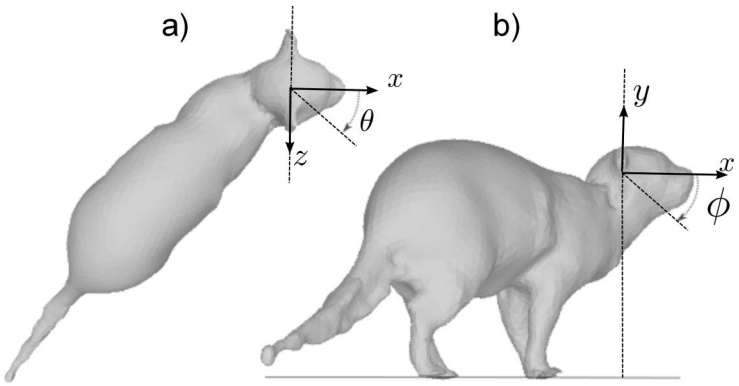

Figure 4.3: Coordinate system definition: a) top view ($\phi = 0°$) with azimuth θ and b) side view ($\theta = 0°$) with elevation ϕ. See [Warusfel, 2002] for more details.

We computed the HRTF for the cat from the previously obtained 3D-model, see Fig. 4.1(b), with rigid boundary conditions. Acoustical sources were placed sequentially at two points situated a few millimeters in front of the 3D-reconstruction of the cat ears, as shown by the red dots in Fig. 4.2(a). The HRTF were computed for 651 positions at a distance of 1.95 m, with a frequency resolution of 43 Hz and a sampling frequency of 44.1 kHz. The spatial resolution of the computed HRTF is 5° in azimuth, from $-175°$ to 180°, and 15° in elevation, from $-40°$ to 90° following the LISTEN coordinates system [Warusfel, 2002] shown in Fig. 4.3. Pre-computation took about 7 hours on a Red Hat Linux, Xeon 3.33 GHz, 8 cores, 48 Gbyte RAM workstation.

4.3.2 HRTF measurements for the cat

In order to validate the HRTF estimation pipeline described above, the actual HRTF of the cat have also been measured experimentally. The HRTF of the cat shown in Fig. 4.1(a) were measured by means of the IRCAM HRTF measurement system [Warusfel, 2002]. Measurements were done in an anechoic chamber covered with glass wool wedges, see Fig. 4.2(c). HRTF were obtained through the blocked ear canal measurement method, which has been shown to be sufficient to acquire directional information in humans [Hammershoi and Moller, 1996, Wightman and Kistler, 2005b]. A pair of miniature microphones have been placed at the entrance of the occluded ear canal of the stuffed animal as shown in Fig. 4.2(b). The cat head was positioned so that the center of the interaural axis was located at the center of the sphere and the horizontal plane was parallel to the support plane. This was accomplished using two low-powered lasers to align the head correctly. The sound source was a speaker placed on a rotating crane. HRTF were measured at the same 651 positions at which computations were performed, with a frequency resolution of 11.7 Hz and a sampling frequency of 192 kHz.

4.3.3 HRTF estimation using a spherical head model

In order to emphasize the benefits offered by the estimation procedure proposed here in comparison with procedures based on simple geometrical models, the HRTF of the cat have also been estimated using a simple spherical model. The implementation of the spherical model proposed by [Duda and Martens, 1998] has been used here. The diameter of the sphere modeling the cat's head has been chosen to be equal to the interaural cat distance, *i.e.* 7.3 cm (see Fig. 4.2). Using this spherical model, the HRTF were estimated for a distance of 1.95 m at the same 651 positions at which computations and measurements were performed, with a frequency resolution of 10.8 Hz and a sampling frequency of 44.1 kHz.

4.4 Results: Comparison of localization cues

In this section the localization cues provided by the different HRTF sets obtained in the previous section are compared. HRTF estimated with the procedure proposed in the paper will be referred to as *estimated* HRTF (see Sec. 4.3.1), experimentally measured HRTF as *experimental* HRTF (see Sec. 4.3.2), and HRTF obtained using the spherical model as *spherical* HRTF (see Sec. 4.3.3).

4.4.1 Definitions

Before comparing the different HRTF sets, the monaural and binaural localization cues that will be used in the following are defined.

The monaural spectral cues for the left and right ears are defined as the left and right directional transfer functions (DTF) as done by [Kistler and Wightman, 1992a]. For each considered HRTF set, the mean across the 651 positions of the HRTF log-magnitude is computed. These mean functions include the direction-independent spectral features shared by all the HRTF of a given set. To remove these features, the appropriate mean function is then subtracted from the log-magnitude of each HRTF of a given HRTF set. With means removed, the resulting 651 log-magnitude functions, denoted $\mathrm{DTF}_l(f)$ and $\mathrm{DTF}_r(f)$ in the following, represent direction-dependent spectral effects.

Let us now consider a pair of HRTF, $H_l(f)$ and $H_r(f)$, corresponding to a given source position. The interaural level differences, $\mathrm{ILD}(f)$ (in dB), and interaural phase differences, $\mathrm{IPD}(f)$ (in radians), are defined according to Eqs. (4.1) and (4.2), where $\angle(.)$ denotes the *unwrapped-phase* operator, *i.e.* the operator that changes absolute jumps greater than or equal to π to their 2π complement:

$$\mathrm{ILD}(f) = 20\log_{10}\left[\left|\frac{H_l(f)}{H_r(f)}\right|\right] \tag{4.1}$$

$$\mathrm{IPD}(f) = \angle\left[\frac{H_l(f)}{H_r(f)}\right] \tag{4.2}$$

Two binaural time localization cues are defined from the interaural phase differences by Eqs. (4.3) and (4.4) following [Roth et al., 1980a]. $\mathrm{ITD}_p(f)$ is the interaural phase delay and $\mathrm{ITD}_g(f)$ is the interaural group delay. They correspond to the interaural delay of the temporal fine structure and of the envelope, respectively. These two quantities generally differ and depend on frequency because of sound diffraction by the head and body [Kuhn, 1977, Roth et al., 1980a].

$$\mathrm{ITD}_p(f) = \frac{\mathrm{IPD}(f)}{2\pi f} \tag{4.3}$$

$$\mathrm{ITD}_g(f) = \frac{1}{2\pi}\frac{d\mathrm{IPD}(f)}{df} \tag{4.4}$$

To quantify this difference, we additionally define in Eq. (4.5) a new index, the *Interaural Diffraction Index* $\mathrm{IDI}(f)$ (in radians). This cue denotes the phase-lag induced by diffraction effects between the envelope and the fine structure of the incoming

sound. For example, an acoustically transparent head (*i.e.* a case without any diffraction effects), leads to equal and frequency-independent inter-aural group and phase delays, and thus would give a zero IDI at all considered frequencies.

$$\mathrm{IDI}(f) = 2\pi f \left[\mathrm{ITD}_P(f) - \mathrm{ITD}_G(f)\right] \tag{4.5}$$

Summing up all the above-defined cues, there are thus 6 frequency-dependent localization cues (respectively $\mathrm{ITD_p}(f)$, $\mathrm{ITD_g}(f)$, $\mathrm{IDI}(f)$, $\mathrm{DTF_l}(f)$, $\mathrm{DTF_r}(f)$ and $\mathrm{ILD}(f)$) to be compared for the 3 HRTF sets obtained in Sec. 4.3. All cues are smoothed in the frequency domain with a third octave sliding window, narrower than the cat auditory filters [Mc Laughlin et al., 2007].

4.4.2 Qualitative comparison

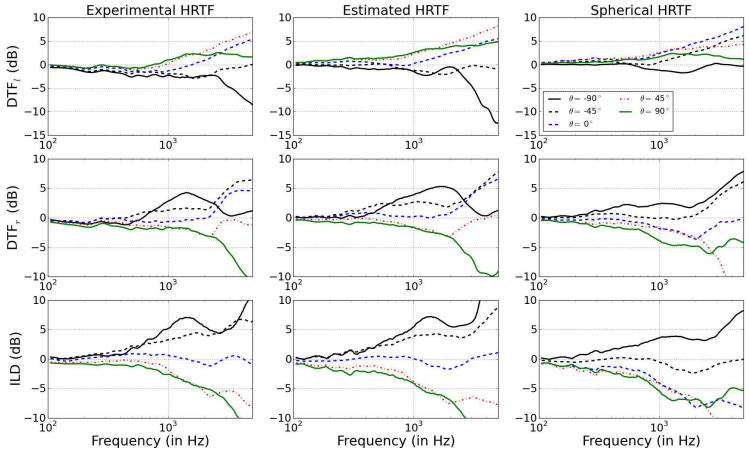

Figure 4.4: Comparison of amplitude-based localization cues $\mathrm{DTF_l}(f)$, $\mathrm{DTF_r}(f)$, and $\mathrm{ILD}(f)$ for the 3 HRTF sets for 5 positions in the frontal azimuth plane and between 100 Hz and 5 kHz. (Color online)

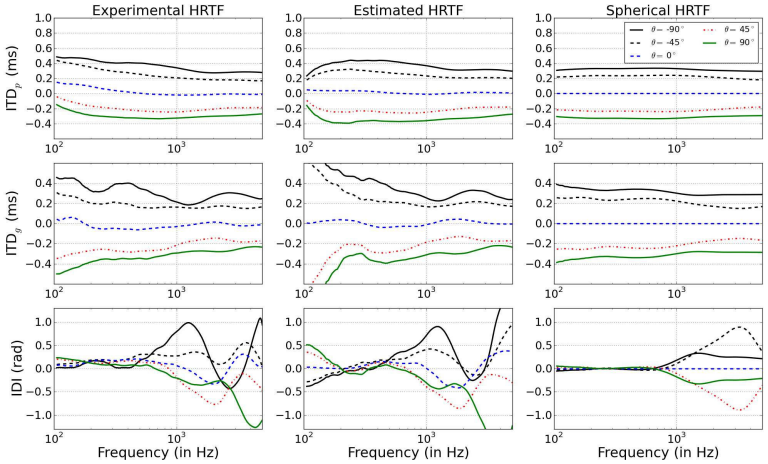

Figure 4.5: Comparison of time-based localization cues $\mathrm{ITD_p}(f)$, $\mathrm{ITD_g}(f)$ for the 3 HRTF sets for 5 positions in the frontal azimuth plane and between 100 Hz and 5 kHz. (Color online)

In Fig. 4.4, amplitude-based loaclization cues $\mathrm{DTF_l}(f)$, $\mathrm{DTF_r}(f)$, and $\mathrm{ILD}(f)$, computed according to the procedure highlighted in Sec. 4.4.1 and to Eq. (4.1) are plotted for the *experimental*, *estimated* and *spherical* HRTF sets for 5 azimuths in the frontal horizontal plane. In Fig. 4.5, time-based localization cues $\mathrm{ITD_p}(f)$, $\mathrm{ITD_g}(f)$, and $\mathrm{IDI}(f)$, computed according to Eqs. (4.3), (4.4) and (4.5), are plotted for the 3 HRTF sets and for the same 5 azimuths.

It can be seen in Fig. 4.4 that there is very good qualitative agreement between the amplitude-based localization cues obtained through the proposed method and those experimentally measured. A much less accurate agreement is observed between those obtained using the simple spherical model and the experimental ones. It is also particularly striking to see that the proposed method is able to render fine changes of the various localization cues with frequency and that the simple spherical model cannot capture such details (see for example the curves for $\theta = -140°$). In Fig. 4.5, it can furthermore be observed that the same comments generally hold for time-based localization cues above 200 Hz. There is thus globally an excellent qualitative agreement between the *estimated* and *experimental* HRTF sets for the 5 positions tested here.

4.4.3 Quantitative comparison

In order to perform a global comparison between the *estimated* and *experimental* HRTF sets over the whole sphere, the differences observed between the monaural and binaural localization cues over the 651 positions have been computed between 100 Hz and 5 kHz. Fig. 4.6 shows the histograms of these differences and Tab. 4.1 provides their

mean and associated standard deviation. In the following, we denote the differences associated with each cue as ΔDTF_l, ΔDTF_r, ΔILD, ΔITD_p, ΔITD_g and ΔIDI.

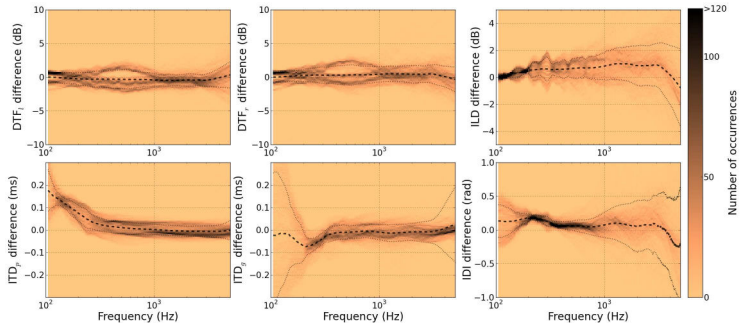

Figure 4.6: Histograms of the differences in localization cues between the *experimental* and *estimated* HRTF sets over the 651 positions and between 100 Hz and 5 kHz. For each localization cue, the dashed line (- - -) stands for the mean difference and the dotted line (·····) depicts two lines lying one standard deviation apart from the mean difference line. (Color online)

Cue	Mean	STD
ΔDTF_l	−0.42 dB	0.95 dB
ΔDTF_r	0.42 dB	1.24 dB
ΔILD	0.84 dB	1.21 dB
ΔITD_p	1.2 μs	23 μs
ΔITD_g	10.1 μs	38 μs
ΔIDI	0.07 rad	0.27 rad

Table 4.1: Mean and standard deviation (STD) of the differences between the different localization cues over the 651 positions between 100 Hz and 5 kHz.

It can be observed that for amplitude-based cues, the differences between the two data sets remains very low between 100 Hz and 3 kHz highlighting the fact that there is quantitative agreement between the amplitude based cues derived from the two HRTF sets in that frequency range. Regarding ΔILD only, it can be observed that the standard deviation of these differences has a global tendency to increase with the frequency. Focusing on ΔDTF_l and ΔDTF_r, it can be seen that the differences are very small between 100 Hz and 5 kHz, with a slight increase around 700 Hz. For time-based cues, the same picture can be drawn highlighting again the fact that HRTF are correctly estimated between 300 Hz and 3 kHz. ΔITD_p exhibits a positive bias below 300 Hz and remains very low above. Standard deviations associated with ΔITD_g and ΔIDI also increase below 300 Hz and above 3 kHz.

In summary, there is thus global quantitative agreement between 300 Hz and 3 kHz between the *estimated* and *experimental* HRTF sets for the 651 positions tested around the whole sphere.

4.5 Discussion

4.5.1 Validity range of the estimation procedure

Results from behavioral experiments in cats indicate that the just-noticeable difference is $\simeq 20$ μs for $\mathrm{ITD_p}$ and $\simeq 1$ dB for ILD [Wakeford and Robinson, 1974]. This is the same order of magnitude as the differences obtained between *experimental* and *estimated* HRTF sets between 300 Hz and 3 kHz (see Sec. 4.4.3). We can thus conclude that for the presented cat example, the HRTF estimation procedure is accurate up to behavioral precision in that frequency range.

For frequencies higher than 3 kHz, it can be seen in Fig. 4.6 that the differences between the *estimated* and *experimental* HRTF suddenly increases for all the localization cues, except $\mathrm{ITD_p}(f)$. As measurements are thought to be reliable in that frequency range, this thus means that the estimation procedure produces unreliable results here. This upper frequency limit is nevertheless in good agreement with the upper frequency limit related to the BEM procedure (see Sec. 4.3.1). We conclude that the 3 kHz upper bound obtained here is a direct consequence of the limited mesh resolution of the 3D-model used here. A solution to go beyond this upper frequency limit could thus be to increase the 3D-model mesh resolution by refining the mesh using smaller triangles or by taking more photographs. Nevertheless, as the influence of fur is supposed to increase with the frequency, the upper frequency limit may not be extended broadly using only mesh refinement.

For frequencies lower than 300 Hz, it can be seen in Fig. 4.6 that the differences between the *estimated* and *experimental* HRTF also suddenly increase but only for the time-based localization cues. By looking closer at Fig. 4.5, it can be seen that there is a systematic increase of $\mathrm{ITD_p}(f)$ as the frequency decreases for the *experimental* HRTF, which is not observed for the two other HRTF sets. As there is no obvious reason to explain this systematic shift, we conclude that the *experimental* HRTF are poorly impacted by the measurement setup and room reflexions below 300 Hz. By comparing the *estimated* and *spherical model* HRTF in that frequency range, one can see that there is a good agreement above 200 Hz. Below that frequency, $\mathrm{ITD_p}(f)$ seems to tend toward 0 and $\mathrm{ITD_g}(f)$ to diverge for the *estimated* HRTF and not for the *spherical model*. This may be a side effect of the BEM procedure which produces

only one point every 43 Hz in the frequency domain (see Sec. 4.3.1) and thus does not provide enough reliable information in the low-frequency range. A solution to go below this lower frequency limit could thus simply be to increase the frequency resolution associated with the BEM procedure.

Finally, our measurements and estimations were done on a stuffed animal, which may differ from a live animal by its acoustical impedance. Nevertheless, the acoustical impedance differences between the stuffed and live animal are expected to be small in comparison with the acoustical impedance differences between the stuffed animal and a rigid model. As the latter differences are expected to have a very small influence on the binaural cues below 3 kHz (see Sec. 4.3.1), results obtained here for a stuffed cat can thus confidently be extended to a live cat.

In summary, the estimation procedure proposed here produces HRTF that can be used for a live animal and that are accurate up to behavioral precision between 300 Hz and 3 kHz. These frequency bounds can furthermore be easily extended as they are directly related to technical implementation details and not to the method itself.

4.5.2 Case study: effects of posture on the cat ITDs

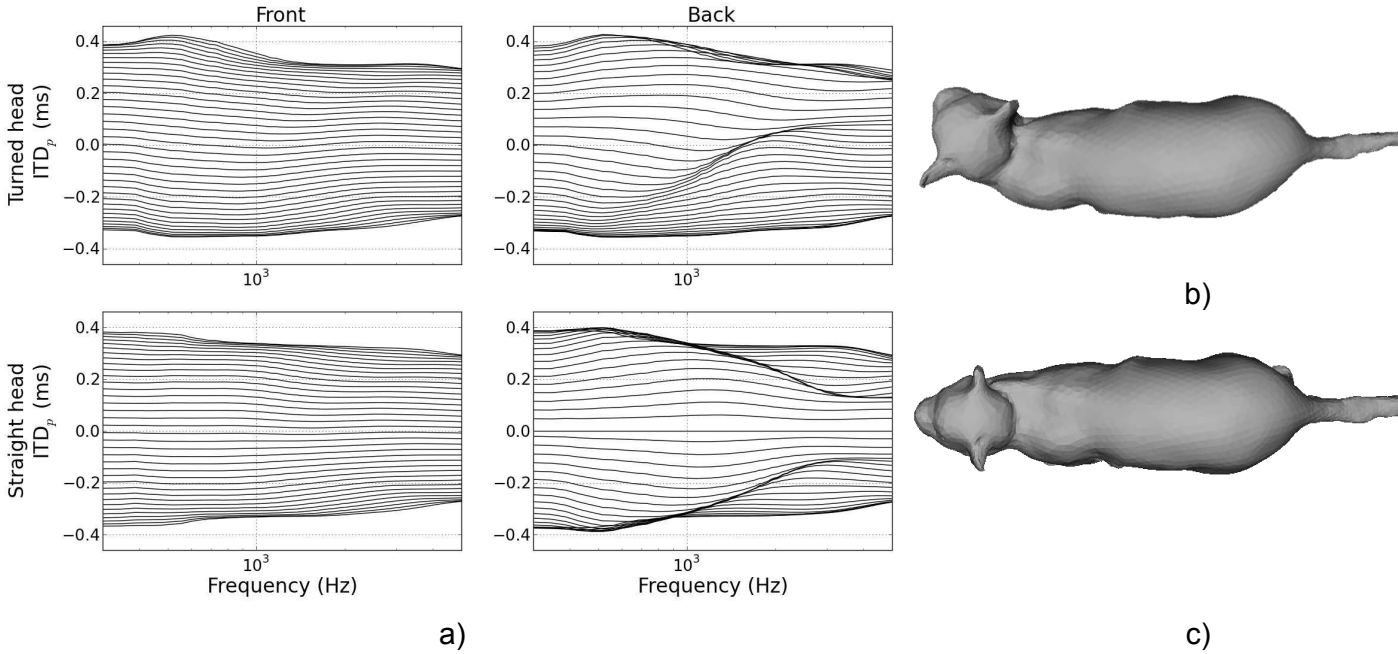

Figure 4.7: Influence of the posture on $\mathrm{ITD}_p(f)$ in the horizontal plane. (a) $\mathrm{ITD}_p(f)$ when the cat's head is straight or turned by 50°. Each line stands for $\mathrm{ITD}_p(f)$ at a given azimuth. The front horizontal plane corresponds to azimuth varying between −90° and 90° by steps of 5°. The back horizontal plane corresponds to azimuth varying between 90° and 180° by steps of 5°. (b) 3D-model of the cat with the turned head. (c) Modified 3D-model of the cat with the head straight.

To illustrate the potential applications offered by this method, $\mathrm{ITD}_p(f)$ is plotted in Fig. 4.7 for the cat model with its head turned by 50°, which was the original posture, or modified to be straight. The front horizontal plane corresponds to θ varying between −90° and 90° by steps of 5°. The back horizontal plane corresponds to θ varying between 90° and 180° by steps of 5°. This figure highlights first the fact that $\mathrm{ITD}_p(f)$ varies in a complex, but organized, manner depending both on frequency and head position. By comparing front and back curves, this figure also emphasizes the fact that the organization of the $\mathrm{ITD}_p(f)$ curves is greatly influenced by the presence of the cat's body. Building on the pioneering work of [Algazi et al., 2002] in that direction, this method thus potentially allows studying on a systematic manner the effect of body posture on binaural cues.

Part II

Consequences of the variation of ITD on the binaural system

In the previous part I characterized the binaural timing cues to position available to animals of different species. I showed that ITD for a given position depends on the frequency of the stimulus, on a global scale (across different auditory filters) and on a local scale (within a single auditory filter). In this part I study the neural structures that encode the azimuth of a sound source, thus underlying the ability of animals to localize sound sources. I am interested in the consequences of the fact that ITD is not constant across frequencies on the understanding of the binaural system.

In the late 60's, the responses of neurons in the cat Inferior Colliculus (IC) were found to be modulated by the ITD in the inputs [Rose et al., 1966]. It is now known that neurons in an afferent nuclei to the IC, the medial superior olive (MSO), are primarily sensitive to ITD (i.e. not to ILD) in stimuli. I propose to approach the study of those nuclei through two classes of questions.

The first class of questions is *mechanistic*: how is the auditory system arranged (in a broad sense) so that the neurons display a sensitivity to ITD. This involves understanding the neural pathways from the ears to the neurons exhibiting ITD sensitivity: from the cochlear function to the morphology and biophysics of the cells. A response to this question was provided very early by Jeffress [Jeffress, 1948], who postulated that axonal delay lines and coincidence detection could provide a mechanism for ITD sensitivity. This is a *mechanistic* explanation in that it says how the particular input-ouput relationship of a neuron can be explained. To this date, this situation is accepted as a fact in the avian realm [Konishi, 1971], even though it is still discussed in the mammalian realm [Joris and Yin, 2007]. In any case, this will never explain how the nervous system further makes sense of the ITD of a sound sourc to create a percept of space, and ultimately a behavioral response.

The second, maybe more interesting class of questions is about *coding* problems: How do ITD-sensitive neurons collectively represent the position of sound sources? This involves a full understanding of the response of the neurons in an MSO population, across all possible stimuli (different levels, different frequency contents, ...). This is a hard problem because simultaneous measurements of a complete population of neurons is hard, yet, hypotheses about the encoding of position provided by binaural neurons exist. An answer to this question was also provided by Jeffress, who proposed that the ITD could be encoded by the identity of the maximally active neuron in a population of heterogeneous ITD-sensitive neurons. This is referred to as a *peak* code in the literature, and again, while this is the accepted answer in birds, it is still a matter of debate in the mammalian realm.

I argue that the variability of ITD for a given position has consequences on the neural functioning both those issues. The local variation of ITD, we showed, introduced

distortion between the signals in the form of different envelope and FS delays. It is a *mechanistic question*: is it possible that single MSO neurons are *sensitive* to both these cues? The global variations of ITD across the spectrum means that there is no longer a one-to-one mapping from the cue (ITD) to the position of the sound source. It is a coding question: How is the position of broadband stimuli encoded by a population of ITD-sensitive neurons?

This part consists in one introductory chapter (Chapter 5) that describes the current knowledge about the binaural ITD processing system in mammals. I examine the mechanistic problem of single neurons being sensitive to ITD in the signals at the ears, and then study the population coding of auditory space, or how populations of neurons collectively represent space. The rest of the part is dedicated to showing that the two types of ITD variability that we observed in the previous chapter (across and within cochlear channels) have consequences on those scales.

Chapter 6 represents the most important result of this disseration. We use recorded from cells in the cat IC and find that neurons encode frequency-dependent ITDs, consistently with previous observations [Yin and Kuwada, 1983]. We test the hypothesis that this frequency-dependence is explained by the fact that ITDs for a given position in space are frequency dependent. We argue that this means that the cells are coding for position instead of pure ITD, in a peak code fashion. In Chapter 7 we provide the basis for a model of neurons sensitive to a *position* in space irregardless of the frequency content of the input. I show that this can be achieved through a combination of axonal delay lines and mismatches in the wiring of binaural neurons to the cochlea on each side. Finally, in Chapter 8, I discuss the notion of place and populations codes. I propose an alternative peak code to the Jeffress-type code. This model predicts that low-frequency ITD sensitive neurons have BDs outside of the *physiological range* of ITDs, as observed in mammalian data.

Chapter 5

State of the art

Contents

In this chapter I introduce the state of the art of the neurophysiology of sound localization. It is divided into two parts, first I review the results on the mechanisms leading to single-cell sensitivity to ITD, and then turn to the models of how populations of neurons encode space.

I argue that ITD sensitivity necessarily arises in binaural neurons because of a handful of *computational primitives*. First, the *transduction* of the acoustical signals into neural signals must preserve the temporal information imposed by the head. Second, it must be that neural signals are *delayed* by different quantities on the left and right pathways. Finally, binaural neurons must act as coincidence detector neurons, able to detect synchronous inputs from both sides. I then explain how those different computational components are thought to be implemented on the neural substrate. The transduction of the signal, from the cochlea through the Cochlear Nucleus is studied. Possible delaying mechanims are then reviewed, from axonal delay to cochlear delays and inhibition. Finally, the ability to detect coincidences is assessed and key features of the MSO are highlighted.

In a second part I introduce the controversial problem of the encoding of source location by populations of MSO neurons. I review the different schemes that have been devised for the coding of auditory space. This is where the Jeffress-type "peak code" is opposed to the recently proposed hemispheric model [McAlpine et al., 2001]. I review the arguments in favor of each alternative, as well as the criticism they received.

5.1 Introduction

Sound source localization is a hard mathematical problem, for which solutions exist in the form of algorithms generally developed in the context of robotics (see e.g. [Durkovic et al., 2011]). Yet, the robustness of the human auditory system to noise, concurrent sources, different acoustical environments, is remarkable and yet to be achieved using traditional signal processing. This is a good motivation to the study of neural sound source localization. Neural computing is different from classical digital computing, for a main part because neurons communicate with discrete asynchronized events. Therefore one may ask: what makes the neural substrate especially suitable to this processing?

Schematically, the peripheral auditory system is an ascending hierarchy of neural structures. At the bottom of this hierarchy, the acoustical pressure waveform is transduced into an electrical neural signal. This signal is carried by the auditory nerve to the auditory brainstem, which consists of bilateral nuclei (that is, symmetrically present on both sides). Source location can be extracted by the processing of interaural cues, which is necessarily done by comparing the signals at both ears. Some brainstem cells receive bilateral inputs, which makes them interesting candidates for the function of sound localization. From these structures on, the neural signal is carried to the medial geniculate nucleus, a structure in the midbrain, which then "projects" the signal into the auditory cortex.

Here I introduce the neurophysiology of sound source localization in mammals, and more specifically the ITD-processing structures. Two types of problems arise when trying to account for the functioning of the binaural system. First, how can the activity of a single cell be modulated by the ITD This entails both describing the nature of that modulation, and how this modulation arises from properties of the neurons and their wiring. I study ITD sensitivity first, in Section 5.2. Such results, though, will fall short of providing a complete understanding of the system. The step from single-cell sensitivity to such *encoding* spatial position is a conceptual leap that is conditioned by and constrains the observed single-cell sensitivities. In Section 5.3, I explore the current views on how the position of the sound source is represented by binaural neuron populations.

5.2 Neural basis of the sensitivity to ITD

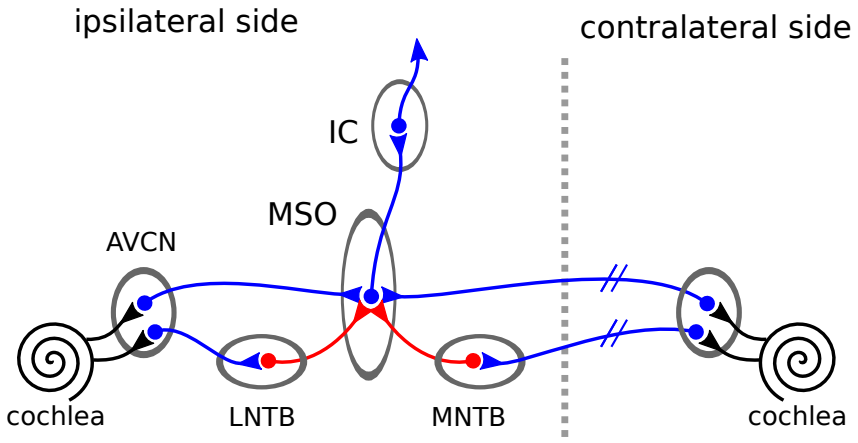

Figure 5.1: Schematics of the auditory pathways leading to Medial Superior Olive (MSO) and Inferior Colliculus (IC). Blue lines signal excitatory projections while red lines signal inhibitory projection. LSO: Lateral Superior Olive, LNTB: Lateral Nucleus of the Trapezoidal Body, MNTB: Medial Nucleus of the Trapezoidal Body, AVCN: Antero-Ventral Cochlear Nucleus

5.2.1 Sensitivity to binaural cues in the auditory brainstem

In most common single-cell electrophysiological setups, an electrode is placed in the vicinity of a neuron in the nucleus of interest, while the animal is anaestetized. The experimenter then presents acoustic stimuli to the subject repeatedly, while the response of one cell is recorded from an electrode. Generally, the stimuli are pure tones (sinusoids), or white noise presented separately at both ears. In each stimulus, the experimenter can introduce binaural cues (ITD, ILD, ...), and then measure the response of the neuron as a function of the cue.

Such work was conducted in the mammalian auditory brainstem in the second half of the XX$^{\text{th}}$ century. An early report [Rose et al., 1966], showed that neurons in the Inferior Colliculus (IC), were *sensitive* to ITD cues and ILD cues in the stimuli. It was later shown that this property is inherited from the Superior Olivary Complex (SOC), which provides inputs to IC [Aitkin and Schuck, 1985,Oliver et al., 2003]. SOC is mostly concerned with localization in the azimuthal plane, as assessed in *lesion* experiments. In this study, [Masterton et al., 1967], the experimenter surgically lesions selected parts of the auditory brainstem and then probes the localization abilities of the animal in behavioral experiments. Results of these studies, and later [Casseday and Neff, 1975], show that lesioning the SOC highly degrades the animal's localization abilities. Overall, reports agree that the SOC is the first locus of binaural cues processing.

The SOC mainly consists in two nuclei, both providing inputs to the IC: the Medial (MSO) and lateral (LSO) superior olivary nuclei. Those structures, much as the IC,

are *bilateral*. Therefore, mammals have two symetrically placed MSOs (resp. LSOs) at both sides of the brainstem. Both are *binaural*, in the sense that they receive inputs from both the ipsilateral and contralateral ear. Moreover, they receive inputs from the contralateral ear through the Cochlear Nucleus (CN) which is almost exclusively a monaural nucleus. Hence, only three synapses away from the ear, SOC structures are the *first* binaural structures in the auditory brainstem. Other periolivary nuclei exist in the SOC (e.g. superior paraolivary, SPON), but will not be discussed in this manuscript.

MSO receives bilateral (from both ears) inputs, via the spherical bushy cells of each side's anteroventral cochlear nuclei (AVCN) [Oliver, 1987, Karino et al., 2011]. A depiction of the pathways leading to MSO is provided on Fig. 5.1. Ipsilateral and contralateral inhibitory inputs also converge to MSO, through the Lateral Nucleus of the Trapezoidal Body (LNTB) and the Medial Nucleus of the Trapezoidal Body (MNTB), respectively. MSO neurons are primarily neurons sensitive to low-frequency signals, relative to the hearing range of the animal. As first demonstrated in the dog by [Goldberg and Brown, 1969], those neurons are sensitive to ITD in the stimulus. In general a given neuron is preferentially activated by free-field sounds played in the contralateral hemifield (e.g. [Aitkin et al., 1985]), hence to ITDs congruent with those positions.

LSO also receives bilateral inputs, but asymetrically. It is excited by spherical bushy cells of AVCN, and inhibited by contralateral inputs through the Medial Nucleus of the Trapezoidal Body (MNTB) [Tollin, 2003]. Neurons in the LSO are primarily excited by high-frequency stimuli. Note that some low-frequency cells in LSO are also sensitive to timing differences in their inputs [Tollin and Yin, 2005]. Yet, the role of LSO seems constrained to the encoding of interaural level differences (ILDs), in a way that is independent of MSO [Tollin, 2003].

Other neural structures are involved in sound source localization, using other sets of acoustical cues. The Dorsal Cochlear Nucleus (DCN) is for example thought to be responsible for the processing of monaural cues (spectral notches) to location [Imig et al., 2000]. These notches are the result of acoustical resonances in the pinnae of the animal, and carry information about elevation even in the median plane. Similarly to MSO and LSO, information contributed by the DCN is then carried on to IC.

ITD sensitivity in different species

Studies of ITD sensitivity in animal hearing have been conducted in a wide array of mammalian species, including the dog [Goldberg and Brown, 1969], cat [Kuwada et al., 1984], rabbit [Kuwada et al., 2011, Batra et al., 1997, Fitzpatrick et al., 1997a]

guinea pig [McAlpine et al., 1996], or mongolian gerbil [Lesica et al., 2010]. Yet, all non-mammal vertebrates also have a sense of hearing and exhibit fine localization abilities. In this regard, the barn owl (Tyto alba) is a very widely studied bird for its localization abilities [Konishi, 1971]. The analog of MSO in the avian auditory system is the Nucleus Laminaris (NL), in which cells are sensitive to ITD [Konishi, 1971, Knudsen and Konishi, 1978]. Other birds share the same organization as the barn owl, for example ITD sensitivity in the chicken [Koppl and Carr, 2008] or emu's NL [MacLeod et al., 2006] is also documented. Some reptilians also have a functional analog of MSO, this is documented in the alligator [Carr et al., 2009] or the gecko. Interestingly, a potential mechanism for ITD sensitivity in the gecko and some birds is the *acoustic* coupling between both ear drums due to the presence a canal between the ears [Christensen-Dalsgaard et al., 2011].

Nature of ITD sensitivity

Remarkably, the response properties of ITD sensitive cells across species are quite similar. Schematically, cells always have a sensitivity similar to that shown on Fig. 5.2, **A**. The neuron discharges maximally around values of ITD that are evenly separated by periods of the Best Frequency of the cell (that is the tone frequency to which the neuron responds the most). The shape of LSO cells tuning to ILD is noticeably different, with a sigmoidal shaped tuning curves. Because ITD sensitive cells in many species have the shape depicted on Fig. 5.2, it was quickly assumed that the basic neural mechanisms were the same [Phillips and Brugge, 1985]. This is despite the fact that the auditory system has evolved differently for, e.g., mammals and birds.

When measured with white noise stimuli, the tuning curve of a neuron to ITD is termed the *noise-delay* curve (Fig. 5.2, **A**). The response of binaural neurons is modulated by the frequency of the input, and the frequency that yields the strongest response is known as the Best Frequency (BF). Typical noise-delay curves have multiple modes (or peaks) separated by one period of the BF of the cell (Fig. 5.2, **A**). The Best Delay (BD) is a common measure of the ITD sensitivity of a neuron, it measures the ITD that elicits the strongest response across ITDs, i.e the peak of the noise-delay curve. The BD is in general within the π-*limit*, that is within one CF period of zero ITD. Therefore, the biggest peak of the noise delay curve is most often the closest to zero ITD (as on Fig. 5.2, **A**), [Joris and Yin, 2007, Fontaine and Brette, 2011]. Sometimes for simplicity, BD is also measured with a pure tone at BF. This is justified because the noise-delay curve is in fact well approximated by a weighed sum of tone-delay functions, with a strong weight on the tone-delay curve at BF [Yin and Kuwada, 1983].

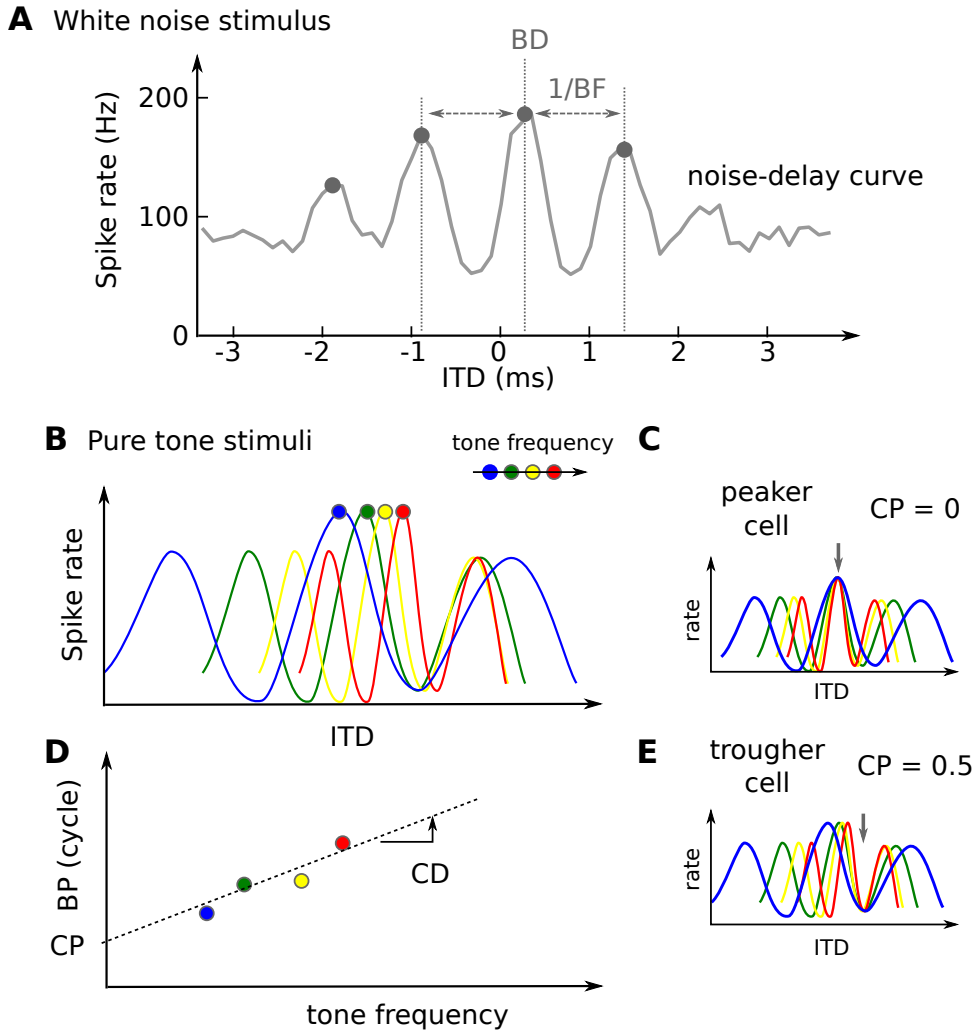

Figure 5.2: Illustration of measures of ITD sensitivity. **A** Schematics of the noise-delay curve of an MSO cell. Grey dots indicate the peaks of the function, they are separated by one period of the BF of the cell, here about 1kHz. The main peak is the BD, here about 200 μs. **B** Tone delay curves are tuning curves to ITD measured with tones at different frequencies (insert). Measured in phases, the BD of each tone-delay curve is the BP of the cell **E**. Linear regression of the BP points provide CP and CD (**D**). **C** and **E** show the tone-delay curves for a peaker (CP $= 0$)and a trougher cell (CP $= 0.5$). Notice how either the peaks (in **C**) or the troughs (**D**) are aligned (downward gray arrows).

Description of stimulus-dependent ITD sensitivity When measured with pure tone stimuli at different frequencies, the tuning curve of a neuron is the *tone-delay* curve (Fig. 5.2, **B**). A "Best Delay" measured on a tone-delay curve at frequency f, is oftentimes reported as a phase quantity (in cycles) the Best Phase, $BP = BD \times f$. Measured on tone-delay curves, the BD generally varies with the frequency of the tone (e.g. in the cat MSO [Yin and Kuwada, 1983, Yin and Chan, 1990]). The relationship between BP and frequency is almost linear, therefore two quantities characterize the delay sensitivity of a cell: the Characteristic Phase (CP) and Characteristic Delay (CD) [Yin and Kuwada, 1983, Fitzpatrick et al., 2000]. These are the intercept (CP) and slope (CD) of a linear fit of the BP data against frequency. This is depicted on Fig. 5.2, **D**. Assuming that the cell has a constant BD, then $BD(f) = CD$, therefore the characteristic phase is zero. Therefore, CP measures the degree to which BD varies with the frequency of the stimulus.

This has lead to the classification of cells into two classes: *peaker* and *trougher* cells. When $CP = 0$, BD is constant so visually the tone-delay curves at different frequencies have aligned peaks. On the other hand, when CP is close to half a cycle, then the minima (troughs) of the tone-delay curves align for different frequencies (see Fig. 5.2, **C**-**E**). Trougher cells, because their response is suppressed for small ITDs, contribute to the encoding of large ITDs in low-frequency sounds [Fitzpatrick et al., 2000]. The distinction in peakers and troughers is somewhat artificial since available data show that in ITD sensitive cells, the distribution of CP is essentially unimodal (e.g. [Yin and Chan, 1990]), that there is a continuum of response types rather than two separate classes of cells.

5.2.2 Computational primitives necessary for ITD sensitivity

The *function* of MSO in mammals and NL in birds is to process the azimuth of a sound source from the binaural inputs. Those ITD-sensitive structures in different species exhibit similar response properties, suggesting that similar computation strategies are used. A theory explaining such strategy had been suggested prior to any electrophysiological evidence, by Jeffress [Jeffress, 1948]. It should be clear to the reader that I am not discussing the "place code" meaning of the Jeffress model here. The population-wise predictions of the Jeffress *code* will be discussed in the next section (Section 5.3). Rather, I explain what core computational components are used in the Jeffress model.

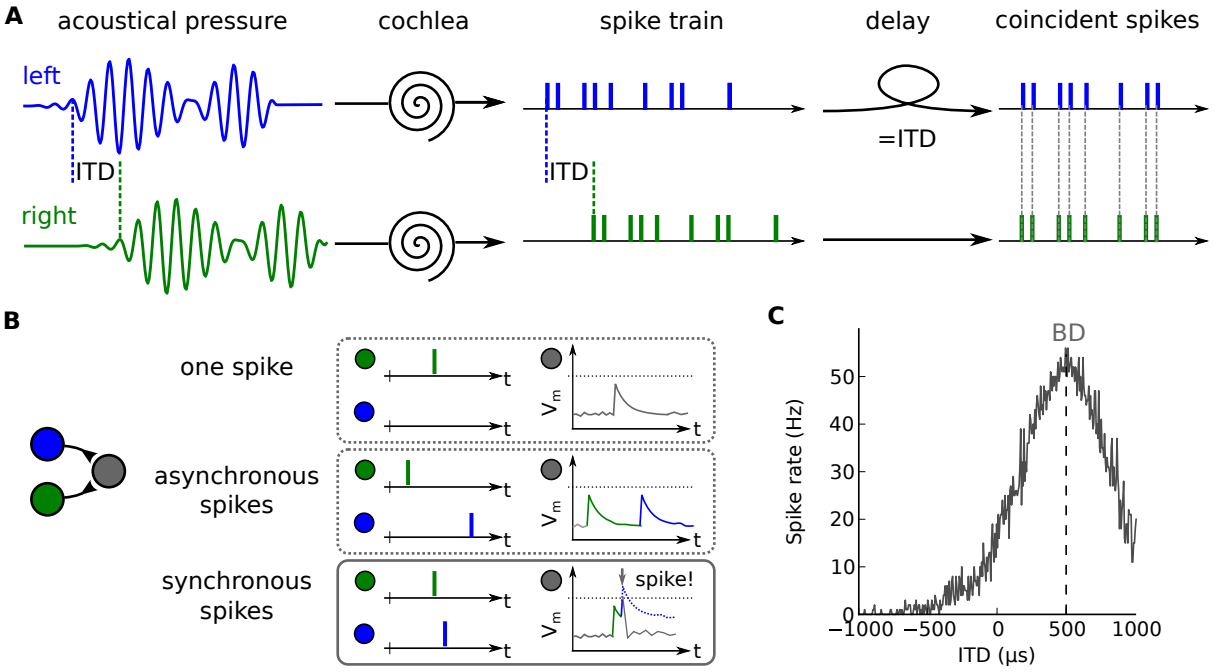

Figure 5.3: Schematics of the Jeffress model of ITD sensitivity. **A** Acoustical pressure waveforms at both ears are transformed into spike trains via the cochlea. The spike trains are also delayed versions of one another. If the leading signal is delayed by an amount equal to the ITD then the spike trains are coincident. **B** Coincidence detection in LIF neurons. Two presynaptic neurons (blue and green) target the same postsynaptic neuron (gray). In each panel, the spike outputs from the two presynaptic neurons are reported alongside a schematics of the membrane potential of the postsynaptic cell (V_m). Dotted line is the threshold. With a single presynaptic spike, or asynchronous spikes, the membrane potential does not reach threshold. On the opposite when the spikes are synchronous, the EPSPs sum and the threshold is reached, and the neuron spikes. **C** shows the result of such a simulation with a neuron whose BD is 500 μs. The neuron spikes more strongly when then acoustical inputs have an ITD close to its BD.

Computational primitives for ITD sensitivity

Let us assume that the acoustical environment is made of stimuli with a frequency-independent ITD depending on the azimuth of the sound source (this is the case when the head is acoustically transparent, Chapter 1). The acoustical signal reaching the tympanii are equal up to a delay, the ITD. This signal is then transduced into a "neural signal" by the cochlea, as shown on Fig. 5.3, **A**. The first important point is that the transduction operation must preserve the structure imposed by ITD on the input. Put simply, if the acoustical signals are delayed versions of one another by a given amount, the neural signals must be delayed as well to maximize their cross-correlation (**A** of Fig. 5.3). Under this condition, a cell is sensitive on ITD if it is sensitive in the difference in delays of the neural signals from both sides s_l and s_r. A cell with interaural neural delay that compensates for the ITD in the inputs will receive perfectly synchronous inputs. This interaural neural delay τ is the difference of the delays at both ears: when it is positive, the delays from the right side are longer than the left side.

In fact, neurons are sensitive to the synchronicity of their inputs [Rossant et al., 2011], a property known as *coincidence detection*. This is shown in Fig. 5.3, **B** for a LIF neurons receiving inputs from two presynaptic neurons. Therefore, a neuron with interaural neural delay τ will fire more strongly when the presented sound has an ITD equal to τ than for other ITDs. The cell is tuned to the ITD that compensates for its interaural neural delay. In summary, ITD sensitivity can be achieved through *faithful transduction* of the physical signal into a neural signal, *delaying* of the neural signals from both ears before they converge to a *coincidence detector* neurons.

In a 1959 paper, [Licklider, 1951] argues that this scheme can also be applied to the computation of the autocorrelation of the monaural signals, which can be used to model pitch perception. Furthermore, a recent paper [Brette, 2012] shows that a similar scheme based on delays and coincidence detection can be used in a variety of sensory tasks, e.g. in olfaction. In the rest of the chapter, we study the experimental and theoretical support for the existence of those different mechanisms in the context of ITD processing.

Illustration: An ITD-sensitive toy model

Those computational properties can be implemented in a simple spiking-neuron instantiation of this model. This subsection is meant to be a quick illustration of how those components can make a spiking neuron sensitive to ITD, and what are the limitations of the different components of such a model.

Transduction The only assumption needed here is that the sequence of spikes at the left and right ear preserve the ITD in the input. In this toy model, we derive the spike train for one input side using a noisy Leaky Integrate-and-Fire neuron (see, e.g. this recent paper from the group [Fontaine et al., 2013]). The acoustical signal is simply rectified, and integrated to yield the membrane potential of a model cochlear hair cell. This hair cell is set to discharge (i.e. emit a spike) whenever the membrane potential reaches a threshold. As usual, the membrane potential is then reset to 0. The result of this model is a sequence of spike times, possibly representing the spike trains propagating along the ipsilateral auditory nerve. The contralateral spike train is then explicitly delayed from the ipsilateral one.

Delay mechanism Monaural sequences of spikes reach a binaural cell with different delays on each side. These could be, for example, due to differences in the presynaptic axon morphology of the inputs to the binaural cell. Due to limited propagation speed along axons, differences in length or diameter of the axon contribute to imposing differential propagation delays onto the spike trains [Seidl, 2013].

Coincidence detection A leaky integrate-and-fire binaural neuron is used, for which the presynaptic spikes have exponential Excitatory Postsynaptic Potentials (EPSPs) (as depicted on **B** of Fig. 5.3). That is, the neuron's membrane potential reflects the integration of the sum of the spike inputs from both ears. This is presented in more details in Appendix 5.A.

Such neurons are sensitive to coincidences. If the binaural I&F neuron has a low enough membrane time constant ($\tau_m = 1\text{ms}$), it can be made to spike only when presented with synchronous inputs. If the spike trains are asynchronous, the depolarization of the membrane is not sufficient to make the membrane potential reach threshold Fig. 5.3, **C**.

Therefore, the ITD for which the neuron responds the most in this model, the Best Delay (BD) corresponds to the difference in delays between the monaural pathways. If a sound is presented with an ITD away from a given cell's BD, the chance of observing coincident inputs decreases, and hence the firing rate of the binaural cell.

The simulation of the spike rate of such a neuron is presented on **D** of Fig. 5.3, for white noise input stimuli and a BD of $500\mu s$. It reproduces the sensitivity of a neuron to an ITD in the stimuli. More details on the model are available in the Appendix 5.A.

Preliminary conclusion

One can identify three key properties of the system that lead to ITD sensitivity. First, the *transduction* of the acoustical signal into a neural signal must preserve the temporal structure of the signal. If the binaural sound has a given ITD, then the neural signals at the output of the cochlea must also be delayed versions of one another. Temporal information can be lost, for example, because of time-variant transformations at the cochlea. Second, mechanisms that impose *delays* on monaural neural signals are necessary so that each cell receives a differently-delayed signal. Those delays must be able to produce sensitivity in cells consistent with ITDs in the environment, that is have effects of the same order of magnitude. Third, it must be ensured that binaural neurons are sensitive to the *synchrony of their inputs*. In particular, because ITDs are typically very small (hundreds of microseconds), the binaural cell's sensitivity to synchrony must be precise enough.

5.2.3 Auditory transduction

The pressure waveform in the ear canal generates a vibration of the tympanic membrane which excites the ossicles in the middle¡-ear. Those ossicles are part of a complex impedance-matching mechanism, that transmits the acoustical energy to the oval window of the cochlea. When the sounds entering the ear have very high levels, middle ear muscles stiffen the ossicles to avoid damage on the cochlea (an acoustic reflex).

The cochlea itself is a coil-shaped tube containing two compartments longitudinally separated by the basilar membrane. When the oval window is set in vibration, the basilar membrane is excited by a travelling wave. The locii of greatest excitation on the basilar membrane depend on the stimulus's frequency content: low frequencies excite the apex of the basilar membrane (where it is narrower), and the higher frequencies at the basal part of the membrane. This property is known as *tonotopy*: it refers to the fact that changing the frequency of the stimulus also changes the *place* where cells are the most excited by it. Tonotopy is a very important property, inherited as we will see in most nuclei of the auditory system. Characteristics and models of this property are described in further details in Subsection 5.2.3.

The basilar membrane still carries a mechanical signal: it can be measured by the displacement of the membrane and/or the fluid in the cochlea. Specialized cells in the cochlea transduce the mechanical vibrations into ionic currents: those are the inner hair cells (IHC). IHC are arranged in bundles on the basilar membrane (stereocilia), when the fluid is set in motion by a sound, the hair cells mechanically vibrate. Due to those vibrations, parts of the cells are mechanically *stretched*. In turn, special-

ized *mechanoreceptor* channels present on the membrane of those cells translate the mechanical stretch into ionic currents.

Inner hair cells, unlike common neurons, do not emit action potentials. Instead the hair cells excite the nearby spiral ganglion cells (SGC, via neurotransmitters). This depolarization in turn leads to a spike being produced and propagated along the axons of SGC, which make up the auditory nerve (AN). AN fibers provide input to Cochlear Nucleus (CN) neurons, which exhibit different types and specializations. Excitatory efferents to MSO are the spherical bushy cells (SBC) of the antero ventral CN (AVCN), which reliably transmits the temporal structure of AN firing [Wu and Oertel, 1984]. Octopus cells (also named onset cells) detect broadband transients in the stimulus [Spencer et al., 2012], and stellate cells encode the amplitude spectrum of stimuli. The Dorsal Cochlear Nucleus (DCN) is thought to process spectral cues to the elevation of sound sources [Oertel and Young, 2004].

Cochlear filtering

The primary property of the cochlea is the fact that it introduces *frequency-selectivity* in the downstream system. Different points of the cochlea are excited preferentially by different frequency bands. For this reason, the cochlea is usually modeled using a bank of band-pass filters: the displacement of the basilar membrane at any point on the cochlea is a band-pass filtered version of the vibration at the oval window (e.g. in [Zhang et al., 2001]). The characteristics of those filters depend on the position of the point along the cochlea, in a *tonotopic* way: nearby points on the basilar membrane are excited by close stimulus frequencies. AN fibers connect on the cochlea at different points, therefore frequency-selectivity is carried in the AN [Narayan et al., 1998]. Similarly, this property is carried onto downstream nuclei, wherein cells are anatomically arranged by their preferred frequency band. As an example, cat MSO neurons are arranged on a dorsoventral tonotopic axis [Guinan et al., 1972].

This frequency selectivity is usually characterized through two measures: a measure of the bandwidth of the filter and a measure of the center frequency of the filter.

Best and Characteristic Frequencies Measurements of the cochlea-induced frequency-selectivity can be based on observations of the basilar membrane displacement. However, it is oftentimes measured via the spiking frequency of auditory nerve fibers, because measurements at the apical part of the cochlea (low-frequencies) are hard. Simultaneous measurements reveals that the spiking frequency in AN fibers is a high-pass filtered version of the basilar membrane displacement [Narayan et al., 1998].

Most auditory neuron's spiking frequency varies when presented with pure tones at different frequencies. The measurement of the spike rate as a function of frequency is known as a rate-frequency curves. The peak of a rate-frequency curve is a measure of the selectivity of the neuron termed the Best Frequency (BF). Rate-frequency curves change with the level of the stimulus [Pickles, 1988], therefore so does the BF: it decreases with the level of the input [Ruggero and Rich, 1983]. This is reflective of cochlear non-linearities [Sachs and Abbas, 1974].

The Characteristic Frequency (CF) is by construction a level-independent measure of frequency selectivity [Pickles, 1988]. The CF is intrinsically a *threshold* measure, defined as the frequency that triggers a significant response at the lowest sound level. Intuitively, it is the BF measured with the softest sound that can be distinguished from looking at the data. To measure a CF, multiple rate-frequency curves have to be obtained at different levels, therefore commonly only the level-dependent BF is measured.

Measures of cochlear bandwidth Tone-frequency curves generally have a pronounced mode, where the cell is most active over a certain bandwidth. It is measured by fixing a threshold on the tone-frequency curve, and often reported as a quality Q measure (where $Q = BW/CF$) [Pickles, 1988]. In typical AN measurements, the rate-frequency curve of a fiber is recorded. The Q_{10} is measured from the frequency bandwidth BW over which the discharge rate is 10 dB above the noise threshold [Evans, 1972, Pickles, 1988]. Recordings in the cat show that these are generally between 1 to 10 (over a range of CFs between 100 Hz and 15 kHz) and between 1 and 4 below 2 kHz [Evans, 1972, Lopez-Poveda, 2005].

In humans, measurements of cochlear selectivity relies on less invasive techniques such as psychoacoustical methods or otoacoustic emissions. Behavioral studies quantify the bandwidth of the filter generally using frequency-masking paradigms (e.g. [Oxenham and Shera, 2003]). These are more prone to variations across paradigms than animal models and it is unclear how exactly they relate to cochlear functioning [Ruggero and Temchin, 2005]. Furthermore, these behavioral studies are hard to replicate on lab animals for which auditory nerve data is available.

Another method to measure cochlear bandwidths is through the measure of sounds emitted by the cochlea. Because of the cochlear amplification, when presented with a sound the cochlea produces measurable sounds termed otoacoustic emissions. Those can be used to infer the cochlear tuning width, and can be used in animals and humans alike [Shera et al., 2010, Joris et al., 2011]. Recent data using otoacoustic emissions in humans and monkeys seem to indicate that humans do indeed have a very sharp

filtering (Q between 10 and 20), sharper than cats (Q between 3 to 10 on the same frequency range) [Joris et al., 2011].

Models of cochlear filtering Models of cochlear filtering have to account for both the basilar membrane dispacement and the auditory nerve firin. Responses of human subjects in psychoacoustical experiments can be explained by a model of the auditory periphery consisting in a bank of bandpass filters followed by a compression operation [Patterson, 1971, Irino and Patterson, 1997]. When AN fiber responses are used, more features have to be accounted for, AN responses typically exhibit adaptation, synchronization, high dynamic range [Lopez-Poveda, 2005]. Therefore, these models must include mechanisms modeling the inner and outer hair cell functioning as well as the spiral ganglion cell spiking. A way to build such models is to compute revcor filters, using the reverse correlation of the spike train and the sound [Carney and Yin, 1988]. This method has been developed into models of AN properties in the Carney model [Carney, 1993] and later Tan & Carney [Tan and Carney, 2003]. Other models include DRNL filters [Lopez-Poveda and Meddis, 2001, Sumner et al., 2003], and are thoroughly reviewed in [Lopez-Poveda, 2005]. All these model architectures are implemented in the Brian Hears package, developed in the lab which was the subject of a publication [Fontaine et al., 2011].

Temporal properties of auditory transduction

A second important aspect of sequences of spikes on single auditory nerve fibers is that they display a high degree of temporal structure. When an AN fiber is recorded in response to a pure tone, the neuron tends to spike more at certain parts of the period of the stimulus, i.e. specific phases. This is reproduced by the toy model response of the introduction, as pictured on Fig. 5.4, **A**. When the stimulus is a pure tone, this property is referred to as *phase-locking*, and was observed in a wide range of animal models. Phase locking is only defined for pure tone stimuli, that have provide a good definition of phase. When a broadband stimulus is presented, then there is no real stimulus phase, but still temporal structure referred to as *time-locking* Because the stimulus undergoes the narrow cochlear filtering, then the cochlear output is best understood as having two components: an envelope and a temporal fine structure (TFS). TFS is essential to many aspects of human hearing, including speech [Lorenzi et al., 2006]. In mammals, phase-locking was first observed in the AN of squirrel monkeys [Rose et al., 1967], and is now very well documented in many nuclei up to IC [Dynes and Delgutte, 1992, Joris et al., 2006, Sullivan and Konishi, 1986].

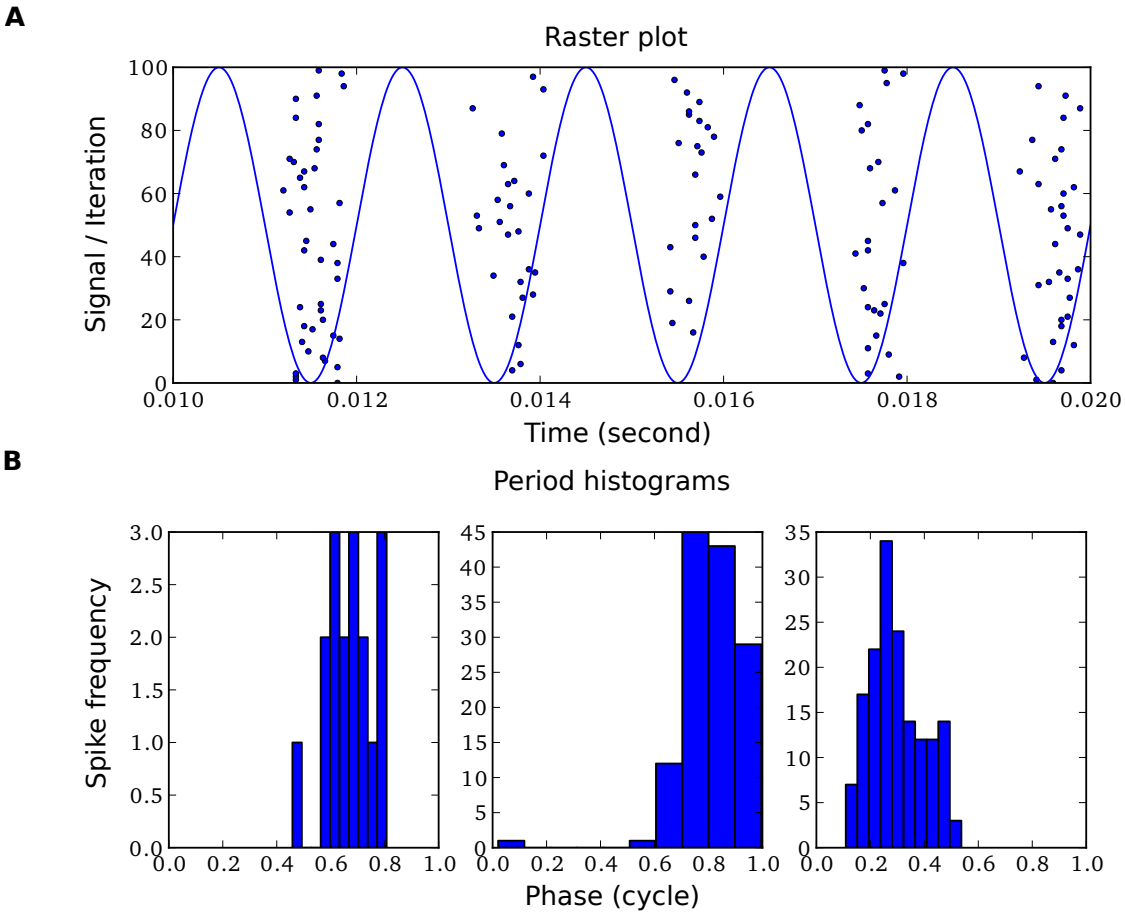

Figure 5.4: Phase locking to pure tones in a model of AN fiber **A** Superimposed raster plot and stimulus (a 500 Hz pure tone). Each model fiber tends to spike at a given phase of the tone. **B** Histograms of relative spike phase (in cycles) of three different model fibers. The distributions have a well-defined mode showing that spiking occurs more at some phases relative to the tone than others.

The Vector Strength (VS) measures the phase locking abilities AN fibers. Spikes are aggregated in *period histograms*, which the histogram of spike phases relative to the stimulus tone (Fig. 5.4, **B**). The VS is the magnitude of the resultant vector of this period histogram [Sullivan and Konishi, 1984]. It is a measure of the inverse of the dispersion of a circular distribution (see, e.g. [Mardia, 1975]), thus quantifies to quality of the phase locking of a spike train to the stimulus.

Phase locking to the stimulus frequency degrades as the CF of the fiber increases. An influential hypothesis for this low-pass behavior is electrophysiological: it is due to limited time-constants in the IHC [Palmer and Russell, 1986]. In this study, the upper limit for phase-locking was found at 3.5 kHz in guinea-pigs. For cats, phase locking cuts off at around 3-5 kHz ([Rose et al., 1967], [Kiang, 1965]). Human data [Joris and Verschooten, 2013] suggests that phase-locking in humans is not higher than a few kHz. For higher frequencies, high-CF AN fibers time-lock to the envelope of high-frequency sounds (in the cat, [Dreyer and Delgutte, 2006]).

The precise value and reason for this upper limit is still debated. For example, some species display phase-locking considerably further up in CFs. Barn owls AN fibers phase-lock up to 10kHz [Koppl, 1997], consistent with the fact that they use ITD information in those high frequencies [Konishi, 1971].

Time-locking, much as tonotopy, is carried on to some extent in the higher nuclei of the auditory system. Intuitively, jitter build up and increasing time constants should contribute to a degradation of time-locking [Koppl, 1997, Anderson, 1973, Paolini et al., 2001]. Still, enhancements of the quality of time-locking have been found in axons of low-CF cells from the AVCN (in the cat [Joris et al., 1994]). In any event, temporal coding in the inputs of MSO neurons is well-preserved, [Wu and Oertel, 1984, Joris et al., 1994], and remarkably invariant on level [Michelet et al., 2012]. Calyces of Held, the biggest synapses in the brain [Cant and Casseday, 1986] seem to be a specialization of the auditory system to preserve spike-timing information. Those synapses are present in the MNTB, that is on the contralateral inhibitory pathway to MSO and LSO [Schneggenburger and Forsythe, 2006] An hypothesis is that their size makes them highly precise and reliable in the transmission of spikes.

As discussed in the introduction, it is necessary for ITD based sound localization that temporal information be preserved in the spike trains leading to binaural neurons. The auditory periphery leading to MSO is a temporally reliable system, that faithfully conveys ITD information [Louage et al., 2005].

5.2.4 Coincidence detection property of neurons

Integrate-and-fire neurons are a good description of biological neurons [Brette and Gerstner, 2005], while being sufficiently graspable, therefore I will be use them to explain the notion of coincidence detection. An integrate and fire neuron has a continuously varying membrane potential and a threshold that is responsible for spike output. In those models, presynaptic spikes have an additive impact on the membrane potential, and in the absence of an input, the membrane potential relaxes to a resting value. When the membrane potential reaches the threshold because of incoming spikes, an impulse is transmitted to the postsynaptic neurons, and the membrane potential is reset. This mechanism makes neurons good *coincidence detectors*, indeed, if the spikes they receive are aligned in time, their contribution will sum and increase the probability of spiking. When they arrive out of timing, the membrane potential might have relaxed enough for the second spike to fail to make the potential reach threshold much as in Fig. 5.3, **B**. This feature is seen as a key computational feature of the nervous system, leading to synfire chains [Abeles, 1991], polychronization [Izhikevich, 2006], [Brette, 2012]. When cells are in a regime where excitatory and inhibitory balance each other, neurons are maximally performant in detecting coincident spikes. This was shown using *in vitro* preparations, and is well explained using integrate-and-fire models [Rossant et al., 2011, Rossant and Brette, 2010].

A typical action potential lasts 1 ms, which is somewhat higher than the maximal ITD in most species. Therefore, in the study of the MSO, the mechanisms underlying the microsecond precision of ITD sensitivity has been the subject of much research [Carr and MacLeod, 2010]. It was shown that MSO neurons receive few number of excitatory and inhibitory inputs, driving neurons in the *balanced regime* [Couchman et al., 2010] where they are most sensitive to ITD. Furthermore, it was proposed that low-threshold potassium currents [Svirskis et al., 2004] enchance coincidence detection, which is also the case in birds [Rathouz and Trussell, 1998]. Voltage-dependent KLVA channels could ensure sharpening of EPSPS in the gerbil MSO, [Mathews et al., 2010] and further ensure temporal fidelity. EPSPs are also amplified at the soma, as shown in [Scott et al., 2005] with voltage and current clamp techniques. Even assymetries between the monaural inputs, for example in the EPSPs [Jercog et al., 2010] seem fine tuned to improve ITD processing. Also of interest is the fact that this coincidence detection ability might be enhanced by dentritic tree geometry [Agmon-Snir et al., 1998].

In the chicken, for example it was found that the coincidence detection abilities measured in NL (the avian equivalent to MSO) was consistent with the observed ITD sensitivity [Kuba et al., 2003]. All in all, it seems that the MSO (resp. NL) exhibits a number of specializations that allow it to be sensitive to coincidences on very short

time-scales of the order of tenth of microseconds, relevant to ITD coding.

5.2.5 Neural delays

As underlined in the introduction, neural delays are an essential prerequisite of the computation of ITDs. Moreover, they are an important matter of debate among the community of binaural auditory processing. Interestingly, the nervous system affords many ways of implementing such delays, more or less specific to auditory processing. The simple conceptual way of introducing neural delays in the transmission between two neurons is to make assumptions as to the geometry of the axon between the two. Indeed, the propagation speed of spikes along myelinated axons of neurons (of the order of 10-40 m/s, [Hutchinson et al., 1970]) is modulated by the length and diameter of the axon (see, e.g. [Seidl, 2013]). Longer axons, as well as smaller diameters lead to longer transmission delays (other parameters, such as the inter-Ranvier node distance also greatly affects the propagation delay).

Axonal delays hypothesis

Anatomical evidence for different geometries of the axons targetting MSO, including length and diameter, exists in mammals [Beckius et al., 1999, Smith et al., 1993, Oliver et al., 2003]. All in all, it is typically found that axons can provide delays, consistent with BDs measurements in bird species, but not in mammals. Data in the cat [Karino et al., 2011] show that axon geometry cannot account for the full range of observed BDs. This is still debated, as in [Seidl et al., 2010], the authors argue that the complex 3D morphology of axons is necessary to account for axonal delays covering the full physiological range of ITDs (this includes both the Ranvier nodes distances, length and diameters of the axons). In smaller mammals (gerbil, guinea pigs), though, some BDs of low-CF cells are very big (circa 1 ms). This is not merely consistent with the axonal delays that are found in all studies, and raises an important problem as to the possible origin of such delays.

The lack of evidence for a map of ITD, and the unclear quantitative match between the BDs and axonal propagation time delays has led to the proposal of other sources of delays in the context of binaural processing. Yet, it must be noted that axonal delays only are indeed able to account for a big part of the BDs in most species, all in the case of birds.

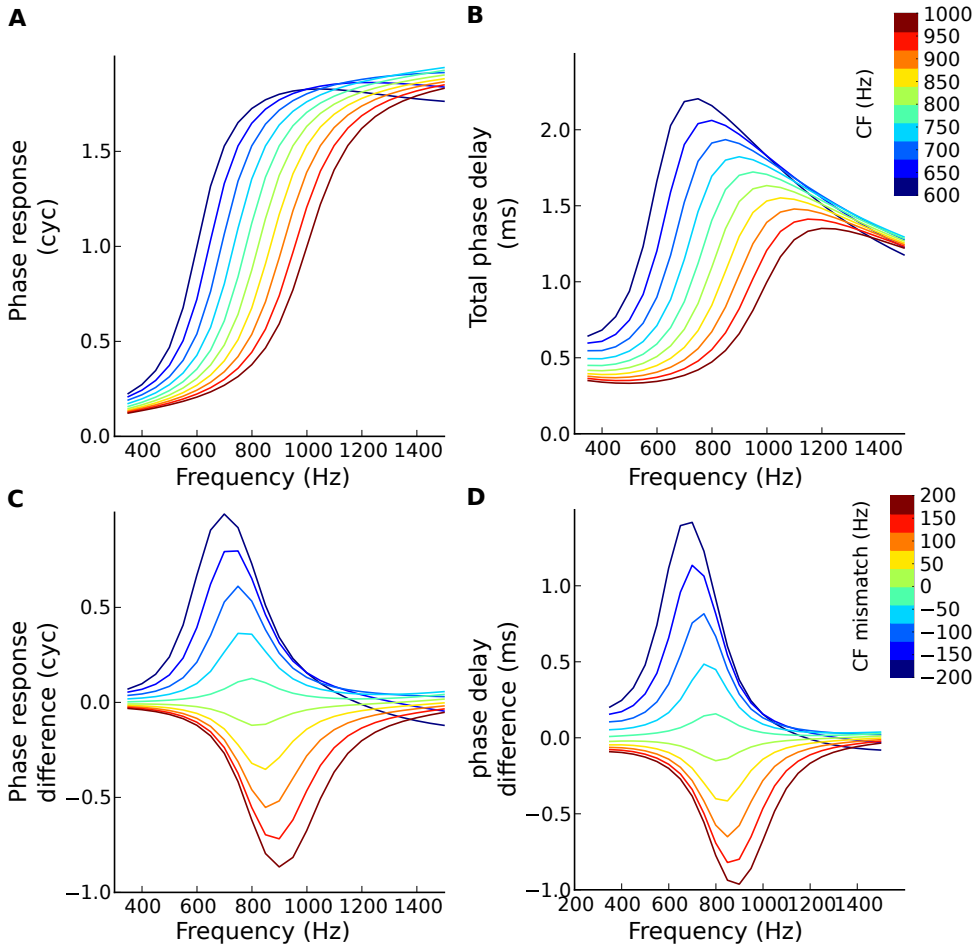

Figure 5.5: Cochlear mismatches in Gammatone filters. **A** and **B** depict the phase response and the total phase delay derived from gammatone filters at different frequencies (color code on the right). Observe that phase responses are non-linear, and all different. On the second row, the difference of the gammatone filter at 800 Hz and at 800 Hz plus a CF mismatch are shown. **C** shows the difference in phase response and (**D**) total phase delay as a function of frequency and CF mismatch (color code to the right). The total phase delay (**D**) shows that: a) it is possible to construct non zero intaural neural delays with CF mismatches and b) that this creates non-linear phase distortions.

Cochlear mismatches hypothesis

Another possible source of neural delays in the binaural system comes from the mechanics of the cochlea. Indeed, it is known that the excitation imposed on the cochlea when presented with a sound takes the form of a *travelling wave*. On top of the tonotopy property already mentioned, points on the cochlea are excited *at a different times* depending on their position. Close to the oval window, at the high-CF base of the cochlea, cells are excited by the travelling wave before the apical part of the cochlea (low-CF). Intuitively, one can introduce a delay between cochlear channels, at the expense of slightly mismatching the monaural CFs. This idea was first proposed by Schroeder [Schroeder, 1977], and later the name *stereausis* was proposed in a more complete development by [Shamma et al., 1989], and [Bonham and Lewis, 1999].

Let us assume that a binaural neuron receives inputs from the contralateral cochlea at CF and the ipsilateral cochlea at $CF + \delta CF$. If the difference δCF is not too big, then it can be assumed that the two frequency ranges on both sides are the same, i.e. the neuron is sensitive to the same left and right frequency bands. But, because of this small difference, the signal it receives on the ipsilateral side will arrive later than at the contralateral side. This is shown on Fig. 5.5, assuming that the neural signals input to a binaural cell are Gammatone-filtered versions of the signal. **A** and **B** of Fig. 5.5 show the phase response and the total phase delay of Gammatone filters. The magnitude of the created delays can be seen from the bottom two plots of Fig. 5.5 (**C** and **D**) which show the expected difference in interaural phase and delay for small CF mismatches (less than 0.2 octave). The range of possible delays is quite wide, and consistent with the magnitude of ITDs: hundreds of microseconds [Bonham and Lewis, 1999].

Cochlear mismatches indeed introduce propagation delays that are consistent with observed ITDs. Furthermore, [Joris et al., 2006] showed that the response of AN fibers could be correlated across frequency channels to predict the inverse BD/BF relationship observed in MSO. In the barn owl, analysis of the monaural revcors (i.e. empirical monaural filters computed from spikes in NL) suggests that cochlear mismatches are not required in order to explain the ITD sensitivity observed in NL [Pena et al., 2001]. The authors show that cochlear mismatches, despite being present, did not correlate with the BD, which cannot formally rule out the contribution of cochlear mismatches in conjonction with other counterbalancing source of delay (e.g. axonal).

More recent studies have even shown that axonal delays and cochlear mismatches could account for finer features of the ITD sensitivity of cells. As an example, it is commonly observed that MSO cells have frequency-dependent BDs (see Chapter 6 for more on this), which can also be accounted for by cochlear mismatches (see, e.g. the recent [Day and Semple, 2011] study). This is because cochlear mismatches intrinsically

provide frequency-dependent interaural neural delays, as shown on Fig. 5.5 **D**. This specific point will be in more depth in Chapter 7.

Inhibition hypothesis

Including inhibition in a model of localization originates from [Lindemann, 1986], as an attempt to predict the lateralization of binaural stimuli in psychoacoustical data. This was at the time also presented to account for the *law of first wave front*. This phenomenon, occuring in the presence acoustical reflections, refers to the fact that perceptual location of a source in the presence of reflections is dominated by the first arriving wavefront.

The presence of ispilateral glycinergic inhibition targetting MSO has been a longlasting puzzle in the sound localization neurophysiology field. This is interesting because inhibition plays a role in the sensitivity to ILD in the LSO (e.g. [Tollin and Yin, 2005]). But it is also present in MSO, wherein MNTB cells also project inhibitory, glycinergic connections to MSO cells [Grothe and Sanes, 1994].

The first observation that glycinergic inhibition may influence the ITD tuning of single cells was brought by in vitro studies of the timecourse of EPSPs and IPSPs [Grothe and Sanes, 1994]. The intuition is that by arriving precisely-timed, inhibition could change the shape of the ipsilateral post-synaptic potential (PSP). Intuitively, if the peak of the PSP is shifted, then the ITD at which the cell is most excitable is shifted accordingly. Therefore, the BD of the cell may also be shifted. Furthermore, if the bilateral inhibition is assymetric this shift ought to be in one direction. Experimentally, the presence of inhibition shifts the BDs to more contralateral values (bigger absolute ITD) by hundreds of microseconds [Brand et al., 2002].

The original *in vitro* study was followed by the very influential [Brand et al., 2002] study, which showed *in vivo* that gerbil MSO cell's ITD sensitivity was modified when glycinergic inhibition was blocked. The model provided in this study also provided insight into the mechanism, which requires highly precisely timed inhibition. Those results were replicated by [Pecka et al., 2008] who also conclude that precisely timed inhibition plays a critical role in the shaping of single-cell BD sensitivity. Modeling studies confirmed these requirements [Zhou et al., 2005], and also attribute a role to the precise anatomy of MSO cells, receiving inhibition at the soma and excitation at the dendrites.

Other interesting properties include the fact that the frequency-dependent BDs observed in cells can be accounted for only by contralateral phase locked inhibition. This was first suggested by a rabbit model [Batra et al., 1997], and later on gerbil data [Leibold, 2010]. The second study also shows that axonal propagation delays

are superfluous, because inhibition was able to explain all their data. This is to be contrasted with the recent [Day and Semple, 2011] study wherein the authors show that cochlear mismatches and axonal delays alone are sufficient to account for the frequency-dependent BDs observed in gerbil data.

Yet, the mere fact that glycinergic inhibition shifts the BD curves is still a controversial issue, even 20 years after [Grothe and Sanes, 1994]. The assumption that contralateral inhibition operates on fast time-scale is debated, which is an essential requirement of the models in [Leibold, 2010, Brand et al., 2002]. Using a different approach, [vanderHeijden et al., 2013] show that inhibitory inputs do not affect the ITD tuning of gerbil MSO neurons. In the avian realm, while inhibition is known to exist, but the contribution to ITD tuning is unclear [Burger et al., 2011].

5.2.6 Conclusion

We have observed that binaural cells of the SOC in mammals are tuned to ITD. This tuning seems to arise by a combination of coincidence detection and neural delays. Interestingly, a number of specializations observed in SOC neurons make this computation easier. Transduction of sound pressure waves at the ears preserves the temporal differences in the signals (i.e. the ITD) by time-locking to the sound wave's TFS. Biophysics of MSO neurons make them especially sensitive to coincidences on a very short time-window. And finally, different mechanisms give rise to neural delays, amongst which the most proeminent seem to be cochlear mismatches and axonal propagation delays.

5.3 Coding for auditory space

The neurons sensitive to ITDs in the mammalian brainstem exhibit a high degree of heterogeneity in their responses. In most cases, a given MSO cell can be characterized by two parameters. A measure of the frequency selectivity of the cell, e.g. the Characteristic Frequency (CF), and a measure of the cell's sensitivity to ITD, the Best Delay (BD). The current view of the MSO population is that it is a sort of *binaural display* [Joris and Yin, 2007]: showing at any time the activity of cells over an array of BDs and CFs. The position of the sound sources present is necessarily represented in this pattern of activities. The Jeffress model was formulated as a *peak code* [Jeffress, 1948]: a cell is the most active of the population provided that the ITD is equal to its BD. In other words ITD is represented in the BD of the most active cell of the

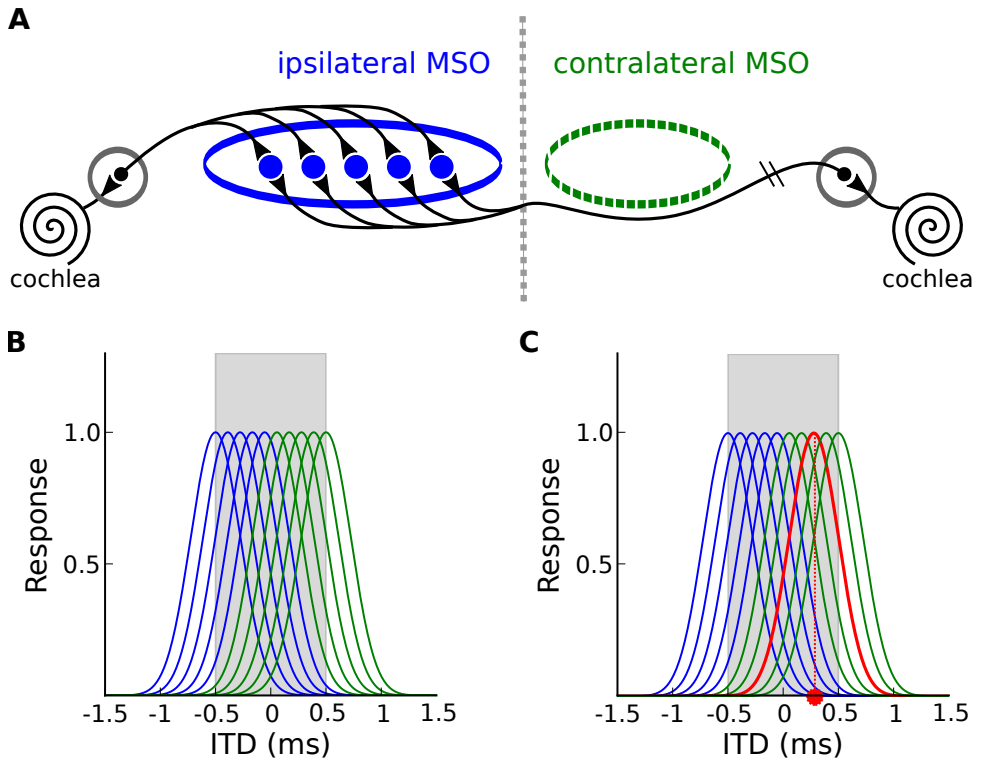

Figure 5.6: A depiction of the peak or labelled-line code for ITD. **A** Drawing of the situation, in each MSO (blue: ipsilateral, green: contralateral) cells have different BDs (materialized by different axonal length). **B** As a result, each neuron's tuning curve peaks for different ITDs, blue curves are the contralateral MSO tuning curves and green curves are the ipsilateral MSO tuning curves. Gray area is the physiological range, between -500 and 500 μs. **C** When a sound is presented, the most active neuron of the population (red tuning curve) signals the ITD in the input (red dot on the x-axis).

population. Peak codes are *place* codes in the sense that the identity of the cell (here the BD) contributes information to the code.

Recent electrophysiological data have raised a debate about the validity of this model, which lead to the introduction of the hemispheric model [McAlpine et al., 2001] which failed to reach consensual status. This code is very different from the peak code idea, for two important reasons. First it is a *population* code in the sense that the only relevant quantity is the summed activity of each MSO. A cell's BD is ultimately irrelevant to the subsequent system. This is to be opposed to *place* codes. The purpose of this section is to introduce both those models, and the arguments that have been put forward in this debate.

5.3.1 The Place code a la Jeffress

Jeffress and peak-coding

[Jeffress, 1948] originally formulated his model as "representing a temporal dimension [ITD]" in the "place of activation" in a population of neurons. It is not clear, however, whether the only relevant information was the identity of the most active neuron of the population. Yet, in one of its modern meaning, a Jeffress-type code is a peak code, as presented earlier. In this manuscript, a peak code (and thus Jeffress) stands for a code in which in any given frequency band, the most active celle of the population signals the ITD of the input.

The Jeffress model is also sometimes taken as predicting that there should be a systematic arrangement of axonal delay lines leading to binaural neurons. As a consequence of this, MSO neurons should be anatomically arranged by BD along a spatial dimension, therefore exhibiting a sort of *BD-topy*. It is necessary to emphasize that those were not explicit predictions by Jeffress, yet this interpretation has driven the field for years. Such properties are actually found in the avian NL (see e.g. [Konishi, 2003]), where the Jeffress peak-code is the consensual view. In the mammalian literature, the evidence for this kind of delay arrangement has proven unconclusive, despite a long seach. For example, many studies report that a gradient of axonal delays from the contralateral CN exists in the cat [Smith et al., 1993, Beckius et al., 1999, Oliver et al., 2003], but the resulting BD arrangement is *coarse at best* [Joris and Yin, 2007]. Furthermore, some argue that axonal morphology leading to MSO [Beckius et al., 1999] can not account for the complete range of observed BDs.

Despite the fact that predictions about a systematic arrangement of delay lines or an anatomical arrangement of BDs were not made explicit by Jeffress, lack of clear evidence in mammals in favor of those predictions was always a motive of suspicion about the Jeffress model. Yet, the peak code idea in itself does not make any specific assumptions as to those points.

The BD anomaly

Low-CF cells in the mammalian MSO sometimes exhibit unusually large BDs, which has long been a puzzle for the field [Phillips and Brugge, 1985]. Data acquired from small mammals, such as the gerbil and guinea pig [McAlpine et al., 2001, Pecka et al., 2008] consistently report neurons with BD $>$ 1ms. A cell with such a large BD, will never be maximally activated because ITDs are naturally constrained to the *physiological range* of the animal. The argument is that cells with large BDs would never be the most active of the population, because they will never be maximally activated. It must be noted outright that this is not necessarily true, because recorded MSO cells typically

have very heterogeneous maximal firing rates. Therefore, a neuron can be the most active of a population, without necessarily being presented with a sound at its BD, argument that will be developped later in this part.

The BD itself is not necessarily a well-defined quantity, for example we have observed in the previous section that it depended on the frequency content of the input. Furthermore, the BD is generally measured on the tone-delay curve obtained at BF [McAlpine et al., 2001]. This poses a number of problems, because BF also varies with the level of the input (as exposed in the previous section). All in all, the BD is not necessarily the best quantity to quantify the sensitivity of MSO neurons, because it only accounts for the response of the cell to a narrow subset of possible stimuli.

The notion of the *physiological range* can also be discussed, because it is often based on a crude simplification of the acoustic space. As I showed previously in Chapter 3, the physiological range is in general broader in low than in high frequencies. This can accout for part of the negative correlation between BD and CF, yet it falls short of explaining the unsually large BDs in low-frequencies.

Generalized place codes

A peak code only uses of the relative activities of the neurons in the population. This means that most of the information bore by the pattern of activities across the population is lost. Some models assume that the full pattern of activity across neurons contributes information to the ITD. Those are generalized *place codes*, in the sense that they do not lose the information about the cell BD, yet do not necessarily rely on the maximally active neuron (peak code).

Developments on such codes stem from the motivation of accounting quantitatively for the psychoacoustical localization performance of human subjects. More often than not, those models include some mechanism to deal with ILDs, but the present development will focus on the ITD coding mechanism. The first attempt at providing a prediction of subjective lateralization based on a model wherein binaural neurons represent the cross-correlation of the input signals was introduced in [Sayers and Cherry, 1957]. This has lead to many developments of binaural models, designed to be tested against psychoacoustical data, as reviewed in [Stern and Trahiotis, 1995], or [Colburn, 1996].

The influential "coincidence-counting" model [Colburn, 1973] used a model of AN firing alongside statistical analysis to show that the Jeffress scheme of representation of the cross-correlation of the neural signals is sufficient to account for the Just Noticeable Differences observed in human subjects. Developments on this model make the mapping from the activity of the binaural display neurons to the position of the sound

source more explicit. The "position-variable" model by [Stern Jr and Colburn, 1978] introduced the idea that the subjective lateral position of the sound source was encoded in the modes of the activity pattern in a single CF-band of the binaural display. In the case of a unimodal activity, this means that the centroid, along the BD dimension, of the activity maps to the lateral position of the sound source.

As discussed earlier, this model takes more information out of the activity pattern of the population of binaural cross-correlators than peak codes do. Those studies typically include additional mechanisms, the most prominent of which compensating for the effects of the ILDs. Another interesting development of the model proposed in [Stern Jr and Colburn, 1978] is that it includes a way to aggregate data across CF-bands according to the *straightness* of the activity, that is emphasizing CF bands that point to the same ITD (idea further developed in [Stern and Trahiotis, 1992, Trahiotis and Stern, 1994]). To account for the increased precision around the midline, a weighing emphasizing more central position can be included, as in the lateral-image model [Stern et al., 1988]. An interesting argument is made in [Fitzpatrick et al., 1997b], showing that the position variable model together with sharpening of ITD sensitivity along the auditory pathway achieves good acuity while recruiting a small amount of cells.

5.3.2 The hemispheric model

Origin and motivation for the hemispheric model

The orignal Hemispheric model was suggested by [Boehnke and Phillips, 1999] in the context of psychoacoustical studies. Because of the presence of large BDs in low frequency ITD sensitive neurons of the gerbil, an alternative model of the coding of space in the binaural display was devised. This model is a *population code*, usually termed the hemispheric model, or the two-channel model. In this coding scheme, ITD is coded in the relative summed activity of each hemisphere, that MSO neurons (or cortical area) in each side. The only relevant statistic of the binaural display is the sum of the activity of all neurons on one side [McAlpine et al., 2001, Grothe et al., 2010, Stecker et al., 2005]. This population activity is bell-shaped as well, [McAlpine et al., 2001], centered on BDs outside the physiological range, with a bandwidth such that the steepest part of the slope centers on zero. This means that on the physiological range, this population activity is an increasing function of the ITD, that can be inverted to decode the position of the sound source.

In fact, the position is read out of the difference (or ratio) of the population responses of two MSO populations, each most active to an ITD widely contralateral. Studies confirm that the performance of this decoder is correct [Lling et al., 2011, Lesica et al.,

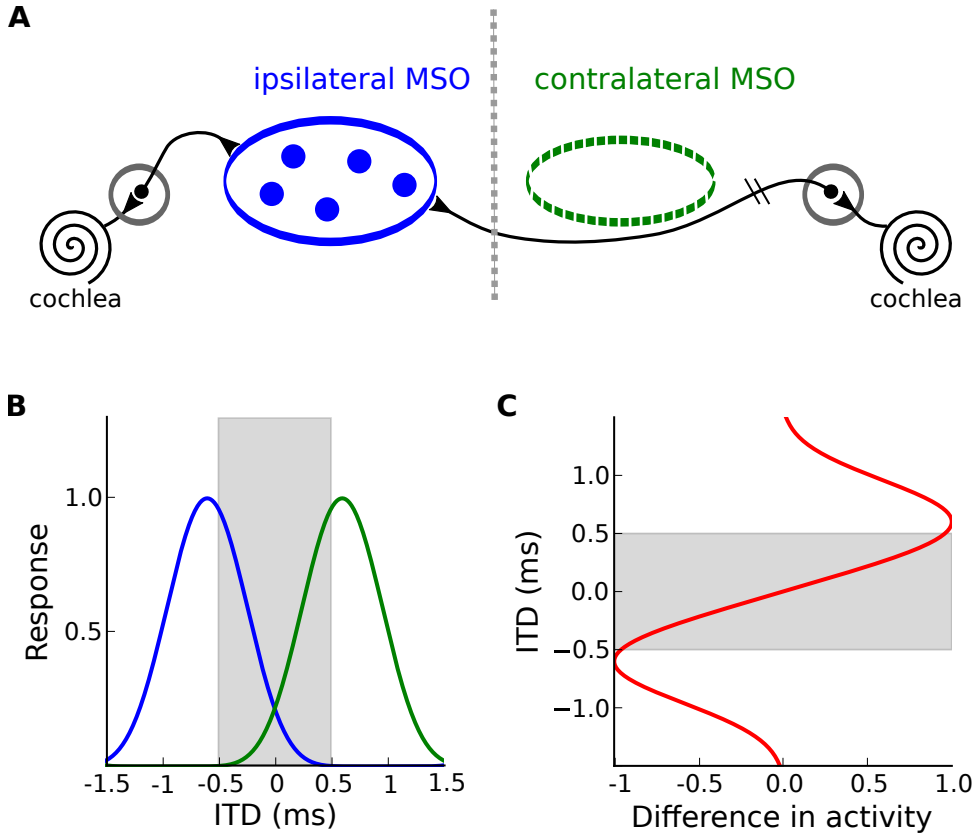

Figure 5.7: A depiction of the hemispheric code for ITD. **A** Drawing of the situation, in each MSO (blue: ipsilateral, green: contralateral) cells have different BDs. The population response (sum of all tuning curves in one MSO) has a single peak outside of the physiological range (**B**). Schematically, the neurons all have the same axonal projections in **A**. As a result, **B**, the population tuning curve peaks for a single ITD on each side, blue curve is the contralateral MSO and green curve is the ipsilateral MSO tuning curve. Gray area is the physiological range, between -500 and 500 μs. **C** When a sound is presented, the difference in the two activities is taken (x-dimension). It uniquely identifies the ITD in the input (y-dimension), provided that the ITD is in the physiological range.

2010]. Those studies suggest that the BD of a single neuron is irrelevant to the decoding of ITD, which is consistent with and "averaging" the hemispheric view.

A prominent fact in human behavioral studies is that human localization abilities in azimuth are better for sounds close to the mid-sagittal plane (see Chapter 9). The hemispheric model predicts that the sensitivity around the midline is optimal, hence consistent with behavioral results. This is was, as we saw, also accounted for by place codes such as the lateral-image model [Stern et al., 1988]. This makes this model a very popular framework to explain psychoacoustical data. The two-channel model is by construction maximally *sensitive* around the midline [McAlpine et al., 2001], and a recent study [Dingle et al., 2013] addressing this precise point shows that the two channel code predicts lateralization judgements without the need for an additional centered channel.

Criticism for the hemispheric model

Optimal code One line of criticism against the hemispheric model is to address the *optimality* arguments made in favor of the model. Indeed, the original study [McAlpine et al., 2001] claimed that the code was *optimal*. In fact, this model is optimally *sensitive* to ITD, that is will allow for optimal discrimination of two ITDs in the same stimulus. Yet, as argued in [Brette, 2010] the model is not optimal if the stimuli are allowed to vary along other dimensions, e.g. different levels, frequency contents. This arises from a misinterpretation of the notion of information.

In studies that adress the optimality of the hemispheric model, the decoder task generally is that of *discriminating* between stimuli with different ITDs [Lling et al., 2011, Lesica et al., 2010]. This is not, however, what the binaural system does, because it is concerned with absolute localization. Therefore the relevant performance is the accuracy (absolute location), rather than the acuity (dicrimination performance).

As an example, in the original study [McAlpine et al., 2001], the activity of an hemisphere *only* varies with the ITD of the stimulus. Specifically, this activity is not modulated by other dimensions of the stimulus: level, spectrum and background noise. If this activity is modulated by other dimensions of the stimulus, hence the inversion operation mentioned requires the prior knowledge of the other dimensions of the stimulus. The decoders presented in [Lling et al., 2011, Lesica et al., 2010] are therefore over-fitted, in the sense that they are performant on the same set of stimuli they are trained with, while being poor on novel data. The hemispheric decoder may be optimally *acute* to ITD, whilst being biased and *inaccurate* [Brette, 2010].

For a model to be *optimal*, it must be taken into account that ITD is only one out of many dimensions of the stimulus that affect the response of binaural neurons.

Therefore, optimality is out of the scope of models that rely on crude descriptions of the acoustical environment.

Low- and high- CF neurons An important remark as to the argument about the presence of large BDs in binaural cells is that it is generally not observed for mid-CFs, higher than a few hundred Hertz. In essence, most studies in mammalian MSO report that the BDs of cells above a certain CF fall within the physiological range (see, e.g. [Joris et al., 2006] for cat data). Therefore, the hemispheric model does not seem appropriate for this range of frequencies. An attempt to explain this discrepancy was pushed forward by the same [Harper and McAlpine, 2004], where it is shown that hemispheric-type and peak-type codes are optimal in different ranges of frequencies. Namely, because of phase-ambiguity, peak-type models are more appropriate in higher frequencies (or smaller heads). This also provides a functional explanation for the discrepancy observed in the coding of ITD in birds (barn owls) and mammals.

First, this study does not avoid the pitfall mentioned in the previous paragraph, as the set of possible stimuli considered consists in pure tones at different frequencies. This represents a quite narrow subset of the possible stimuli that an animal will encounter in his environment. Second, this study provides a unifying explanation of across-species differences in ITD tuning, it is actually not consistent with electrophysiological data. A good animal model is the chicken, because chickens and gerbils both use ITD cues on the same frequency range, while having similar-sized heads. According to [Harper and McAlpine, 2004], both those species should use the same ITD coding strategy. Yet, this was shown to be by [Kppl and Carr, 2008, Schnupp and Carr, 2009]. Indeed, the chicken's NL does not display large BDs in low frequencies, therefore is inconsistent with a hemispheric code. Rather, they are thought to be implementing a place code. On the other hand, gerbil MSO neurons are known to exhibit large BDs in low frequencies, and implement a hemispheric model [McAlpine et al., 2001]. Therefore functional constraints alone (or at least those considered in [Harper and McAlpine, 2004]) cannot account for the coding differences in different species.

Lesion studies Another line of criticism against the hemispheric model can be found in lesion studies. It has been found indeed that unilateral lesions of binaural structures (SOC) [Jenkins and Masterton, 1982] result in degradation of the localization only in the contralateral field. This is consistent with the widespread idea that sound source location is coded in the contralateral structures. More importantly, it is in direct contradiction with the hemispheric model, as ITD is extracted by comparing the activities of the binaural structures on both sides.

5.3.3 Discussion

Frequency-dependence of BD

As wa have seen in the previous section, the BD of ITD sensitive cells varies with a number of factors. In the cases where it is measured on the tone-frequency curve at BF, it is dependent on BF therefore on stimulus level. Furthermore, the BD also varies with the frequency of the input tone. Therefore, the BD is an incomplete measure of the tuning of a binaural cell to ITD. That is, the response of cells is much more heterogeneous than BD data seem to indicate: they vary with more dimensions of the stimulus.

Modeling arguments in favor of the hemispheric model [McAlpine et al., 2001,Harper and McAlpine, 2004,Lesica et al., 2010] also rely on a frequency-independent description of the BD. Furthermore, the response of any given cell is considered constant for all frequencies. In conclusion, those models cannot account for the precise frequency-dependent response of MSO neurons.

An alternative hypothesis is that neurons are not sensitive to ITDs, as summarized by the BD, but rather to *position* I have shown that ITDs are variable across small frequency ranges, could it be that encode these frequency-dependent ITD? In this case, the relevant characteristic is no longer a Best Delay, it is a Best Position. In the next Chapter 7 we explore the hypothesis that BDs across freuqency can be predicted on the basis of acoustical measurements alone.

Place and population codes

At the core of the arguments against the Jeffress model is the presence of large BDs in the low frequencies. This, we argue, is not a direct contradiction with the idea of a place code. Indeed, a neuron can be the most active *of the population* for a given ITD, while being more active at another ITD. In essence, a neuron is not necessarily the most active neuron of the population when the stimulus is presented at its BD, and a neuron is not necessarily the most active of a population when the ITD is its BD. This is a confusion between the measure of BD which is the ITD that leads to the highest response across ITDs and the place code readout which is the most active neuron of a population presented with a given ITD. Single unit recordings are measurements of one cell across stimuli, and care should be taken when generalizing those results to populations of cell. This precise idea is discussed in Chapter 8, in which we present an alternative place code that exemplifies this phenomenon.

5.A A simple ITD sensitive model

We us a leaky integrate and fire neuron with unitless membrane potential. The dynamics of the membrane potential (here adimensional) are defined by the linear equation:

$$\tau_m \frac{d}{dt} V_m = -V_m + \sigma\xi(t)$$

Thanks to the linearity, the EPSPs have fixed exponential shape and sum exactly. The neuron spikes iff $V_m(t) > V_{th} = 1$ and the membrane potential is subsequently reset to $V_r = 0$. The time constant of the neuron is picked as $\tau_m = 1ms$. The noise $\xi(t)$ is a gaussian white noise term, with a sigma value of 0.1.

The inputs to the neuron are the output of a simple auditory periphery model. The output of a gammatone filter with center frequency 600 Hz is rectified and compressed, $max(H * s, 0)^{\frac{1}{3}}$, and fed to an integrate and fire neuron at each side that model the AN fiber. Those AN fiber inputs are then fed to the neuron with a zero delay on the left and BD delay on the right.

Simulations are performed for 100 ms long white noise stimuli. The number of output spikes of the binaural neuron is recorded for each ITD in the input, mimicking an noise delay measure. All simulations were performed using the Brian simulator [Goodman and Brette, 2009] and the Brian Hears toolbox [Fontaine et al., 2011].

This can be set by setting the monaural weights to $w = 0.55$, which means that each spike in the monaural input creates an increase of 0.55 of the membrane potential. On the opposite, two synchronous spikes are enough to trigger an output spike from the binaural neuron ($1.1 > 1$).

From these considerations, one can work out a combination of delays that makes the neuron spike for a given ITD. Indeed, if the cell's neural delays are such that:

$$-ITD - \tau_r = -\tau_l \Rightarrow ITD = \tau_l - \tau_r$$

Then the neuron receives synchronous input signals, hence produces spikes.

5.B Behavioral studies in animals

Animal models for the electrophysiological study of sound source lcoalization are carefully chosen according to different criteria. Essentially there are two criteria, first the match between the animal's hearing range and that of humans. This is the case mostly

in mammals (because humans are mammals), e.g. cat, guinea pig, chinchilla, gerbil, that hear on roughly the same range as humans. A second criterium is the behavioral ability to localize sound sources. For this reason, birds such as the barn owls are often used because they strongly rely on acoustical cues to hunt. Yet, because behavioral data is hard to gather in animals, the extent to which usual lab animals are really able to localize sound sources, in this section we review the literature on behavioral studies with animals, that provide information as to the localization abilities of those different animal species.

5.B.1 Minimum Audible Angles in animals

The most common experimental paradigm for assessing the localization performance of animals is the measure of Minimum Audible Angles (MAA). This is a threshold measure (generally defined on a hit rate of 75%) for the discrimination of two free-field sources symmetrically placed around the midline [Casseday and Neff, 1973].

This paradigm was applied by Heffner to a wide variety of common lab animals, and less common species (blind mole rats, cattle, etc...). In cats, and for reasonably long ($>$ 40ms) broadband sounds, the MAA is measured at 5° [Casseday and Neff, 1973, Heffner and Heffner, 1988b]. This value is similar to the Macaque monkey and relatively close to human psychoacoustics [Blauert, 1997]. Furthermore, in the cat study [Casseday and Neff, 1973] using pure tones, the sound localization ability does not seem to degrade over the range of frequencies where ITD is used (below 4 kHz). Gerbils, however, seem to have especially poor localization abilities [Heffner and Heffner, 1988a]. This result was recently confirmed [Carney et al., 2011], which questions the relevance of data gathered in the context of sound-localization studies in this species.

Similar experiments were performed on barn owls, e.g. [Knudsen et al., 1979] show that the MAA is of about 2°. Absolute localization judgement precision in barn owls is also greater around the midline [Knudsen et al., 1979]. In [Fischer and Pea, 2011], the authors argue that barn owl have a systematic bias to judge position closer to the midline. This behavior can be explained by assuming Bayesian model that puts a prior on central positions: representing the fact that the animal assumes that sound sources are close to the midline.

5.B.2 Across-species prediction of localization precision

Another interesting point is made by Heffner, who is interested in the evolutionary pressure constraining sound localization abilities. In a thorough across-species study, [Heffner and Heffner, 1992] show that the best predictor of a mammal's ability to

localize is the width of the *width of the field of best vision*. The field of best vision is defined using the spatial distribution of retinal ganglion cells, and is thus quite small in mammals with foveas. More precisely, the localization threshold increases with the width of the field of best vision. Surprisingly, this correlation is stronger than other studied dimensions, e.g. interaural distance, visual precision, This results suggests that sound localization is essentially useful for directing gaze for visual inspection of a sound source.

The hypothesis that vision is essential to proper auditory localization is emphasized by the fact that early-blind humans have auditory localization deficits [Zwiers et al., 2001]. Furthermore, current data overall agree on the influence of sounds on visual search performance. Presentation of auditory stimuli, simultaneously to visual targets can help direct gaze [Doyle and Snowden, 2001]. Yet, the spatial position of the sound does not seem to influence the accuracy of gaze direction [Doyle and Snowden, 2001, Zou et al., 2012]. In other words, it is the mere presence of a sound, not the position of the sound that helps directing gaze in [Zou et al., 2012, Doyle and Snowden, 2001]. Therefore, the hypothesis that sound source separation helps direct visual gaze is not completely supported by the human behavioral data.

Chapter 6

Local variations of ITD: The fiction of the Best Delay

Contents

This chapter is the result of a collaboration with Philip Joris, who kindly provided us with binaural-beat data from single IC cells. All the data was therefore gathered by Philip Joris, while I performed the analysis together with Bertrand Fontaine and Romain Brette. The work presented in this chapter is in submission.

The relative time it takes for a sound to travel from its source to the two ears varies systematically with direction. In the auditory system of many animals, there are neurons sensitive to this interaural delay. It is thought that their firing rate provides a

measure of this delay, from which sound direction can be inferred. However, we find by analyzing in vivo responses of 200 neurons in cat that tuning to delay is highly dependent on sound frequency. On the other hand, we show that interaural delays measured acoustically at the ears of a cat also depend on frequency, because of diffraction of sounds by the head. We demonstrate that the frequency-dependence of neural tuning to delays matches that found in acoustical recordings. We propose that binaural neurons are tuned to space rather than to time, and that sensitivity to interaural delay derives from spatial tuning.

6.1 Introduction

Acoustical waves produced by a sound source reach the two ears at slightly difference times depending on its direction. Interaural time differences (ITDs) are used by many species to localize sounds in the horizontal plane. In mammals, neurons in the medial superior olive, just three synapses away from the ear, are sensitive both to ITD and to sound frequency. It is thought that their activity encodes ITD in each frequency band. They project to neurons in the inferior colliculus (IC), which inherit these properties. The firing rate of these cells is strongly modulated by the ITD of a tone presented binaurally through earphones (Fig. 6.1, **A**). The "best delay" (BD) is defined as the ITD where the cell fires at highest rate. Equivalently, the firing rate represents the similarity between the sounds ITD and the cells BD, and therefore may form the basis of a neural code for ITD. However, the BD of a cell is not a fixed quantity: it can vary with sound frequency. Fig. 6.1, **A** shows normalized ITD selectivity curves of a neuron in the IC of a cat for tones at frequencies between 800 Hz and 1400 Hz, where the cell responds strongly. The BD varies between 400 and 100 μs over this frequency range (Fig. 6.1 **B**). We examined such responses for cells tuned to characteristic frequency (CF) between 97 and 3341 Hz. We found that the BD of a cell spanned on average an interval of 128 μs (+- 240 μs) over the relevant frequency range (Fig. 6.1 **C**). This extent is large, considering that the maximum natural ITD in cats is about 350 μs and that cats can discriminate ITD differences of just 20 μs. This observation challenges the view that these neurons encode the ITD of sounds, since the code seems to change depending on the nature of the sound.

Frequency-dependence of BD has been observed in many species. To quantify it, the standard method consists in calculating a linear regression between the best phase (BP = BD times frequency) and the tones frequency (Fig. 6.1 **D**). If the BD is constant, then the BP depends linearly on frequency, i.e., the intercept is zero. Thus this intercept, called the characteristic phase (CP), quantifies the frequency-dependence of the BD. The slope of the regression is called the characteristic delay (CD). In the special case when CP=0, the BD is constant and equals the CD. Linear circular fits were highly significant for almost all our cells (Fig. 6.5). The distribution of CPs, wrapped within one cycle (-0.5 to 0.5), is shown for all 203 cells on Fig. 6.1 **E**. As observed in other species, many cells have CPs near zero, but the distribution is broad, meaning that many cells do not have a fixed BD. There is also a small bias for positive CPs. Finally the distribution is unimodal, suggesting that the differences in CPs are not due to the cells being of different types. The CDs are mainly positive (corresponding to contralateral leading sounds) and mostly within the natural range of 350 μs, but a

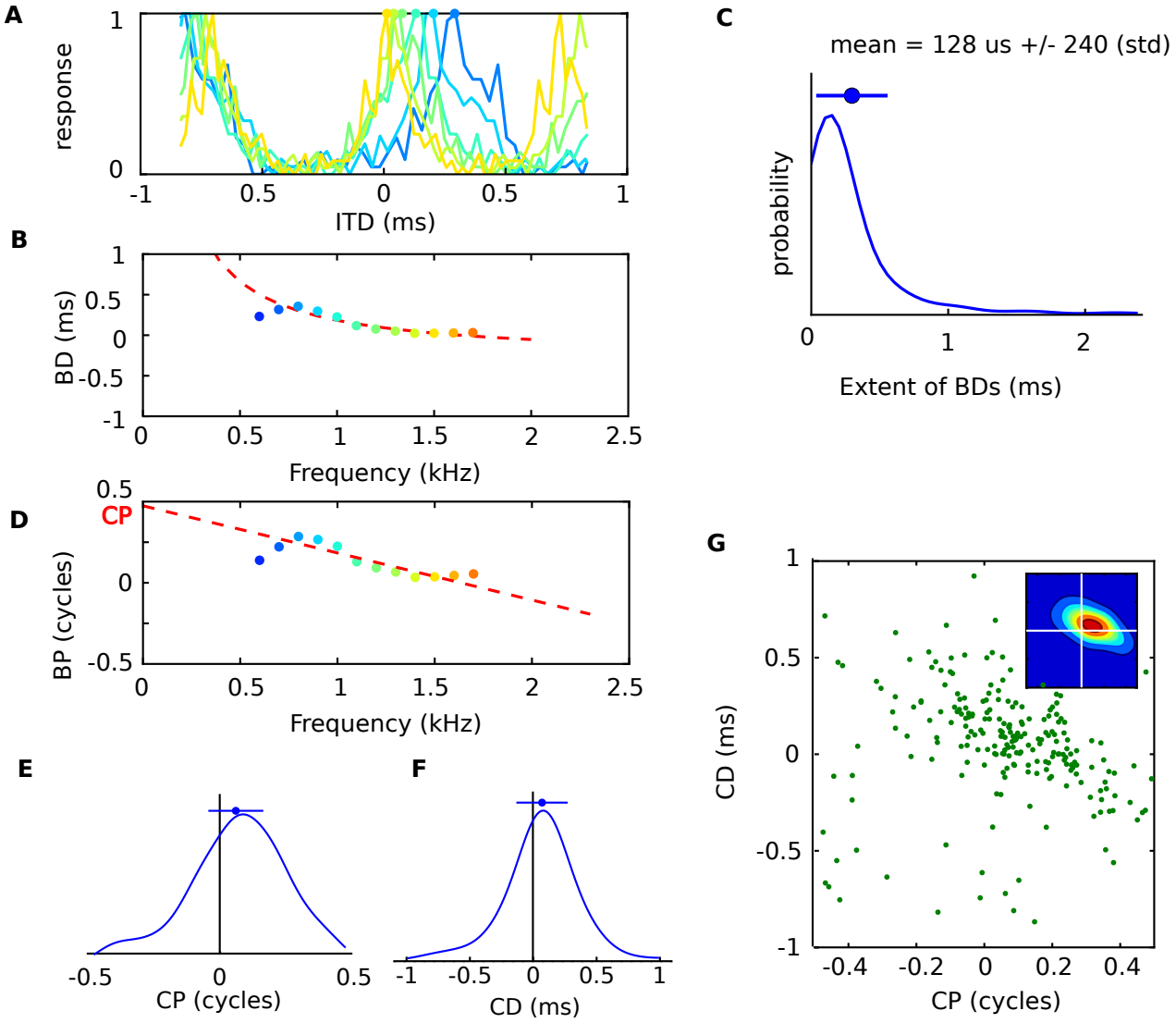

Figure 6.1: Frequency-dependence of best delays. **A** Normalized firing rate vs. ITD for one neuron, with tones between 600 Hz (blue) and 2000 Hz (orange). **B** Best delay (BD) vs. frequency for the same cell, and hyperbolic fit (dashed red curve). **C** Distribution of BD range across all 186 cells. **D** Best phase (BP) vs. frequency and linear regression. The characteristic phase (CP) is the intercept, the characteristic delay (CD) is the slope. **E** Distribution of CP across all cells. **F** Distribution of CD. **G** CD vs. CP across all cells (inset: color-coded density).

large minority of cells have negative CDs and a smaller minority have CDs larger than 350 μs (Fig. 6.1 **F**). Most intriguingly, as observed in other species, CDs are negatively correlated with CPs (Fig. 6.1 **G**). We checked with bootstrap analysis that this negative correlation is not due to measurement artifacts (Fig. 6.9).

These features are difficult to comprehend if cells are assumed to encode frequency-independent ITDs. However, it has been demonstrated with physical models and acoustical measurements that for a given source direction, ITD varies with frequency, because of sound diffraction by the head. This variation can be fully quantified by analyzing head-related transfer functions (HRTFs), which measure the acoustical filtering of the head for sources at various positions (Fig. 6.2, **A**). Fig. 6.2, **B** shows the variation of ITD with frequency for three source directions in an anesthetized cat. These variations are consistent with previous acoustical measurements in cats (Fig. 6.7, **A**). To make sure that these variations are generic, we also measured HRTFs for a taxidermist model of a cat in a natural posture (Fig. 6.7, **B**), which we compared with numerically computed HRTFs from a 3D model of the same cat (Fig. 6.7, **C**-**D**), so as to make sure that these variations were not due to artifacts of acoustical recordings. In all cases, similar variations of ITD with frequency were observed. A simple spherical head model also displays this phenomenon (Fig. 6.7, **E**), particularly when a ground is included (Fig. 6.7, **F**). If a cell were tuned to a constant BD, then its best azimuth (azimuth at which it fires most) would vary with frequency (Fig. 6.2, **C**, top). On the other hand, if the cell had a constant best azimuth, its BD would vary with frequency to match the variation of ITD with frequency at that position (Fig. 6.2, **C**, bottom). Therefore, we tested whether the features seen in the electrophysiological data could be explained by the hypothesis that cells are tuned to an azimuth rather than to an ITD.

6.2 Results

For each source direction in the horizontal plane, we examined the variation of interaural phase difference (IPD) with frequency using linear regressions, in the same way as we previously examined the variation of BP with frequency in cells (Fig. 6.1 **D**). For example, at 70°(Fig. 6.2, **D**), the linear regression gives two different results in low frequency and in high frequency. Acoustically, the slope of these regressions corresponds to the group ITD (ITD_g), defined as the derivative of IPD with respect to frequency. It corresponds to the ITD of the sound envelope (Fig. 6.10, **A**-**B**). When there is no sound diffraction, the ITD of the fine structure (IPD divided by frequency, or ITD_p) equals the group ITD, and the linear regression intercepts the vertical axis at zero cycle. But these two ITDs are not always equal with sound diffraction, as is seen in Fig. 6.2, **E**

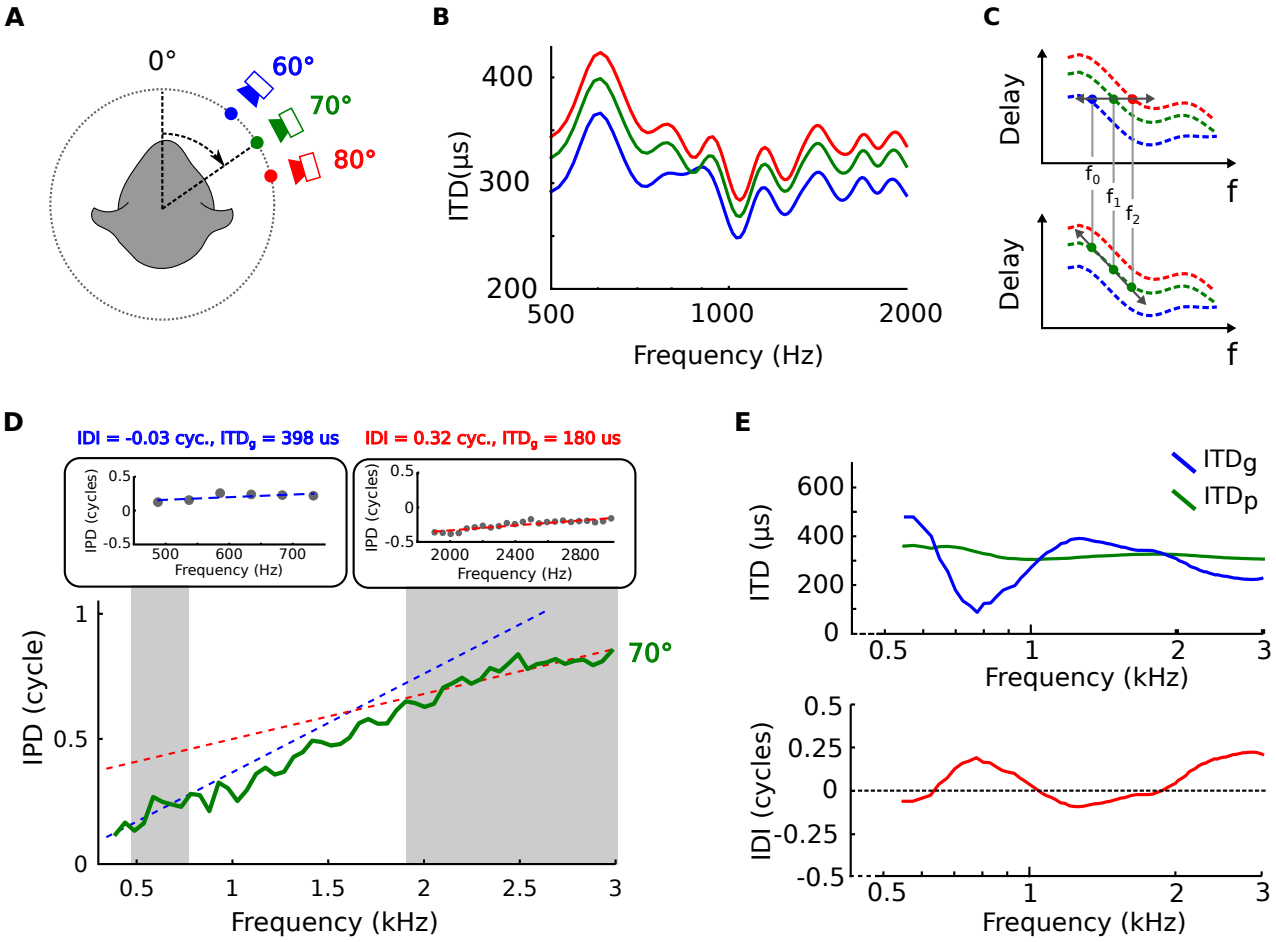

Figure 6.2: Frequency-dependence of ITD. **A** Head-related transfer functions are measured binaurally at different source directions. **B** ITD vs. frequency in cat at 60, 70 and 80°. **C** A neuron with fixed best delay (top) is tuned to different azimuths depending on frequency, while a neuron with fixed azimuth tuning (bottom) has a frequency-dependent best-delay. **D** IPD vs. frequency and linear regressions in two frequency ranges, centered around 600 Hz (top left) and 2400 Hz (top right). The slope is the group ITD and the intercept is defined as the interaural diffraction index (IDI). **E** Top, group and phase ITD vs. frequency for a source at 70°. Bottom, IDIvs. frequency.

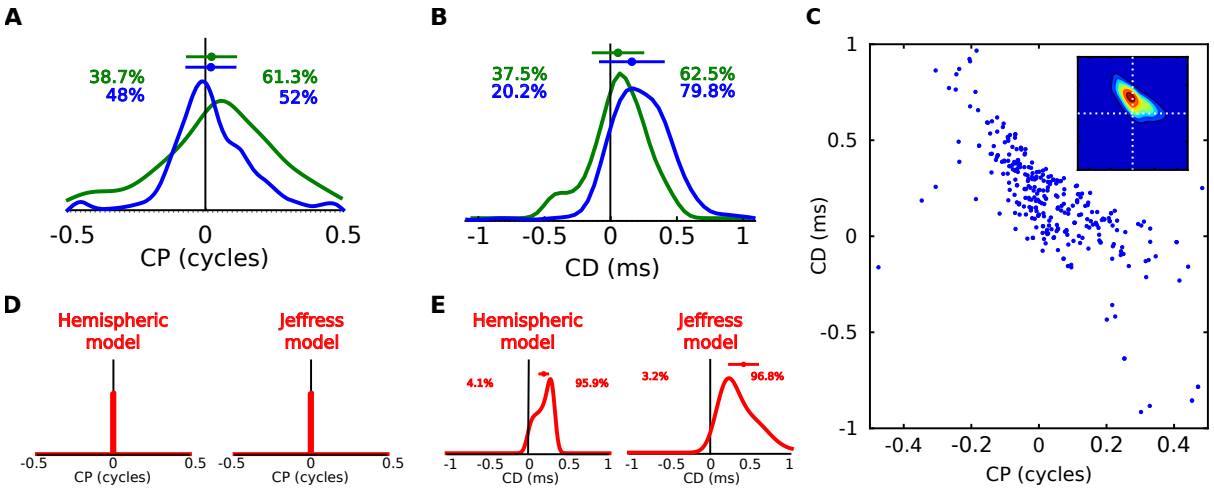

Figure 6.3: Model predictions. **A** Distribution of CP in the cells (green) and in the model cells based on acoustical measurements (blue). **B** Distribution of CP in two models with fixed BD, hemispheric and Jeffress model. **C** Distribution of CD in recorded and model cells. **D** Distribution of CD in hemispheric and Jeffress model. **E** CD vs. CP in the model cells (inset: color-coded density).

(top). For this reason, we call the intercept "interaural diffraction index" (IDI). It equals the difference between $ITD_p$and ITD_g, in cycles of the frequency (Fig. 6.2, **E**, bottom; Fig. 6.10, **C**). In our cat HRTFs, $ITD_g$and ITD_phave different distributions (Fig. 6.10, **D**) and IDIare broadly distributed around zero (Fig. 6.10, **E**).

If a cell were tuned to a constant azimuth, then the variation of BP with frequency, measured electrophysiologically, should match the variation of IPD with frequency at the preferred azimuth, measured acoustically in the HRTFs. Therefore its CD should equal the $ITD_g$for that azimuth, and its CP should equal the IDI. We generated predictions by considering that azimuth is uniformly represented across cells in the contralateral hemifield (front and back), and that characteristic frequencies (CF) are distributed as in the recorded cells. We obtained a set of model cells characterized by a CD and CP extracted from the acoustical data, which we compared with the electrophysiological measurements (Fig. 6.3). Quantitatively, the distributions of CD and CP in the model cells depend on the chosen distribution of best azimuths, which could alternatively be uniform over directions in the front only (Fig. 6.8, **A**), or inferred from the electrophysiological recordings (Fig. 6.8, **B**), or biased towards the side (Fig. 6.8, **C**) or the center (Fig. 6.8, **D**), as suggested in the barn owl. Despite the quantitative differences, the same qualitative features remain. In particular, in agreement with measured data, the model predicts that the distribution of CP is unimodal and broad, with a small but significant positive bias (Fig. 6.3, **A**). In contrast, a model with constant BD would predict that all CPs are zero (Fig. 6.3, **B**). As in the measured cells, the predicted

CDs are mainly positive and mostly within 350 μs, but with a minority of cells with negative CDs and with large BDs (Fig. 6.3, **C**). In contrast, a model based on a peak code for ITD predicts that almost all CDs should be positive and smaller than the maximal natural ITD of 350 μs (Fig. 6.3, **D**, left). Another model in which BDs are 1/8th the characteristic period (1/CF) also predicts that most cells should have positive CD (Fig. 6.3, **D**, right); an alternative interpretation of this model is that cells have a constant BP of 1/8th of cycle, meaning that CP = 1/8 and CD = 0 for all cells, which is also not in agreement with the data.

Finally, the model based on HRTFs is the only one to correctly predict that CD is inversely correlated with CP (Fig. 6.3, **C**). This non-trivial feature can be explained by the variation of ITD with frequency, in a simple spherical head model (Fig. 6.4, **A**). In this model, ITDs are larger at low frequency than at high frequency (Fig. 6.4, **B**; Kuhn). More variations are introduced by the complex shape of head and body of real animals (Fig. 6.7), and also by early reflections on the ground (Fig. 6.4, **B**, green; [Gourevitch and Brette, 2012]). As a result of these variations, the IPD vs. frequency curve is non-linear (Fig. 6.4, **C**). When a tangent is moved along this curve (dashed lines), the slope (ITD_g) decreases at the same time as the intercept (IDI) increases. It follows that $ITD_g$and IDIare inversely correlated for a given position (Fig. 6.4, **D**).

6.3 Discussion

A number of studies have shown that there is a strong negative correlation between BD and best frequency (BF). We found the same trend in our data, with a slope of -560 μs/kHz (Fig. 6.6, **A**). However, since BD can vary widely within the bandwidth of a binaural cell, this relationship quantitatively depends on the choice of definitions. The BD can also be calculated as the fixed interaural delay that gives the best fit to the BP vs. frequency relationship. In this case, the slope of the linear regression with BF is -329 μs/Hz (Fig. 6.6, **B**). Another slope value, -181 μs/kHz, is found when the linear regression is calculated against characteristic frequency (CF), which is generally higher than BF (Fig. 6.6, **C**). Spearman coefficient, which quantifies nonlinear correlation, is also smaller (ρ = -0.22 vs. -0.40). Finally, we can select the pair of HRTFs for source directions in the frontal horizontal plane that gives the best fit to the BP vs. frequency relationship for each cell (Fig. 6.6, **D**). We call the corresponding azimuth is the best azimuth (BA). In this case, the slope of regression with CF is about 12°/kHz, with a small Spearman coefficient ($\rho = -0.16$). Previous measurements in free field did not show strong correlations either. In Fig. 6.6, **E**, we split the cells into two equal groups with CF¡804 Hz and CF¿804 Hz. It appears that the distribution of BAs is similar in

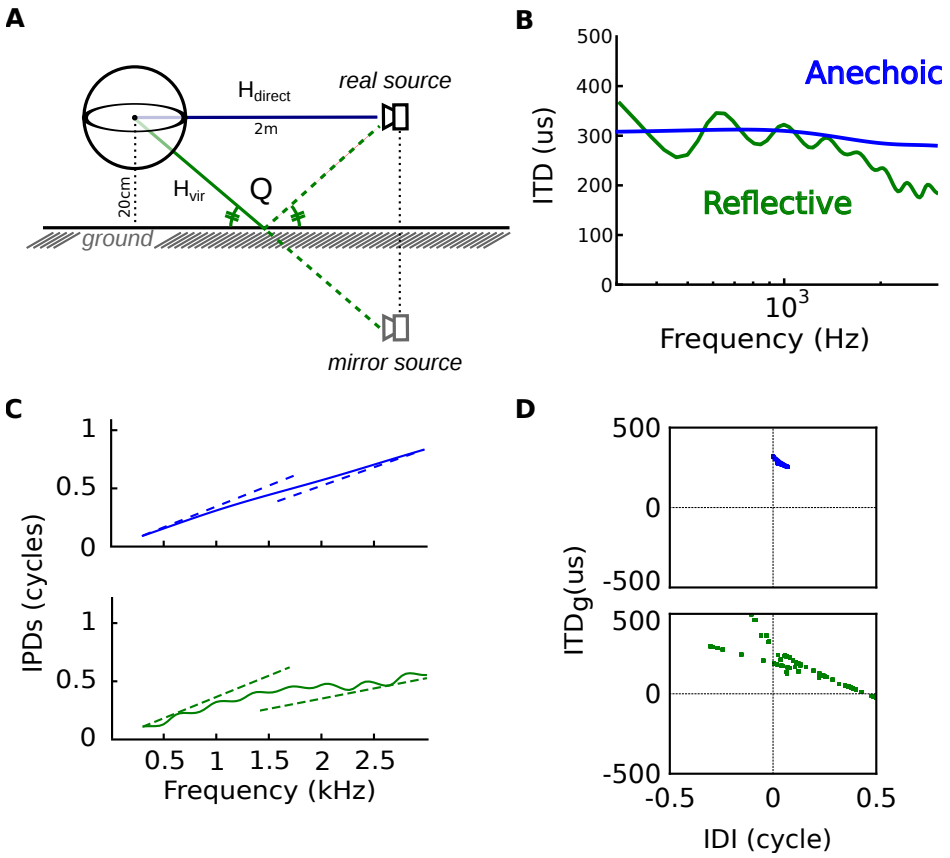

Figure 6.4: Theoretical explanation of inverse CP-CD relationship. **A** A spherical head model with a ground reflection. **B** ITD vs. frequency in the spherical model for a source at 70°, with (green) and without (blue) ground reflection. **C** IPD vs. frequency in the two cases. **D** Group ITD vs. IDI, corresponding to predicted CD vs. CP.

the two groups.

These results support the hypothesis that ITD-sensitive cells in the cat are tuned to a specific location, or more generally to a feature in the interaural phase spectrum of natural sounds, rather than to a fixed ITD. Since the frequency-dependence of ITD is a general feature due to sound diffraction, it should also apply to other species. We applied our analysis to barn owls HRTFs (Fig. 6.11, **A**-**C**), which use ITDs in high frequency (2-10 kHz), and we also found some variability in predicted CP. This variability was broader in human HRTFs (Fig. 6.11, **D**-**F**), which suggests that the frequency-dependence of ITDs may have behavioral consequences.

Finally, by which mechanism can the best delay of cell depend on sound frequency? This observation implies that the internal source of delay is not only a conduction delay. Frequency-dependent best delays can be produced by small mismatches in the CFs of the monaural inputs to a binaural cell, in addition to conduction delay mismatches. The binaural cell then responds preferentially when the mismatch in its monaural filters corresponds to the mismatch in the directional acoustical filters, a principle that can be used to accurately estimate sound location with realistic acoustics. This requires precise correspondence between monaural filters, but such correspondence has been demonstrated in the barn owl, and can be obtained with activity-driven plasticity. Therefore we suggest that, rather than measuring timing differences across the ears, ITD-sensitive cells are tuned to complex acoustical features of natural ecological environments, which are acquired during development

6.A Electrophysiological recordings

All electrophysiological recordings were performed by Philip Joris in K. U. Leuven., according to the method presented in [Joris, 2003]. Single-cells in the cat IC central nucleus were measured in response to binaural beats at different frequencies. All neurons presented in this study were recorded at 60 dB SPL. Characteristic frequency (CF) (frequency of lowest threshold) was determined with a threshold tracking algorithm to contra and/or binaural stimulation. All procedures were approved by the K. U. Leuven Ethics Committee for Animal Experiments and were in accor- dance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals.

6.A.1 Analysis of electrophysiological data

Calculation of CP and CD

The characteristic Phase and delays are respectively defined as the [Yin and Kuwada, 1983] intercepts and phase of a circular linear fit of the BP spectrum of the cell. The computation was done here according to [Lling et al., 2011].

The regression operation is not linear because a circular variable (BP, a phase in cycles) is regressed against a linear variable (frequency). Hence, the fit is obtained by minimizing a circular distance to the affine fit:

$$BP(f) \approx CP + CD \times f \ [1 \ cycle]$$

The BP spectrum of the cell consists of measurements of BP(f) at different frequency points. This represents an average of 9.5 points per cell (std = 3.96), and cells with less than 5 measured frequency points were discarded.Those frequency points were in general not uniformly spaced in frequency. Moreover the cell might have a sensitivity to IPDs in tones at some frequency and not others. Hence the response of the cell as a function of IPD at all frequencies was tested for uniformity using a Rayleigh test. Frequency points where the null hypothesis of uniformity couldnt be rejected ($p < 0.001$). Additionally, we weigh each BP measurement by the reliability of the response. This is usually done [Yin and Kuwada, 1983] by using the product of the discharge rate and the Vector Strength (VS) of the response of the cell. This quantity is known as the sync rate $SR(f) = r * vs$.

This quantity can be for example used to reconstruct an approximate noise-delay curve (the composite curve) using tone-delay curves [Yin and Kuwada, 1983].

Finally we are faced with solving the following non-linear least square problem:

$$(CP, CD) = argmin_{CP,CD} \sum_f ||SR(f)BP(f) - (CP + CDf)||^2$$

Where the norm used here is a circular norm:

$$||x||^2 = \frac{1 - cos(2\pi x)}{2}$$

The fit is obtained in a two-steps process: first the parameter (CP, CD) space is uniformly sampled (CP in the full range, CD between 2ms and -2 ms). This yields a distance per point. The best-fitting point is used as a seed to the second part of the fitting procedure. In this second part a gradient descent approach is used to solve the least square problem.

The reason for such a two steps fit is that the energy space has a lot of local minima in which the gradient descent procedure can get caught. This approach ensured that the computed CP and CD be the same, and was better suited to reconciling the simulated to the measured cell results. Also, not using a linear fitting procedure allows one to circumvent the unwrapping of the IPD pattern, which is in general an unreliable operation.

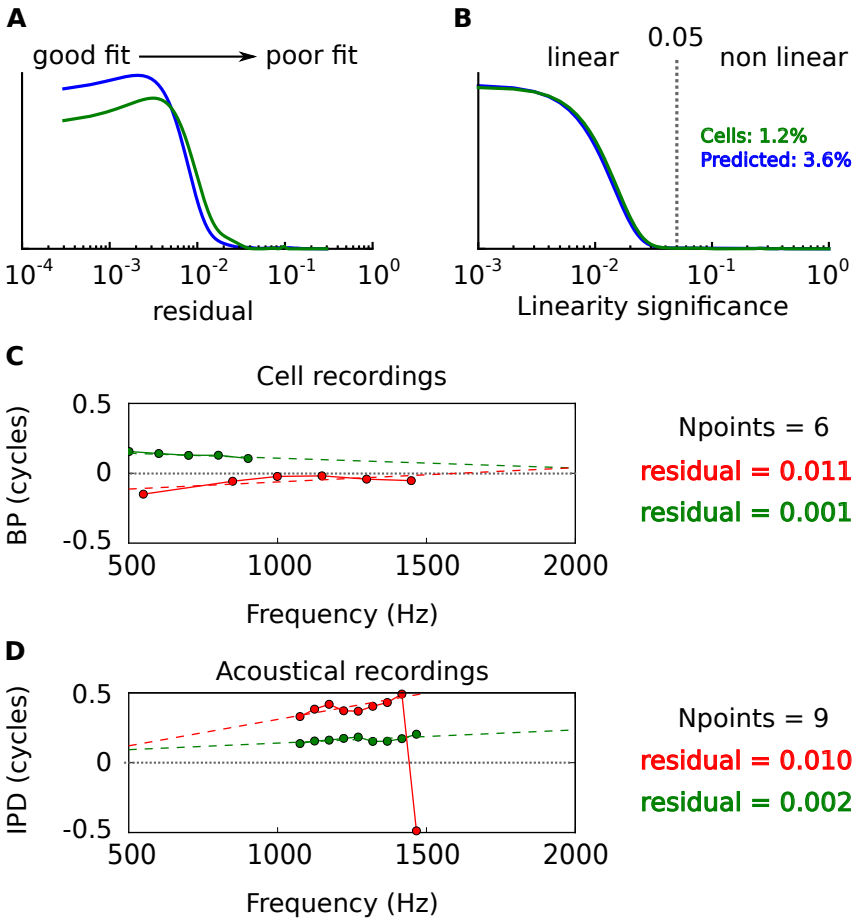

Figure 6.5: Linearity of best phase (BP) vs. frequency curves. **A** Distribution of residual error in linear fits of BP vs. frequency in cells (green) and in acoustical predictions (blue). **B** Statistical significance of linear fits. Cells are selected for further analysis when $p < 0.05$. **C** BP vs. frequency and linear regressions for two cells with the same number of data points (different frequencies) and different residual errors. **D** Same as **C**, for acoustical recordings.

The quality of the circular linear regression (that is, the fit of CP and CD) can be assessed by measuring the distance of the BP(f) spectrum to the regression line. This is in general known as the residual error. In our case this is a circular root mean square measure defined as:

$$residual(BP(f), CP, CD) = \sqrt{\frac{1}{N}\sum_f ||SR(f)BP(f) - (CP + CDf)||^2}$$

This quantity measures the adequation of the circular linear fit of the BP(f) data. To test whether a cell had a linear BP(f) spectrum, we used the same approach as

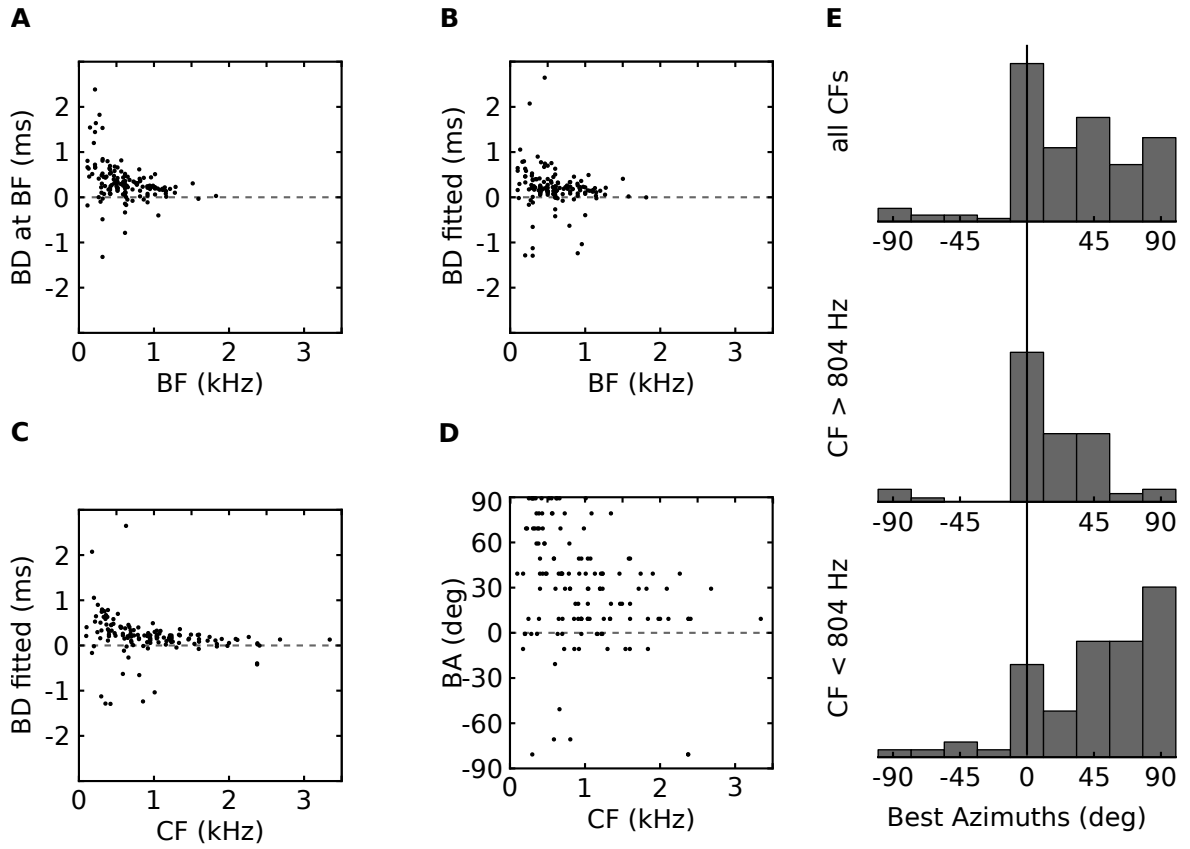

Figure 6.6: Distribution of best delays (BD) and best azimuths (BA). **A** BD measured at best frequency (BF) vs. best frequency (BF). **B** BD of the linear fit to BP vs. frequency curves, plotted against BF. **C** BD of linear fits vs. characteristic frequency (CF). Note that CF is generally larger than BF. **D** BA of the fit of BP vs. frequency curves to HRTF, plotted against CF. **E** Distribution of BA over all cells (top), over high frequency cells (middle, CF greater than median CF), and over low frequency cells (bottom, CF lower than median CF).

in [Yin and Kuwada, 1983], Fig. 6.5. We want to reject the null hypothesis that the cells have uniformly distributed BP(f) at each frequency points between 0.5 and 0.5 cycles, using this residual measure as a statistic. We perform simulations of cells under the null hypothesis with variable degrees of freedom (i.e. number of measured frequency points), perform the circular linear fit and measure the residual. This in turns yield a probability distribution of observing a given residual under the null hypothesis for different degrees of freedom. The cumulative of that distribution gives the linearity significance measure (Fig. 6.5, **B**). The null hypothesis could not be rejected in only 1.2% of the case for the measured cells.

Best delay, best frequency, characteristic frequency, alternative best delay

The Best Frequency of the cell was measured as the frequency of the tone at which the cell responds maximally when presented at 0 ITD. This is, unlike the Characteristic Frequency, variable with level, because in general a cell has different BFs at different stimuli level. The characteristic frequency is measured as the frequency at which the cell has a response significantly above noise for the lowest level. It can be understood as the minimal observable BF in that it is akin to the BF measured at the lowest level that the cell responds to.

The Best Delay is usually measured as the ITD that elicits the highest response in a cell, when presented in a tone at the Best Frequency (BF). This means it can be computed out of our binaural beat data, by converting the BP at the BF to a time:

$$BD_{BF} = \frac{BP(BF)}{BF}$$

This has been shown to be the same value as the maximum of the noise delay curve [Yin and Kuwada, 1983]. Yet, as it depends on BF, it in turn depends on the level of the stimuli presented.

We also measured BDs at CF using the formula:

$$BD_{CF} = \frac{BP(CF)}{CF}$$

Which is a measure that does not depend on the sound level of the stimulus presentation. Because some ITD sensitive cells display ITD sensitivity only at frequencies lower than CF (in the tail of the cochlear filter), sometimes binaural beats measurements werent available around the CF which prevented us from computing BP(CF). Those cells were excluded from the analysis in Fig. 6.6.

An alternative measure of BD is proposed here, that does not rely on measurement of BP at a single frequency, but rather the complete BP(f) spectrum. If the cell has a constant BD, then we show that the BP(f) spectrum is linear. Consequently, the best possible constant-BD representation of a cell is in fact given by the circular-linear fit

$$BP(f) \approx BD_{fit} f \ \ [1 \ cycle]$$

That is the same fitting procedure as in the CP/CD analysis, except that CP is constrained to be zero. This BD measure we term BD_{fit}. It was measured exactly in the same way as for the CP/CD analysis, for the same frequency points etc...

6.B Acoustical measurements

The HRTF data used for the main figures of the paper was already published by another group [Tollin and Koka, 2009b]. It consists of 26 measurements of HRTFs in the horizontal plane evenly spaced in azimuth. Each monaural filter was represented by 1999 samples obtained at roughly 96.0 kHz.

Other HRTF measurements reported in Fig. 6.7 were measured or computed in the lab. The naturalized cat (Panel B) was borrowed at the Paris Museum of Natural History, and measured in the IRCAM Anechoic chamber in Paris, using the same experimental setup as the IRCAM LISTEN database [Database, 2002] using the sine-sweep method. Miniature microphones were placed in the entry of the meatus, which had been occluded by the stuffing procedure.

The anechoic BEM simulations were performed using 3D meshes extracted from pictures (**C**, **D**, inserts). The BEM method is that of Chapter 4, and provided 1024 measurement samples at 44.1 kHz per monaural filter. Similar simulations were performed on the 3D model after manually tilting the head 45°so as to align it with the body (Fig. 6.7, **D**).

The spherical model HRTFs were computed according to the 1998 paper [Duda and Martens, 1998], based on the analytic solution of the wave equation by Lord Rayleigh [Rayleigh and Lodge, 1904]. The head diameter was measured on the stuffed cat 3D model (d = 7.3 cm), and the HRTFs were generated for 1024 points at 44.1 kHz. Naturalistic ground reflections were included in the spherical model (Fig. 6.7, **F**) using the method described in [Gourevitch and Brette, 2012]. The situation modeled was that of a cat with its head 20 cm above ground. The sound source was placed at a 1 meter distance from the animal at 0°elevation, and then the azimuth was varied. The ground was modeled with flow resistivity of 5.10^5, which was picked so as to roughly correspond to situations from grass fields (10^5) or sand (10^6), as opposed to highed values corresponding to rougher grounds, e.g. concrete is 2.10^9 [Gourevitch and Brette, 2012].

The rest is relative to Fig. 6.11. The Human HRTF measurements used to generate the predictions of Fig. 6.11, are that of one randomly picked subject of the openly available IRCAM LISTEN HRTF database [Database, 2002]. They consisted of 4096 data points measured at 44.1 kHz, for 72 positions on the horizontal plane.

The barn owl measurements were kindly obtained from H. Wagners group in Oldenburg. They consisted of pairs or monaural filters of 4800 samples at a sampling frequency of 96 kHz for 34 positions.

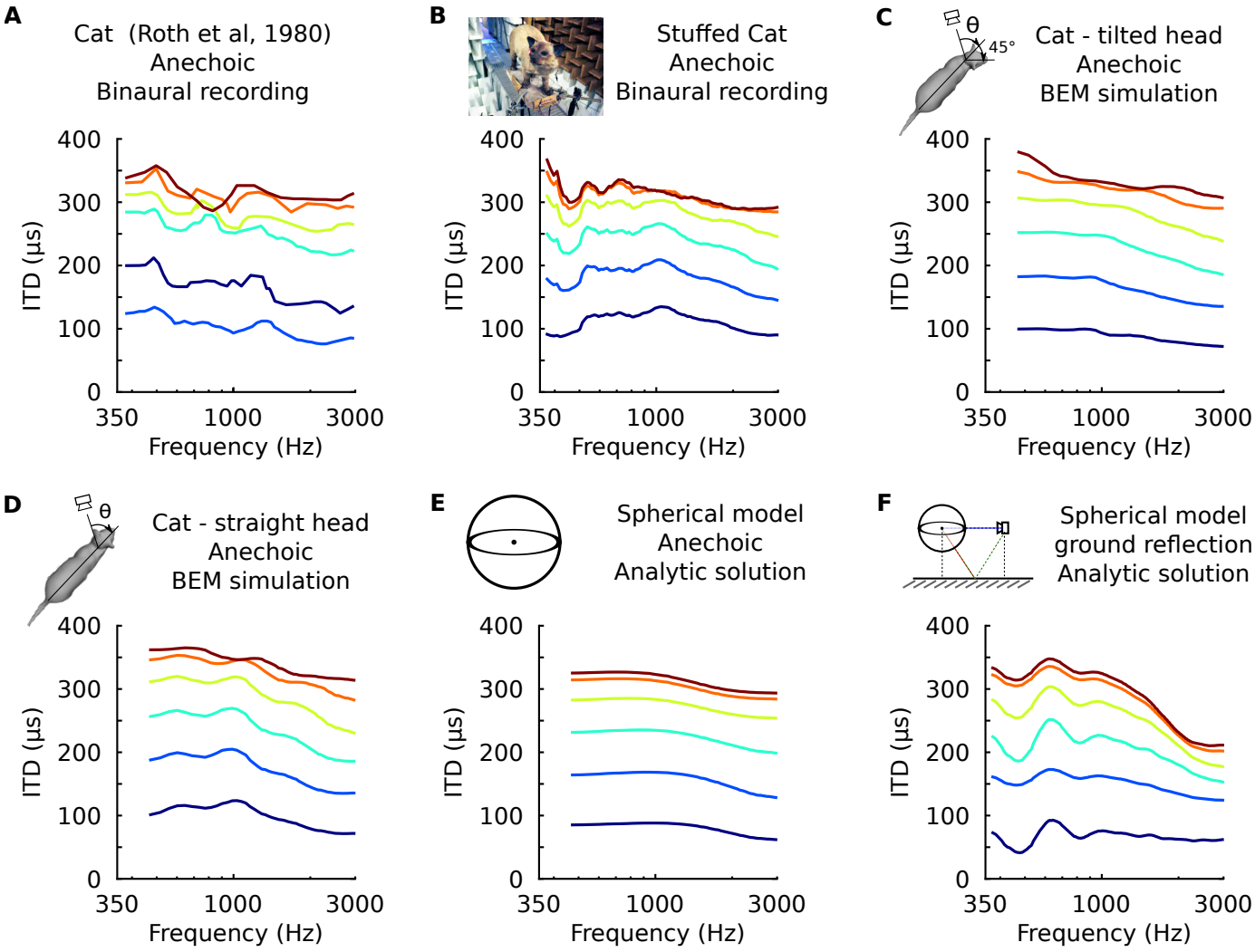

Figure 6.7: Frequency-dependence of ITD in several acoustical datasets. **A** ITD vs. frequency for 6 sound directions on the horizontal plane (15-90°, spaced by 15°), measured in a live cat and previously reported in (Roth et al., 1980). **B** Acoustical measurements on a taxidermist model of a cat in a large anechoic room (same sound directions). Note that the head is tilted on the left; directions are relative to the head (not the body). **C** Numerical calculation of ITDs by boundary element method (BEM) simulation on a 3D model of the same cat as **B** (grey shape), obtained from photographs. **D** Same as **C**, but with a straightened head. **E** Analytical calculation of ITDs for a spherical rigid head. **F** Same as **E**, but with an additional reflection on the ground. Head and source are placed one meter from each other and 20 cm above the ground.

6.B.1 Acoustical analysis

Consider the HRTF filter at position θ with the interaural phase response $\phi(f)$, measured in cycles. There exist two common definitions of interaural delays: The group delay: measures the delay of the envelope

$$\mathrm{ITD}_{\mathrm{g}}(\theta, f) = \frac{d}{df}\phi(f)$$

The phase delay: measures the total delay of the signals fine structure at any frequency.

$$\mathrm{ITD}_{\mathrm{g}}(\theta, f) = \angle\frac{d}{df}\phi(f)\rangle$$

Where the brackets operator $\angle.\rangle$ is the unwrapping operation. On discrete phase spectra, it replaces phase jumps $||\phi_k - \phi_{k-1}$ greater than 1 cycle to their 2π complement. If the phase is linear, then the IPD expression reads $\phi(\theta, f) = \mathrm{ITD}_{\mathrm{g}}(\theta)f$, and both those delays are equal. Conversely, if the phase is non-linear, then there is a difference between the two quantities. In measurements and in models of HRTFs, the ITD is found to vary with frequency [Roth]. This is indicated by a non-linear IPD spectrum (see Fig. 6.2). This arises because of diffraction of the sound wave around the head, as modeled by a spherical model [DudaMartens]. When the sound wavelength is the same as the circumference of the head of the animal (i.e. when f=c/d), then the ITD transitions between a low and high frequency regimes. We thus define the Interaural Diffraction Index (IDI), a frequency dependent quantity that represents how much the ITD varies because of diffraction effects. We define it as:

$$\mathrm{IDI}(\theta, f) = f(\mathrm{ITD}_{\mathrm{p}}(\theta, f) - \mathrm{ITD}_{\mathrm{g}}(\theta, f))$$

It is a phase quantity (measured in cycles) that has an interesting interpretation in terms of signal processing. Even though in general it is hard to analyze the effects of such phase distortions, a simple case can be analytically solved. When the source emits a band pass noise one can show [MarpleJr] that $\mathrm{ITD}_{\mathrm{g}}(\theta, f)$ measures the interaural envelope delay, that is the difference in delay imposed on the envelope of the signal at both ears. The IDIis the additional phase ITD imposed between the fine structure of the source signal, here because or diffraction. We argue that because the time delay $\mathrm{ITD}_{\mathrm{p}}(\theta, f) - \mathrm{ITD}_{\mathrm{g}}(\theta, f)$ varies with the inverse of the frequency, it is in fact better formulated in terms of additional interaural phase delay (as opposed to the description in [Roth]).

6.C Model and predictions

6.C.1 Labeled line code for sound location (HRTF-based)

We hypothesize that cells are sensitive to the frequency-dependent ITD cues associated with their preferred position BA. This can be achieved by assuming that, for all frequency points, the BD tuning of the cell match the ITD in the acoustics at the BA. In phase terms, the BP(f) spectrum (as defined in Computation of CP and CD) should be the opposite of the IPD spectrum at all frequency points. This ensures that the cell receives a maximal number of coincident inputs from both sides, hence is maximally active, when the sound is presented at the BA [GoodmanBrette].

To create a population that codes uniformly (in terms of solid angle) the position of the sound source, one has to create a population such that the distribution of BAs across the population is flat. This is very similar to the labeled-line idea as introduced by [Jeffress], even though it generalizes to non-axonal delays (because it allows for variable BDs across frequencies). In this simple scheme, the position of the sound source can be read out by finding the most active cell of the population. The azimuth of the source is then the BA of that cell.

We design a model of IC cells wherein each cell has two parameters, the CF (Characteristic frequency) and the BA (Best Azimuth). The CF of the cell defines the range of frequencies over which this cell responds. The passband of the cell is defined as a centered interval of width BW(CF)=CF/Q(CF), the Q function was a linear function of CF derived from the cell recordings. Model cells were created for 256 different CFs, from 300 Hz to 3 kHz, roughly corresponding to the range CFs in the recorded cells.

The BA of a model cell is the azimuth to which the cell is tuned. This means that, for a cell with a given CF and BA, its BP spectrum is defined as:

$$BP(f) = -\text{IPD}(BA, f) for\ f \in \left[CF - \frac{BW}{2}, CF + \frac{BW}{2}\right]$$

That is, the BP(f) spectrum of the cell exactly compensates for the IPD spectrum acoustically recorded for a source at the BA. This ensures that the cell receives maximally coincident inputs when a sound is emitted at the BA [GoodmanBrette]. From this BP(f) spectrum, we compute CPs and CDs according to the same method described in Computing CP and CD. The frequency points considered in the circular-linear regression are those in the passband of the model cells. The VS of the response of the cells was modeled as a constant, hence in essence no weighing in the CP/CD computation was considered. Additional statistics (non linearities, etc) were computed in the same way as in cells.

A mother population of such cells was generated for 256 CFs, and the 36 BAs for which acoustical recordings were available (see Acoustical Measurements). This means that the mother population contains 9216 cells sensitive to different positions and frequencies. We sample randomly from this population to report statistics, as explained next.

In order to compare with the recordings in IC, we have to equate the distribution of CFs in the modeled and measured populations. This is especially important because the distribution of interaural delays (IDIand ITD_g) varies with frequency. The predicted populationwise CP and CD distributions depend critically on this population CF distribution. We use a bootstrap approach to generate populations of cells with the right distribution of CFs. We first estimate the distribution of CFs in the data using a Gaussian kernel estimation. Then, we sample cells from the mother population using the prior distribution of CF estimated before, and a uniform distribution of BA. An example of such a population is represented in Fig. 6.3 of the main text, with N = 200. This operation can be repeated multiple times to obtain average distributions of CPs and CDs, as reported on Fig. 6.3, **A** of the main text.

Effect of azimuth prior

In the procedure presented above, the populations of modeled cells encode space uniformly. That is, the probability of selecting a cell from the mother population with any BA is equal. Some data suggest that this might not be the case, at least in the barn owl [Fischer and Pea, 2011]. To study the effects of changing the marginal distribution of BAs, we slightly modified the bootstrap algorithm presented above to sample the (CF, BA) with a non-uniform marginal distribution of BA. The CFs and the BAs of the cells were still chosen independently. Four situations were modeled, Panel A describes a situation where cells only encode from frontal-contralateral positions, in a uniform way. Panel B describes the situation were the marginal prior of BAs is matched with the one in the electrophysiological recordings (see Best Azimuths). That is, this prediction should be the closest from the recorded measures. Panel C describes the situation wherein cells are selective to the most contralateral positions, that is the positions around 90°are overrepresented. The marginal distribution of BA is a raised cosine distribution (before being normalized):

$$P(BA) = (1 + cos(BA - \pi/2))^4$$

D is the situation in which the marginal is that found in [Fischer and Pea, 2011], that is a central preference modeled with the same raised-cosine distribution as for **E**, but

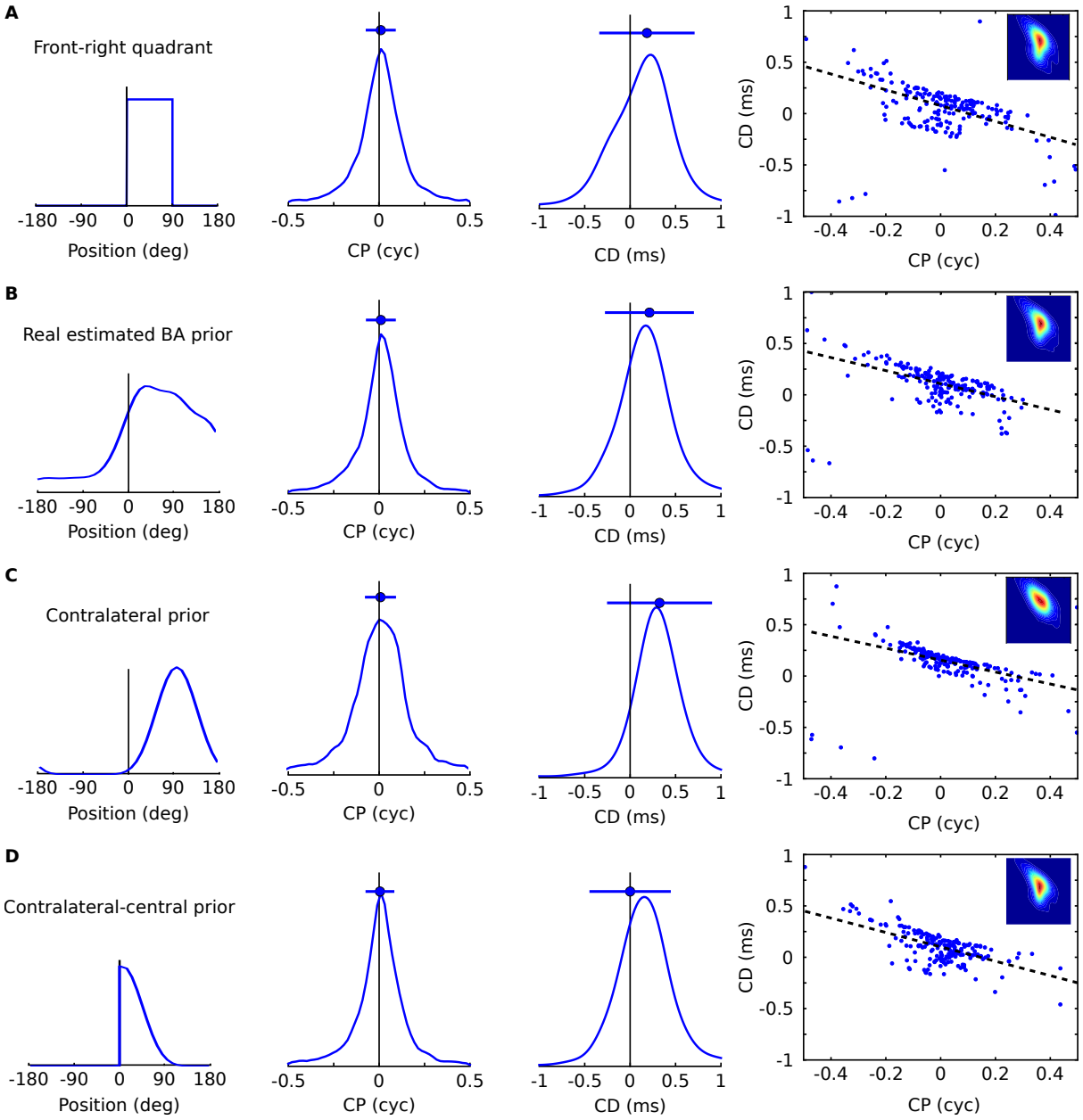

Figure 6.8: Acoustical predictions of CD and CP distributions for various prior spatial distributions. **A** First column: Uniform distribution of preferred positions in the 0-90°quadrant. Second to fourth columns: Prediction of CP, CD and joint CP-CD distributions (100 sample cells drawn at random; inset: color-coded probability density). **B** Distribution of preferred positions inferred from cell recordings (best fits to HRTFs). **C** Bias for positions near 90°. **D** Bias for positions near the center.

strictly contralateral.

Other models

Labeled line code for ITD (Jeffress) We modeled a population of IC cells coding uniformly for azimuth around different Center Frequencies (CFs). This means that at any given CF, we built 36 cells sensitive to evenly spaced Best Azimuths in the contralateral hemifield. The CFs were chosen according to the distribution of CFs observed in the data. The Jeffress model is usually formulated as a labelled line code, wherein the cell responding maximally to a given ITD in the stimulus. In this description, the Jeffress model is a special case in which each neurons INFs phase response is linear, i.e. the delays are pure delays (e.g. axonal). This has the immediate consequence that the CPs are all zero, which predicts a dirac function-like distribution. To construct the whole population of cells, we computed ITDs for each position around the animal using a broadband approximation. Because the INF is linear-phase, the CD distribution is exactly the distribution of ITDs across position and frequencies. Note that, because ITD varies as the sine of the azimuth, the distribution isnt flat, but has a peak at higher CD values.

Hemispheric model The hemispheric model is usually formulated as a pure delay model (hence the CP distribution, see previous paragraph). In this model, though, a cell at CF has BDs of at $\pm\pi/CF$, outside of the π-limit. As in the previous model, the distribution of BDs is the same as the predicted distribution of CDs. The distribution has a long tail because in this model, lower frequency cells have BDs that diverge to $+$ as the frequency goes to zero.

Best azimuths

For all the cells, the Best Azimuth is defined as the azimuth position on the horizontal plane that elicits the biggest response in that cell. According to the model, we expect this to occur when the IPD and BP compensate each other. Hence the best azimuth BA of the a cell is defined as:

$$BA = argmin_\theta \frac{1}{N} \sum_f ||BP(f) - IPD(f)||^2$$

The norm used here is a cosine-like because the quantities are circular (and measured in cycles). We used the same norm as for the circular linear regression used for computing

CP and CD (see Calculation of CP and CD). Similarly, the frequency points over which the sum is taken were picked according to the same criteria.

Since best phases and IPDs are not necessarily measured at the same frequency points, the IPD spectrum was approximated to its nearest neighbor. Because the HRTF measurements have many frequency points, the nearest neighbor was never further than a few hertz away. This is not reported here because it yields qualitatively the same results and complicates the statistical comparison with pure delay HRTFs.

CP CD correlation analysis

It could be argued that the CP/CD inverse correlation that is found in the data is just a statistical artifact. Indeed, if we consider the simple case of a linear regression, then it is known that the intercept estimate and the slope estimate covary negatively. That is to say that when the intercept is over-estimated, then the slope will be underestimated. The same reasoning can apply to our circular-linear regression used to estimate CP and CD. It could be argued that this is reflected on the correlation of estimates of CPs and CDs on different cells, even though they are in fact independent, and hence the observed correlation is artefactual.

We performed statistical simulations so as to show that this is in fact not the case. On Fig. 6.9 we show the measured distribution of CP/CD including the global linear regression (Panel A). For any given point (i.e. cell) in Panel A, we can estimate the covariation of CP and CD by using a bootstrap approach. This is illustrated on Panel B. We pick frequency points at random (for the sake of the example, just one out of two points), and compute the CP/CD on this subset of the actual data. This yields two different estimates of CPs and CDs, and we can see on this example that the intercept gets bigger (red intercept is bigger than blue intercept) as the slope gets lower (red slope is smaller than blue slope). Panel C shows the result of this procedure on 4 cells (each color is a cell, and each point the result of the estimate on a different subset of frequency points). From Panel C we do observe the expected inverse correlation between estimates of CPs and CDs for a given cell. That is, the noise distribution evaluated for each cell is not independent in CP and CD. What we see though, is that it is perfectly sensible that there is a genuine correlation between CP and CD across cells despite the negative correlation between the estimates within cell. Here, a positive correlation was intentionally included, so as to illustrate this counter intuitive fact.

Here we want to show that the negative correlation observed in the data is not due to an artifact. We try to reject the null hypothesis that the CPs and CDs are independent across the population of cells, despite the covariation of CP and CD estimates for each cell. To do so, we run simulations wherein we sample CPs and CDs independently

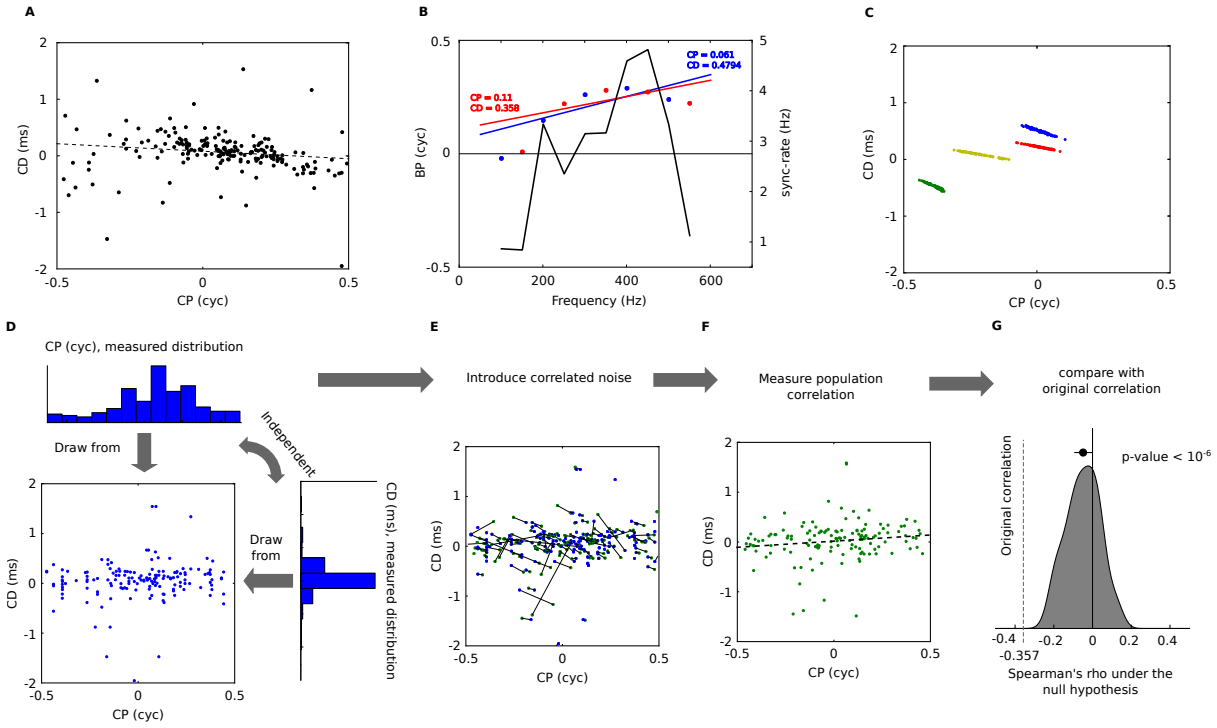

Figure 6.9: Stastistical signifance of CP-CD correlation. **A** Characteristic delay (CD) vs. characteristic phase (CP) for all cells, and linear regression (dashed). Spearmans rank correlation ρ is -0.357. **B** Illustration of spurious correlations due to noise. Two subsets of BP vs. frequency data points (red and blue dots) are fitted to lines: intercept (CP) and slope (CD) are inversely correlated. The solid curve shows the sync-rate. **C** Linear regression performed on bootstrap samples for 4 cells: CD and CP are inversely correlated for each cell, but positively correlated overall. **D**-**G** Statistical test of CP-CD correlation. **D** Data points are generated at random under the hypothesis that CP and CD are independent, using the distributions measured in cells. **E** Correlated noise is added to each (CP, CD) point, with the distribution previously measured in bootstrap samples, as shown in **C**. Each new point is shown in green, connected by a line to the original point (blue). **F** Correlation is measured across all data points (dashed: linear regression). The procedure **D**-**F** is reiterated many times with new sets of random samples. **G** The distribution of Spearmans ρ in the generated data points has a small negative bias, much smaller than in the original data points (dashed, $p < 10^{-6}$).

from the measured marginal distributions of CP an CD. This yields a set of CP/CD measurements that are not correlated (Panel D). We then introduce the measurement noise in each of this estimate (Panel E). This is done by using a randomly chosen noise distribution amongst that of measured cells (as evaluated using the bootstrap method of Panel B/C), centered so as to keep only the second-and-higher moments of the noise. The result of this procedure is that all the data points are moved in a way that reflects the covariation of CP and CD estimates (Panel F, most black bars have negative slopes). A Spearman correlation test is then run on the data, and the rho statistic is recorded. On the example of Panel F, there is a positive correlation, which appeared by chance. Repeating this procedure provides us with an estimate of the distribution of Spearmans rho values under the null hypothesis that the original population of CPs and CDs displayed no correlation (Panel G). As expected, this is on average a negative correlation, even though occasionally positive correlations may occur. Comparing the measured correlation in the original data to the cumulative of the distribution on Panel G provides us with a p-value of $10^{-}6$. This leads us to reject the hypothesis that the observed CP/CD correlation is due to a covariation of the estimates of CPs and CDs.

6.D Interpretation of variable ITDs

6.E Model predictions for different species

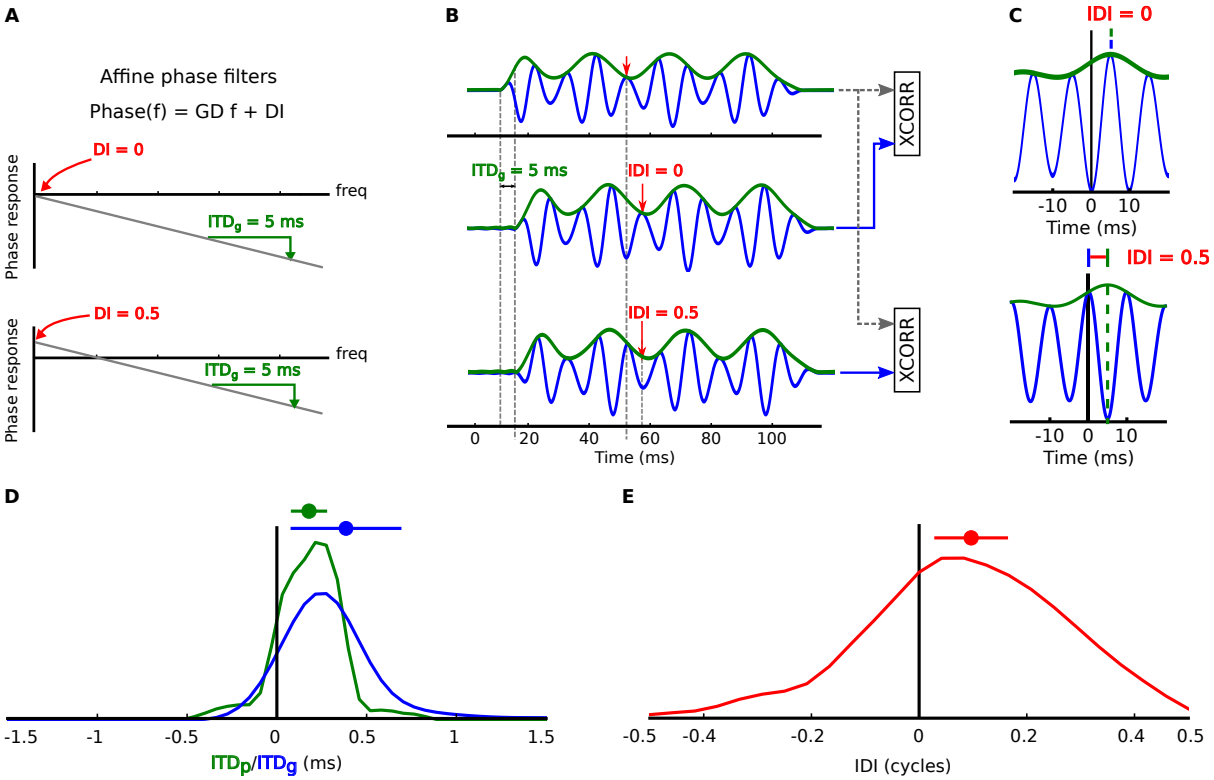

Figure 6.10: Envelope and fine structure ITDs. **A** Top: A filter with linear phase response delays signals by a constant group delay (GD). Bottom: A filter with affine phase response introduces an additional frequency-dependent delay, equal to a constant fraction of cycle, the diffraction index (DI). **B** The group delay is the delay applied to the signal envelope, while the diffraction index is the additional delay applied to the fine structure, in cycles. Two monaural signals at the left and right ear are shown (blue). The two envelopes (green) are delayed from each other by the group ITD (ITD_g, difference in GD). If the interaural difference in DI (IDI) is 0, then the entire signals are delayed by ITD_g. If it is non-zero (third signal), there is an additional delay in fine structure equal to the IDI(in cycles). **C** When cross-correlating the signals from the ears, the fine structure and the envelope peak at the same lag when IDI= 0 (top), otherwise they are delayed by a fraction of cycle equal to the IDI(bottom). **D** Distribution of phase ITD (ITD_p, green) and group ITD (ITD_g, blue) in the cat HRTFs over the 300-3000 Hz frequency range.

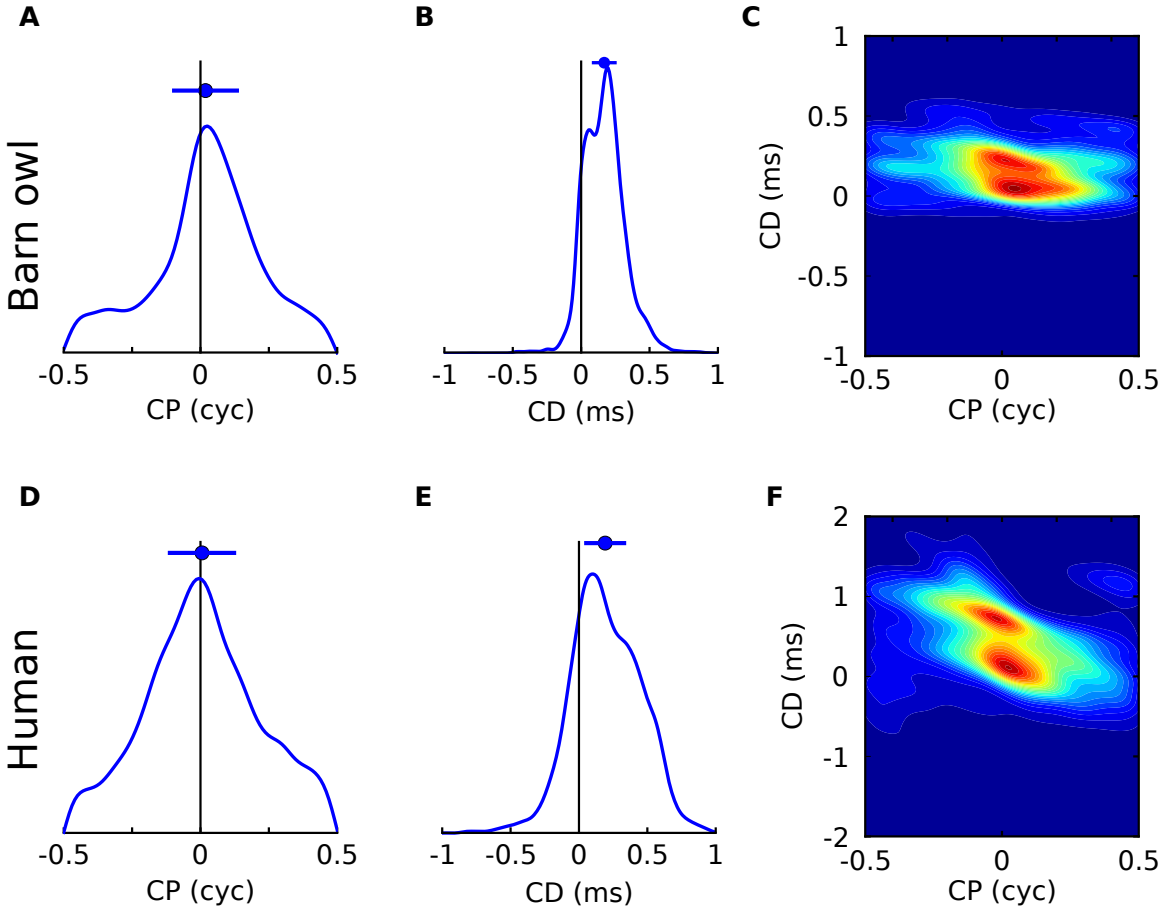

Figure 6.11: Acoustical predictions of CD and CP distributions for barn owl and human. (**A**-**C**) Predictions using measured HRTFs of a barn owl (ref), for the distributions of CP **A**, CD **B** and joint CP-CD distribution **C**, over a 500-8000 frequency range. **D**-**F** Predictions for human HRTFs Chapter 1, over a 100-1500 Hz frequency range.

Chapter 7

A model of spectrum-invariant sound localization

Contents

This part of the thesis lays the basis for a full spiking model of spectrum-invariant sound source localization. We hypothesize, that binaural cells are sensitive to the local ITD variations (with respect to) associated with a source at their *best position*. We show theoretically how this enables the cell to represent the position independently of the frequency of the input.

It is known that the neural sensitivity to frequency-dependent ITDs can be attained by a combination of axonal and cochlear delays (e.g. [Day and Semple, 2011]). Assuming that the population is a place code, the distributions of such delays and cochlear mismatches can be predicted from acoustical HRTF data. I expose a method performs this task, present the results in the case of a model of cat binaural cells and discuss their validity.

7.1 Introduction

In the previous chapter we showed that cells are sensitive to the fine ITD cues associated with one position. Given a set of acoustical recordings, we demonstrated that cells can be tuned so as to be maximally active for a given position irregardless of the source stimulus's frequency content. What neural mechanisms could give rise to such selectivity ni binaural cells have not been developed in Chapter 6. The problem is that of constructing a model of binaural cells that are sensitive to frequency-dependent ITDs. Interaural neural delays, i.e. differences in delays imposed on the signals on each monaural pathway, are necessary for ITD tuning to emerge in binaural cells (see Chapter 5). Hence a mechanism for introducing frequency-dependent neural delays is necessary for implementing this model.

The precise mechanism by which neural delays arise in the context of ITD processing is the topic of much discussion (reviewed in Chapter 5), because it constrains the plausible neural codes. For example, axons converging onto binaural IC neurons in the cat have differential lengths and diameters that can account for binaural neurons' BDs [Cant and Casseday, 1986], although not the unusually long BDs low-CF neurons. Other sources of delays were sought, and it was found that cochlear mismatches could give rise to interaural neural delays [Shamma et al., 1989]. Cochlear mismatches have other interesting properties, for example AN fibers with mismatched frequencies show maximal synchrony when they were delayed to each other [Joris et al., 2006]. This may also contribute to the BD and BF relationship observed in many mammals. Another property is that cochlear mismatches provide a mechanism for frequency-dependent neural delays. In a recent study, [Day and Semple, 2011] showed that the responses of cells in the guinea pig MSO could be fully accounted for by a model comprising cochlear mismatches and axonal delays. This is because mismatched cochlear filters have non-linear phase responses which explains nonzero CPs in the tuning of MSO cells (i.e. the frequency-dependent component of the selectivity to ITD) [Bonham and Lewis, 1999]. The [Day and Semple, 2011] study suggests that axonal delays and cochlear mismatches together can account for the full range of observed ITD tuning in mammals.

When the inputs to a binaural neuron have slightly different CFs, they originate from positions on the cochlea that are excited by the travelling wave at different times. In fact, because cochlear filters have non-linear phase responses (with respect to frequency), the interaural delays imposed on the neural signals will also depend on frequency. This is very similar to what we have seen with acoustical delays in Chapter 2. This property cannot in principle arise with axonal delays alone, which by definition impose constant BDs. In the [Day and Semple, 2011] study, the authors fit a model

consisting of cochlear mismatches and axonal delays to the Best Phase (BP) response of the cell in the guinea pig MSO. Furthermore, we have shown in Chapter 2 that the BP response of a cell could be predicted by the frequency-dependent ITD cues in the animal's acoustic environment. Taken together, these results suggest that a labelled-line code for azimuth as described in the previous chapter could be implemented by means of axonal delays and cochlear mismatches. Furthermore, because both do not have a completely interchangeable role, acoustical data can predict the relative contribution of axonal delays and cochlear mismatches in interaural neural delays.

In this chapter, we start by developing the theory of the model presented in the previous chapter. Then we show that it can be implemented using axonal delays, cochlear mismatches and cross correlator neurons. We show that cochlear mismatches introduce non-zero CPs in binaural cross correlators, and quantify this effect as a function of the mismatch and the center frequency. We then use this scheme to fit a population of cross correlator neurons to a set of synthetic cat HRTF data. Finally, We show the relative contribution of neural and cochlear delays, relative to the acoustical delays.

7.2 Spectrum invariant azimuth representation

7.2.1 Binaural cross-correlators model

Along the lines of the model described in the previous chapter, we construct a model of binaural cells that are tuned to frequency dependent ITDs. Each neuron receives inputs from both ears that are filtered by a different monaural filter and compute the cross-correlation.

In this model, the monaural filters describing the transduction of the sound signal from the eardrum to the neuron have been termed "neural" filters [Goodman and Brette, 2010, Fischer et al., 2008]. These filters represent both the cochlear filtering and further linear processes such as delays (axonal or not). A cell thus is defined by its two neural filters, that will be assumed linear for simplicity $\mathrm{N}^L(f)$ and $\mathrm{N}^R(f)$ (and their time-domain counterparts n^R and n^L). In a cross correlation model, the firing rate in response to the signals s^L and s^R (and their frequency-domain counterparts S^R and S^L) is given by:

$$r(\mathrm{s}^R, \mathrm{s}^L) \propto XCORR(n_L * \mathrm{s}^L, n_R * \mathrm{s}^R)(t = 0) \tag{7.1}$$

Where $XCORR$ designates the cross correlation function of two signals, and is taken at the time lag zero in Equation 7.1. When the two monaural signals are equal, the cross correlation at time zero is maximal. Hence the cells best stimuli are signal pairs

$(\mathrm{s}^R, \mathrm{s}^L)$ such that:

$$\mathrm{N}^L \times \mathrm{S}^L = \mathrm{N}^R \times \mathrm{S}^R$$

We define the Interaural Neural Filter $INF = \mathrm{N}^L/\mathrm{N}^R$, that describes the response of the binaural cell (up to a gain factor). The INF is closely linked to common measures of binaural cells's ITD sensitivity. The phase spectrum of the INF is the Best Phase spectrum $BP(f)$ of the cell. The modulus is proportional to the binaural sync-rate, because the noise-delay curve of a cell is the sum of tone-delay curves weighed by the VS (e.g. [Yin and Chan, 1990]).

We assume that the cell is tuned to respond maximally when the sound is at its Best Azimuth (BA). This means that Equation 7.1 holds iff the signal s is emitted from the BA. Because of the filtering of the head and body, the signals are filtered by the HRTFs at this position $\mathrm{H}^L(\theta, f)$ and $\mathrm{H}^R(\theta, f)$, which introduces binaural cues:

$$S_R(f) = \mathrm{H}^R(\theta, f) \times S(f)$$
$$S_L(f) = \mathrm{H}^L(\theta, f) \times S(f)$$

A cell thus responds maximally *iff*:

$$INF(f) = \frac{\mathrm{H}^R(\theta, f)}{\mathrm{H}^L(\theta, f)} \tag{7.2}$$

Note the similarity to the Interaural Transfer Function definition [Blauert, 1997]: $ITF(\theta) = \mathrm{H}^L/\mathrm{H}^R$. Interestingly, this formulation shows that the response of the cell is maximal under a condition that *does not depend on the sound source signal*. This is the reason why in this model cells can be said to achieve sound source localization *invariant on the spectrum* of the source. In particular, when presented with pure tones in free field at different frequencies, such a cell will respond maximally to the same position in space, albeit to frequency-dependent BDs.

In the present context, binaural neurons are only sensitive to the timing cues, and thus we are only interested in equating the phase spectra in Equation 7.2 The phase spectrum is often referred to as the Interaural Phase Difference (IPD). It follows thus that the Best Phase response of the cells exactly compensates for the IPD associated with its BA:

$$BP(f) = IPD(f)$$

In a spiking network formulation, an interesting object is the Synchrony Receptive Field of the two presynaptic monaural cells. It represents the set of positions that lead to synchronous firing of those two cells (as introduced in [Brette, 2012]). In our

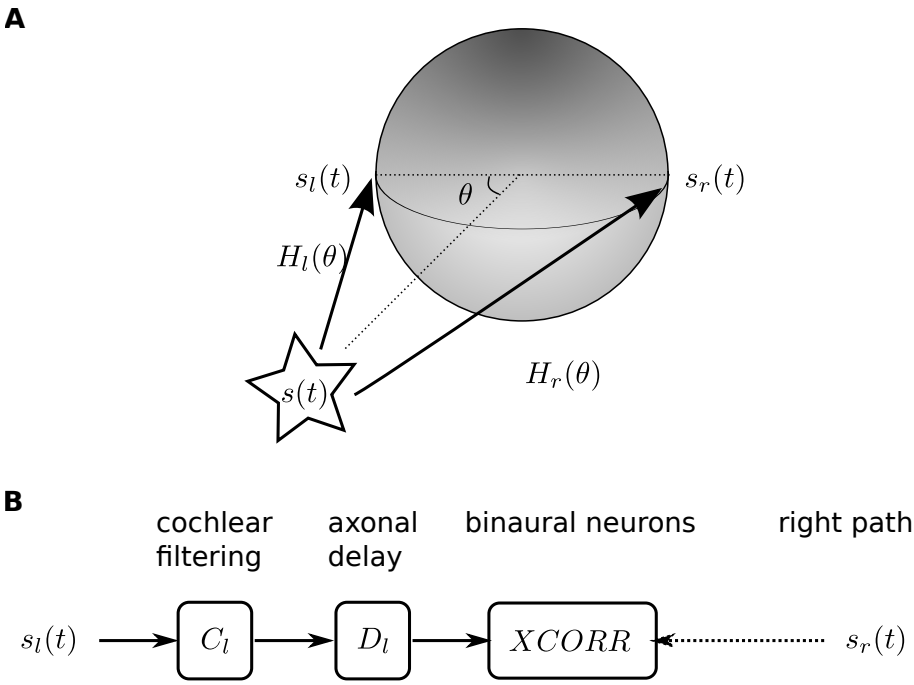

Figure 7.1: Simulations of cochlear mismatches and axonal delays **A** a depiction of the situation, the listener (a cat) receives two inputs, filtered by the HRTFs at that azimuth. **B** The signal is transduced into a neural signal by the cochlear filtering. Binaural neurons compute the cross correlation of the input with the right pathway.

formulation, by construction the SRF of the two monaural presynaptic cells is exactly the best azimuth of the binaural postsynaptic cell.

In the present study, we simulate the BP spectrum of model cells in the presence cochlear mismatches and axonal delays. Using acoustical data recorded in the cat (as part of Chapter 3, we fit the sensitivity of individual cells to the acoustical IPDs. Therefore, we can predict the optimal cochlear mismatches and axonal delays distributions for a population of binaural neurons in the cat.

7.3 Methods

7.3.1 Model of the cat auditory periphery, and INF simulations

Monaural inputs to cells are modeled as the succession of a non-linear phase cochlear filter [Tan and Carney, 2003] and a linear-phase axonal delay. The left monaural filter reads:

$$N_L = C_L(CF) \times D_L(\delta_{\text{ax}})$$

Where $C_L(CF)$ is the cochlear filter as a function of the CF and $D_L(\delta_{\text{ax}})$ the axonal delay. For any given cell, the INF can be expressed:

$$INF = C_L/C_R \times D_L/D_R = \underbrace{M(CF, \delta_{\text{CF}})}_{\text{cochlear mismatches}} * \underbrace{A(\delta_{\text{ax}})}_{\text{axonal delays}}$$

The axonal delay component A of the neural filtering is modeled as a simple delay:

$$A(\delta_{\text{ax}})(\omega) = e^{j\omega d_{ax}}$$

Where $\delta_{\text{ax}} = d_{\text{right}} - d_{\text{left}}$ is the delay imposed by the differences in axonal lengths on the monaural pathways. The cochlear mismatch component $M(CF, \delta_{\text{CF}})$ depends on two parameters: the CF of the left monaural cell, and the CF mismatch, that is the difference $\delta_{\text{CF}} = (\text{CF}_R - \text{CF}_L)$.

Unfortunately, the phase spectra of $M(CF, \delta_{\text{CF}})$ for arbitrary CFs and mismatches cannot be computed analytically from the model of cochlear filtering. This is easy in linear models like Gammatone filters. In our case, so we simulate the response of the Tan & Carney model to compute the phase spectrum of this filter as a function of the CF and the cochlear mismatch for a broad range of parameters.

Simulating cochlear mismatches

We used a Tan and Carney [Tan and Carney, 2003] model of the auditory periphery modeled which accurately predicts the response of AN fibers for cats. Because such a model has a non-analytic phase response, evaluating the INF means computing numerically the phase response of the model for each pair of parameters we consider. This part was done using Brian.Hears, a toolbox developed in the lab to do this kind of simulations [Fontaine et al., 2011].

INFs of [Tan and Carney, 2003] model alone were computed for 128 positions on the left cochlea arranged between 100 Hz and 3 kHz in a logarithmic scale. For these CFs, simulations were done for $N = 100$ frequency mismatches δ_{CF}, that is 100 different points on the right cochlea. The mismatches spanned an octave centered around the CF in a logarithmic scale. For each CF computed, we have access to 100 frequency mismatches between one half octave up and one half octave down.

For any given CF and δ_{CF} we compute the response of the Tan & Carney model [Tan and Carney, 2003] on 20 CF periods-long white-noise stimuli at a constant level (60 dB SPL). We then cross-correlate the left and right responses to measure the phase response of the INF for each combination of parameters. Cross-correlation functions

are upsampled 4 times, and the phase response is then averaged (in a circular sense) over 100 iterations.

Measure of CP and CD

The BP spectrum of a model cell can be inferred, for each combination of axonal delays and cochlear mismatches (as the phase of the INF filter). In order to quantify the influence of cochlear mismatches on the frequency dependent ITD sensitivity of cells, we can compute the BP of model cells with different cochlear mismatches. From the BP spectrum, we compute CPs and CDs as a function of the cochlear mismatch parameters. Those are the center frequency of the left cochlear, and the difference in CF such that $\delta_{\mathrm{CF}} + CF_l = CF_r$. The CPs and CDs are computed using the same method as was applied to IC cells in the previous chapter (Chapter 6). The bandwidths over which fits were computed is defined by a Q-factor of 2.

7.3.2 Fitting to acoustical data

Acoustical data

We used simulations of the BEM method described in Chapter 4. It is the synthetic cat recordings, with the head aligned to the body (i.e. modified 3D model). This allows us to make use of the phase response of the filters down to a very low frequency, which is more complicated with the acoustical data due to noise, and very low-frequency reflections in the room.

Fitting procedure

We construct cells such that their BP spectrum matches the IPD at their Best Azimuth (BA):

$$BP(CF, \delta_{\mathrm{CF}}, \delta_{\mathrm{ax}}) = IPD(CF, BA)$$

Where the BP spectra are computed from the measured INFs in the simulations of the Tan & Carney model [Tan and Carney, 2003], and adding the contribution of an interaural axonal delay. We now look for the parameters $delta_CF$ and $delta_ax$ such that this relationship is met, across center frequencies and positions in space. This is done by minimizing a circular measure as presented in Chapter 6. The values of the IPD and BP are compared over an octave centered on the CF, the passband PB. Those fits are performed with a broader range ($Q = 1$) than used to measure CP and CDs on

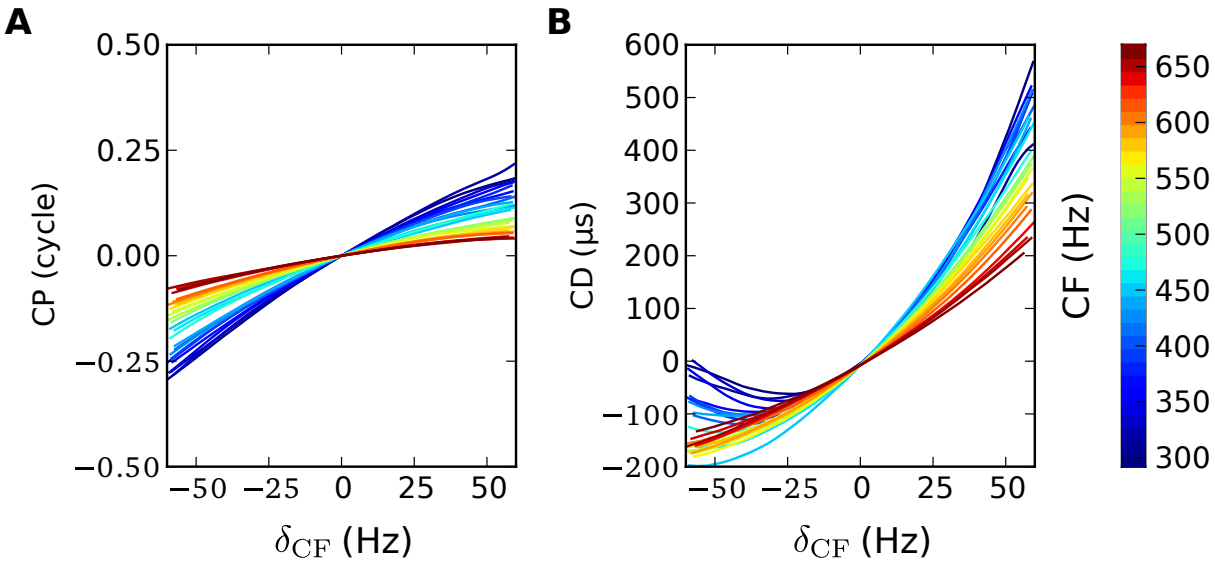

Figure 7.2: Influence of cochlear mismatches on CP and CD. **A** CP as a function of cochlear mismatch (x axis) and center frequency (line color). The color bar is on the right. **B** CD as a function of cochlear mismatch (x axis) and center frequency (line color).

cells. Hence the non-linear least squares problem reads:

$$\widehat{\delta_{\mathrm{CF}}}, \widehat{\delta_{\mathrm{ax}}} = argmin_{\delta_{\mathrm{CF}}, \delta_{\mathrm{ax}}} \sum_{f \in PB} d(-IPD(CF, BA), M(CF, \delta_{\mathrm{CF}}) - \delta_{\mathrm{ax}} f))^2$$

Where the circular distance is defined as in the previous chapter (see Chapter 6).

The algorithm used here is different, because we use an exhaustive search approach: computin the distance function for all combinations of parameters on a wide range, and the taking the minimum. Therefore we are also able to inspect the distance function on the whole space of parameters to check the validity of the fits. Also, we can get an idea of the relative importance of each parameter. The ranges of possible parameters were chosen wide, interaural axonal delays δ_{ax} could span ± 1 ms range, and the cochlear mismatches half an octave up and half an octave down the CF.

7.4 Results

7.4.1 Cochlear mismatches and neural CPs and CDs

We first reproduce the results (also obtained with a Tan & Carney model) in [Bonham and Lewis, 1999], by setting $\delta_{\mathrm{ax}} = 0$ in our model. Fig. 7.2 shows the influence of cochlear mismatches on the CP and CD of a binaural cell. The BP spectrum is in general well approximated by a circular-linear regression. The Characteristic Phase

(CP) of the cell is the intercept of the fit and the Characteristic Delay (CD) is the slope of the fit (see Chapter 6). Results in Fig. 7.2 report the predicted CP and CD as a function of the cochlear mismatch, for 100 linearly spaced CFs between 300 and 700 Hz. The result of these simulation is reported on Fig. 7.2, around $CF = 500$ Hz and with small mismatches $|\delta_{\text{CF}}| < 50$ Hz.

As expected, cochlear mismatches introduce significant CDs (Fig. 7.2, **B**), reaching up to 500 μs for relatively low mismatches (roughly semitone). This effect is much more pronounced in the low-CF range, than it is in the high-CF range. In essence, this suggeststst that cochlear delays can replace axonal delays, in that they impose a linear-phase component to the monaural signals. Notice that it can account for delays biggers than the physiological range for cats $\approx 400\mu s$, even in this restricted range of mismatches. Consistent with the [Day and Semple, 2011] study, cells with mismatched monaural CFs have non zero CP in general (Fig. 7.2, **B**). Again, this phenomen is more clearly observed for the low-CF neurons. Yet, the magnitude of the phenomenon at those frequencies is limited, only about a quarter cycle advance for the range of mismatches considered here (a 100 Hz band). However, greater CPs can be predicted assuming bigger mismatches (not shown).

The prediction of non-zero CPs for non-zero cochlear mismatches expresses that in fact cochlear and axonal delays are not completely equivalent. Axonal delays are linear-phase and cannot account for the presence of non-zero CPs. On the other hand, the effect of a cochlear delay is more complex and introduces non-linearities in the monaural phase responses, leading to non-zero neural CPs.

7.4.2 Fit to the positions

Given a set of HRTF measurements for a position, we can fit a cell such that it is sensitive to the precise ITD variation present in the acoustical data. This cell will then be said to have a Best Azimuth equal to that position in space. This is done by minimizing a circular distance between model BP spectra and the IPD pattern at the BA. For each position and center frequency CF, many combinations of $\delta_{\text{ax}}$and δ_{CF}parameters are simulated, and their fit to the acoustical data is computed. The fit value on the whole simulated parameter space for two positions and two center frequencies is reported on Fig. 7.3. We observe that for one given BA and CF, the range of parameters that yields a good fit is broad (blue valley), and that there is an obvious interplay between $\delta_{\text{ax}}$and δ_{CF}. The aboslute value of the fit quality is in general very low, about 0.02. Further, from visual inspection of Fig. 7.3 it appears that in the range of parameters considered, the minimum is global and therefore is well-defined.

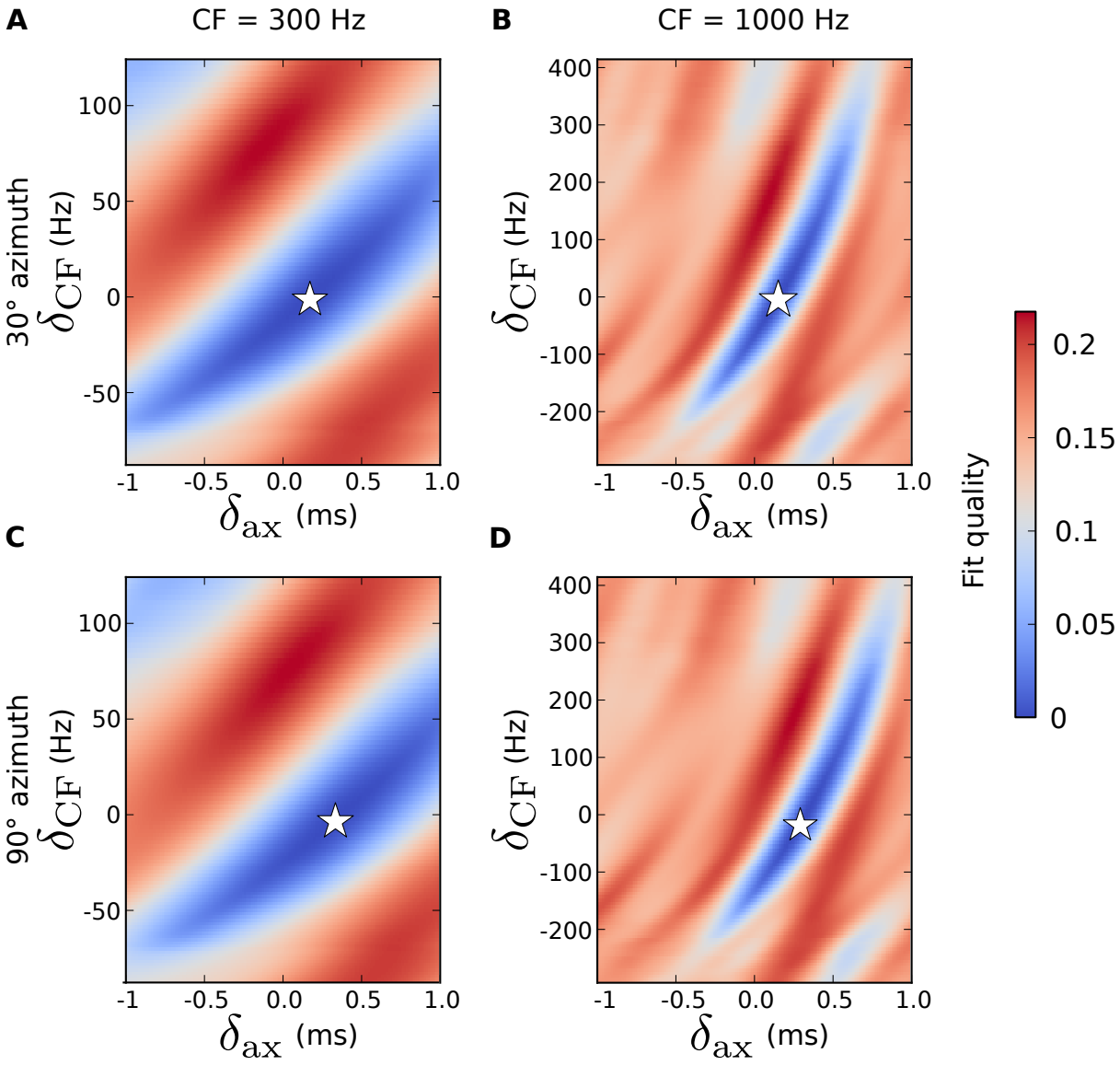

Figure 7.3: Fitting to a position. Each panel shows a color plot of the distance between the IPD at the BA and the cell's BP spectrum (see Methods). The fit quality is color-coded as a function of the axonal delay mismatch d_{ax} and the cochlear mismatch δ_{CF}. White stars on each panel show the optimal parameter value (minimum of the function). Values of the color code are common across panels (color bar to the right) and simulations were performed in four conditions. Rows correspond to cells with different Best Azimuths θ while columns are the same cells with different Center Frequencies: **A** $\theta = 30°$, $CF_l = 300Hz$, **B** $\theta = 30°$, $CF_l = 1000Hz$, **C** $\theta = 90°$, $CF_l = 300Hz$, **D** $\theta = 90°$, $CF_l = 1000Hz$.

As shown on Fig. 7.2, cochlear mismatches introduce both a neural CP and a neural CD. Axonal delays, on the other hand, only contribute additively to the CD, the linear component of the phase response. Therefore, ignoring CPs, a cochlear mismatch is equivalent to an axonal delay in terms of phase response. Consistent with those observations, on Fig. 7.3 the δ_{CF} and the δ_{ax} seem compensate each other, the blue "valley" of good parameters has a positive slope, indicating that positive cochlear mismatches compensate for positive axonal delays. Yet, this valley has a limited extent in both directions, which reflects the fact that $\delta_{\mathrm{ax}}$and $\delta_{\mathrm{CF}}$are not exactly interchangeable. This justifies the use of a fit to predict the *relative* contribution of axonal delays and cochlear mismatches, as is discussed in the next section. Indeed, had their role been totally indentical, then there would be an infinite number of optimal parameter combinations, corresponding to equivalent effects. Yet, this is not required here, and the relative contribution of the two delay mechanisms are obtained solely on the basis of acoustical data.

7.4.3 Predicted distributions of cochlear mismatches and axonal delays differences

For each CF and BA, we compute the optimal combination of axonal delays and cochlear mismatch for our frequency-invariant cells (Fig. 7.4, **A** and **B**). In Fig. 7.4, **A** the optimal axonal delay difference is color-coded as a function of frequency and position on the contralateral hemifield. As expected, the most excentric positions (i.e. away from the midline) necessitate large axonal delays, and at all frequencies, this relationship is almost monotonous. The parameters for which the axonal delays are greatest are low frequency and about 100°azimuth. The analysis of the optimal frequency mismatches displays a different story: predicted CF mismatches are mainly negative, and the only positive δ_{CF} occur for positions on the back and at quite high frequencies. This plot is consistent with the idea that cochlear mismatches are necessary when the cell has a non zero CP. In turn, this means that the CP should correlate with the acoustical IDI measure introduced in Chapter 2. This happens when early reflections are present on the body, so for positions originating from the back of the animal (see Discussion of Chapter 4). Also, IDI is nonzero in the cat for mid-frequencies in the range of 2-3 kHz, which is consistent with the frequency mismatch prediction.

Population-wise statistics of the cochlear mismatches and axonal delays are obtained by assuming that the cells in the population have linearly spaced BAs (every 5°), and arranged in frequency bands with logarithmically scaled CFs (128 frequency bands between 100 Hz and 3000 Hz). This is broadly consistent with the distribution of CFs

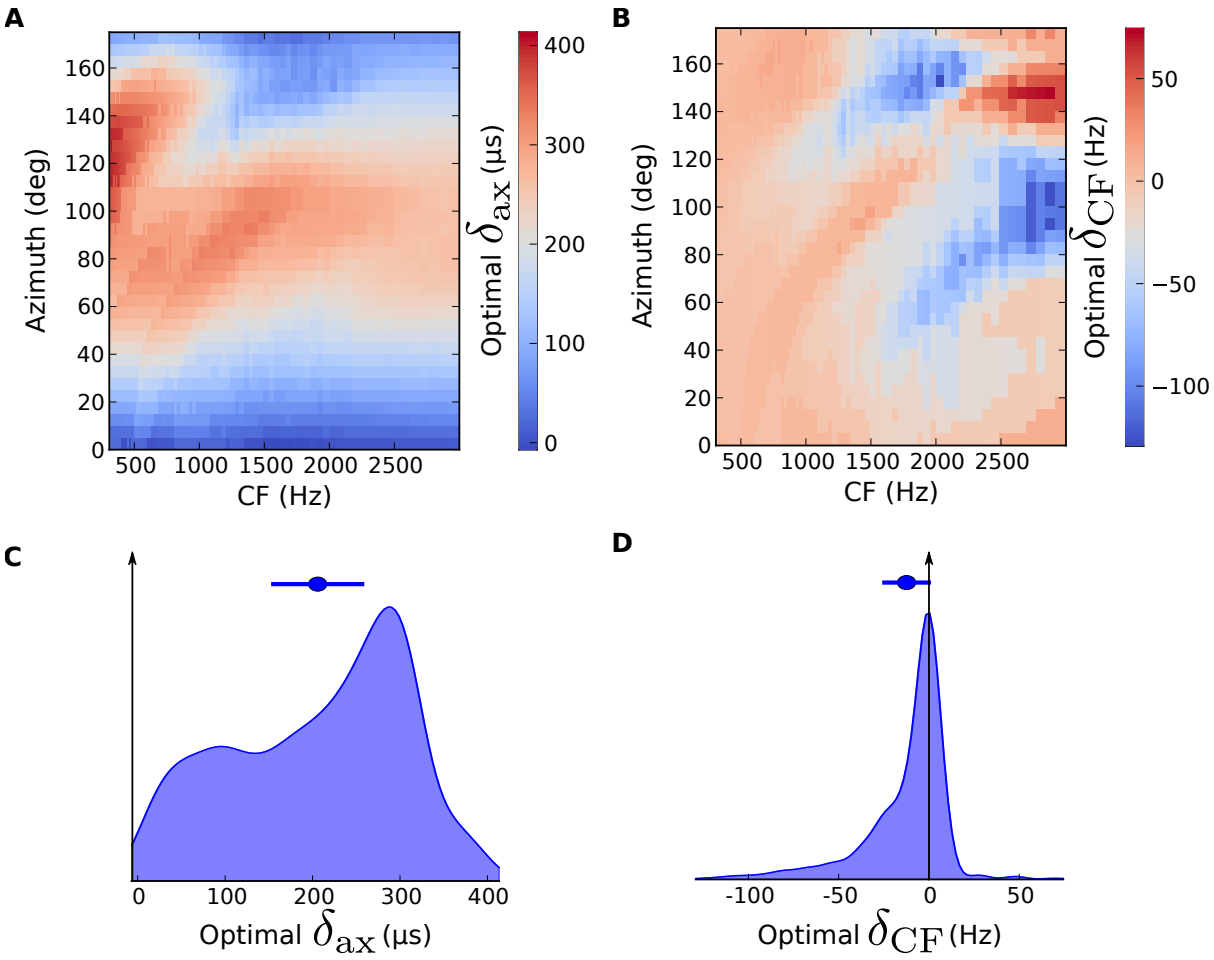

Figure 7.4: Distribution of optimal δ_{ax} and δ_{CF} for a model population of cells. **A** and **B** are color plots of the optimal δ_{ax} (**A**) and δ_{CF} (**B**) as a function of the BA and CF. **C** Distribution of δ_{ax} over the population of cells. **D** Distribution of δ_{CF} over the population of cells.

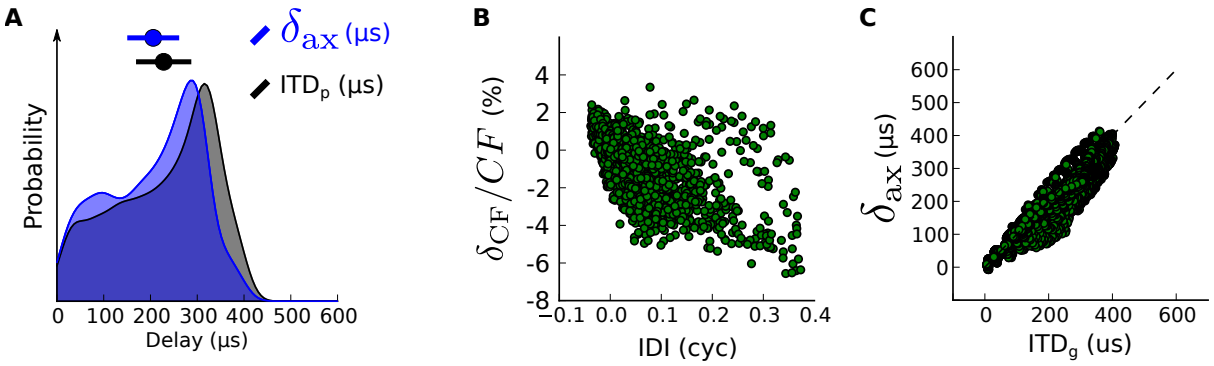

Figure 7.5: Comparison of predicted delay mechanisms and ITD cues. **A** Total phase ITD distribution (physiological range) of a cat (gray area) alonside the predicted distribution of axonal delays. **B** Each point is the predicted cochlear mismatch magnitude for a cell (in percentage of the CF), plotted against the IDI in the HRTF at the cell's BA, around the cell's CF. **C** Each point is the predicted axonal delay mismatch for a cell as a function of the group ITD ITD$_g$ (similar as **B**).

found in IC recording data (see, e.g. Chapter 6). On Fig. 7.4, **C** and **D** we report the population-wise distributions of predicted axonal delay and cochlear mismatch. The predicted distribution of δ_{ax} is also consistent with data, within the range or observed axonal delays [Joris et al., 2006]. The δ_{CF} are also in general quite small, because they occur only for the small amount of CF and position combinations that lead to frequency-dependent ITDs.

7.5 Discussion

7.5.1 Neural and acoustical delays

Axonal delays and cochlear mismatches are sufficient to compensate for the acoustical delays in the environment of the animal. As it is shown here, they have non-interchangeable roles, and some of the contribution of the axonal delay can be compensated for by cochlear mismatches. Thus, it is possible that they are involved in compensating for the two different interaural timing cues defined in Chapter 2.

In **A** of Fig. 7.5, we show the distribution of total phase ITDs ITD$_p$ (blue area) alongside the predicted distribution of axonal delays. The black area is the distribution of the ITDs when the positions are uniformly distributed in azimuth. Axonal delays are consistently smaller than the total phase ITD range, which suggests that some of the neural delaying is performed by the CF mismatches. Therefore, it is not surprising from our model that observed axonal delay mismatches cover a smaller range than the physiological range of the species. Our model predicts that axonal delays of the value

of (or greater than) the biggest ITDs should not occur. Instead, cochlear mismatches always contribute a small portion of the total neural delay.

Cochlear mismatches are also hypothesized to be responsible for the frequency-dependent BDs observed in MSO cells [Bonham and Lewis, 1999, Day and Semple, 2011]. As we also report in Fig. 7.2, cochlear mismatches can explain the presence of non-zero CPs in binaural cells, and furthermore, as we reported in Fig. 7.4, they seem to occur only in some range of positions and frequencies. We argue that axonal delays and cochlear mismatches are separate mechanisms involved in separately compensating for the two types of acoustical inteaural delays available to the cat: group ITD and IDI (Chapter 3). Therefore, we expect that CF mismatches occur for BAs and CFs wherein IDI are non-zero. This is indeed the case in the data reported on Fig. 7.5, **B** where a strong nergative correlation is found between the relative CF mismatch (δ_{CF}/CF) and the IDI (Spearman's $r = -6$, $p < 10^{10}$). For each cell, the IDI (and later the group ITD, $\mathrm{ITD_g}$) are computed on the pair of HRTF measurements corresponding to a source at the cell's BA, and around the cell's CF. We also expect that the axonal delay, because it only contributes a linear-phase delay to the neural signals, compensates for $\mathrm{ITD_g}$ and therefore is correlated with it on a population scale. This is shown on Fig. 7.5, **C**, where similarly to **B**, the cell's $\delta_{\mathrm{ax}}$is plotted against the corresponding $\mathrm{ITD_g}$in the acoustical data, and a positive correlation is observed. In conclusion, CF mismatches and axonal delays do compensate separately for IDI (CF mismatches), and group ITD (axonal delay differences).

7.5.2 Models of ITD based localization

The present study shows that cells can be made to be maximally active when presented with a sound at their BA using a combination of axonal delays, cochlear mismatches and coincidence detector neurons. This scheme implements a labelled-line code for frequency-invariant ITD based sound localization. Because all positions in the acoustical data can be fitted using biologically plausible axonal delay and cochlear mismatch parameters, the primary result of this study is that it lays a basis for future spiking implementations of this scheme.

The mathematical framework developed here is very close to ideas from the signal processing realm [MacDonald, 2008, Durkovic et al., 2011], where algorithms that encode the precise variations of ITD across frequency are shown to be more robust to noise and reverberation. Those results could well apply to a neuromorphic instantiation of this scheme, wherein the cochlea is implemented *in silico* [Chan et al., 2007]. Such hardware was sucessful in implementing an axonal delays only ITD-based localization

mechanism [Yu et al., 2009], which could be extended to making use CF mismatches with the developments described in this chapter.

Chapter 8

A normalized peak code for ITD

Contents

The textbook model of ITD coding is that the most active neuron of a population signals the ITD in the inputs. This model was recently challenged because it was observed that neurons were maximally active for ITDs too large for the animal to experience. In this chapter I argue that this is in general not in contradiction with the idea of a place code. To show this, I provide an instantiation of a peak code that is consistent with the observation of large BDs in low frequencies, without postulating large neural delays. In this model, the response of each cell is normalized so that the average population response is constant whatever the ITD in the input. I show that

this simple model is indeed a peak code of ITD, in which the cells are the most active of the population for ITDs away from the Best Delay as observed from the noise delay curve.

I explore the predictions of BD distributions for this model across populations of cells having different CFs, and for different animals. In the low frequencies, the observed BDs are consistently well oustide of the physiological range, in a manner consistent with the data in different species. I discuss the various possible implementations of this model, and its biological plausibility. Further, I argue that this model reveals a fundamental misinterpretation of single cell sensitivity data.

8.1 Introduction

In the mammalian MSO, the activity of populations of cells is modulated by ITD. The way in which the position is encoded in this ITD-dependent pattern of activity is still a matter of debate (as argued in Chapter 6). In his 1948 paper, [Jeffress, 1948] considers that his model transforms a "difference in time" (the ITD) represented as a "difference in place": it is a *place code*. In the present meaning, a place code does not necessarily refer to physical space, i.e. does not postulate a particular spatial arrangement of neurons. Rather, in a *place code*, it is only assumed that the neuron's individual identity contributes information to the read-out. As an example, in the code proposed by Jeffress, the ITD is represented as the Best Delay of the most active neuron in the population. The read-out depends on the "identity" of the cell in the sense that if we shuffle the neurons by randomly shuffling the BD and response measures of the population, then the readout result is different. Alternatives to place codes exist, and maybe provide a better understanding of the meaning of place codes. As an example, in a *population code*, the position of the sound source is encoded in the summed activity of all neurons across the population. The relevant quantity is a scalar (the summed activity), and if the BDs of the cells are shuffled, the result is identical. I such codes, the read-out does not not depend on individual characteristics of the cells, it is not a *place code*.

The Jeffress-like code is in fact a specific instance of a place code termed a *peak* code, because only the peak of the activity pattern across cells matters. This has to be contrasted with other codes that take into account the activity of the whole population, such as the weighed-image model [Stern et al., 1988]. In such models, the activity of the population represents the short-time cross-correlation of its inputs (the BD). Each neuron's activity corresponds to this cross-correlation function measured at a given time-lag. From this activity it is possible to read-out the position using a centroid function, that is more or less finding the modes of the cross-correlation, weighed according to other factors (such as centrality or consistency of ITD across bands, [Stern et al., 1988]). In the sense exposed here, this is also a place code, because the weighing and centroid operations depend on the BD of each individual cell. Interstingly, the current debate on the code relevant to the binaural ITD-sensitive cells functioning in mammalian MSO opposes place and population codes.

In mammalian MSOs, it is consistently found that some (not all) low-CF cells have very large BDs, well outside of the range of ITDs that the animal naturally encounters. This was argued to be in contradiction with peak codes, which lead to the introduction of a population code, the hemispheric model (see Chapter 6). Here we argue otherwise,

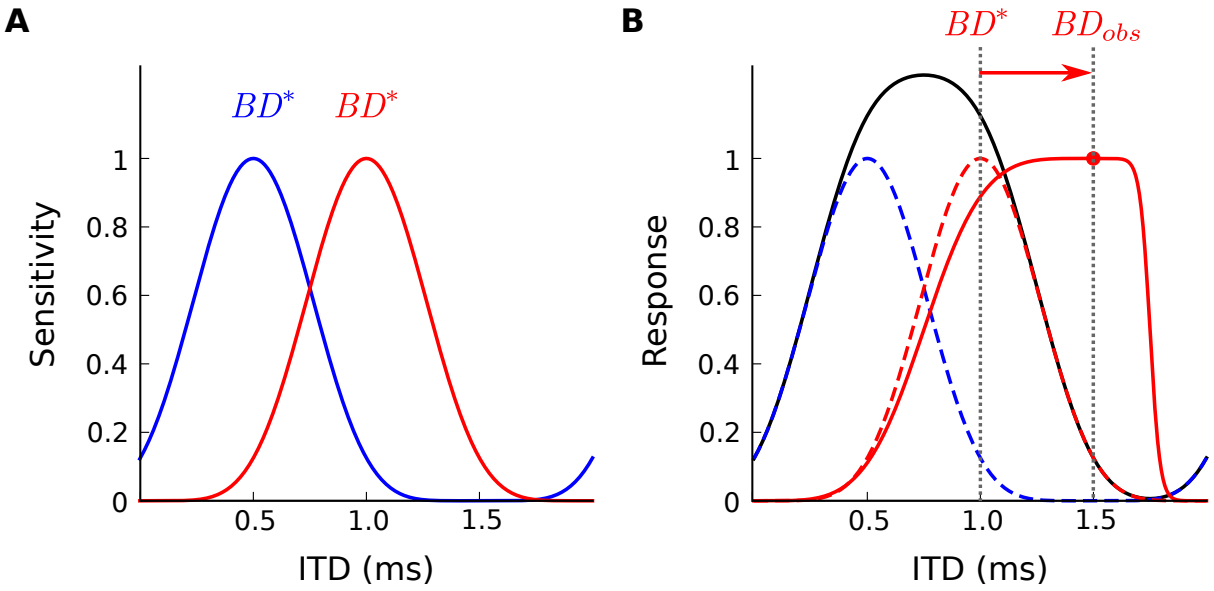

Figure 8.1: A depiction of the model for a population of $N = 2$ cells. **A** The sensitivity curves of the cells in the absence of normalization. $BD^* = 0.5$ ms for the blue cell and $BD^* = 1.$ for the red cell. **B** The responses of the cells (plain curves) are the sensitivity (response without normalization, dashed curve) divided by the summed response of the N = 2 neurons (black curve). This normalization depends on ITD, but is the same for both cells. Because of this term, the *observed* BD of the red cell is shifted away from BD^*.

by providing an example of a peak code that is consistent with the observations of large BDs in low frequencies. The reasoning is that in a place code, the readout is not necessarily exactly the BD of the most active cell, rather, it is a function of it (much like in the lateral-image model). In the model presented here, a cell can have a large BD, while being the most active of the population for ITDs within the physiological range.

8.2 A normalized peak code for ITD

8.2.1 Single cell sensitivity in the model

A mammalian MSO is modeled as a population of cells sensitive to the same band of the spectrum of the inputs, as identified by the Characteristic Frequency (CF). The population consists in N binaural cells, identified by a *neural Best Delay* quantity, noted BD^*. This neural BD corresponds to the difference in neural delays converging to the cell from the two ears, that gives rise to the neuron's sensitivity to ITD. Therefore it is, for example, the difference in axonal propagation delays leading to an MSO neuron,

if we assume that propagation delays only contribute to MSO sensitivity. Yet, be aware that the development presented here does not depend on the nature of this delay.

The response of a cell can be expressed as a function of the cell identity (its neural BD) and the stimulus value ITD. We let $s_{\mathrm{BD}^*}(\mathrm{ITD})$ bet the spiking frequency of the cell with BD when presented with a stimulus with ITD, presented here in an adimensional unit. This response value is termed *sensitivity* of the cell, because it solely depends on BD^* the difference in monaural delays leading to the binaural cell. It represents the activity of a cell as a function of ITD, without the normalization that will be devised in the rest of this chapter. In other words, it is the tone-delay curve of the cell in the absence of normalization. The sensitivity is modeled as a raised-cosine function of the ITD:

$$\mathrm{s}_{\mathrm{BD}^*}(\mathrm{ITD}) = \left[\cos(2\pi CF(\mathrm{BD}^* - \mathrm{ITD}))\right]^6 \tag{8.1}$$

The sensitivities of two model cells are presented on **A** of Fig. 8.1, in the case where CF = 500 Hz. This shape reproduces the allure of noise-delay curves, namely: it respects the 1/CF periodicity of electrophysiological noise-delay curves (e.g. in gerbil data [Day and Semple, 2011]). Further, it exhibits a maximum at $\mathrm{ITD} = \mathrm{BD}^*$, that is when the ITD in the stimulus compensates for the neural delays leading to the cell.

8.2.2 Population normalization

Crucially, we assume that the total activity of the population when presented with an ITD sums to a constant value. This means that a cell's response is modulated by the response of all other cells in the population. Therefore, the activity of the cell with BD BD^* reads:

$$\mathrm{r}_{\mathrm{BD}^*}(\mathrm{ITD}) = \frac{\mathrm{s}_{\mathrm{BD}^*}(\mathrm{ITD})}{\sum_{\mathrm{BD}} \mathrm{s}_{\mathrm{BD}}(\mathrm{ITD})} \tag{8.2}$$

The numerator is the sensitivity of the cell, and the denominator is the summed activity corresponding to the normalization factor. With normalization, the response is modulated by a constant term for all cells, that depends solely on the ITD. This normalization could correspond to homeostatic constraints, enforcing, in particular, that the energy budget attributed to the population is constant for all ITDs (see the Discussion Section 8.4.1).

In the presence of normalization, the noise-delay curve of a cell is modeled by r_{BD^*} of Equation 8.2. I will sometimes refer to a situation "in the absence of normalization". In this case a cell's response is equal to its sensitivity $\mathrm{s}_{\mathrm{BD}^*}$, with both functions modelling the electrophysiological noise delay curve.

8.2.3 Peak code

In the introduction, we used the word *code* to refer to a decoder, that is a mapping from the activity of the population to an ITD. Therefore, a *code* is a function G of the activities of all N cells, at a given ITD:

$$\widehat{\mathrm{ITD}} = G(r(\mathrm{BD}_0^*, \mathrm{ITD}), ...r(\mathrm{BD}_N^*, \mathrm{ITD})) \tag{8.3}$$

In a peak code, the *estimated ITD* $\widehat{\mathrm{ITD}}$ given the responses of the population, is the *Best Delay* of the cell that has maximal activation. In essence, this is an argmax operation *across the population of cells*:

$$\widehat{\mathrm{ITD}} = \mathrm{argmax}_{\mathrm{BD}^*} \mathrm{r}_{\mathrm{BD}^*}(\mathrm{ITD})$$

For simplicity, we assume that the reponse function r is reasonably regular, and that the neurons are arranged in a continuous array of BD^*. In this case, computing $\partial r / \partial \mathrm{BD}^*$, and looking at the zeros of this function that correspond to maximas of r provide the identity of the most active neurons in the population.

Without normalization, we are in the same situation as the classical Jeffress model. The estimated ITD, $\widehat{\mathrm{ITD}} = \mathrm{argmax}_{\mathrm{BD}^*} \mathrm{s}_{\mathrm{BD}^*}(\mathrm{ITD})$, is exactly the neural delay BD^* of the most active cell of the population. This is because, in Equation 8.1 the two variables BD^* and ITD have interchangeable roles. Hence, the maxima along the ITD dimension and along the BD^* dimension occur at the same place. In other words, each cell is most active of the population when it is presented with an ITD at its BD^*.

When the responses are normalized, this intuitive result may not be so clear. Yet, we oberve that the normalization term, $1/(\sum_{\mathrm{BD}} r(\mathrm{BD}, \mathrm{ITD}))$ does not depend on identity BD^* of the cell: it only depends on the stimulus dimension value (here ITD). Therefore, the derivatives of $\mathrm{r}_{\mathrm{BD}^*}(\mathrm{ITD})$ and $\mathrm{s}_{\mathrm{BD}^*}(\mathrm{ITD})$ with respect to BD^* are equal up to a factor (that depends on ITD). In particular, s and r vanish for the same values of BD^*. Therefore, the maxima of the responses $\mathrm{r}_{\mathrm{BD}^*}(\mathrm{ITD})$ across the population (i.e. across BD^*) are the same as the maxima of the sensitivities $\mathrm{s}_{\mathrm{BD}^*}(\mathrm{ITD})$. Thusly, the decoded ITD $\widehat{\mathrm{ITD}}$ is not affected by normalization. In conclusion, in the present peak code scheme, each cell is the most active of the population for ITDs close to its neural BD^*, whether normalization is present or not.

8.2.4 Observed BD and neural BD

The usual definition of the Best Delay in electrophysiology is the maximum of the response of the cell, *across ITDs*. In the absence of normalization, it is obvious that

the observed BD should be the same as the neural BD (that is, BD^* for all cells), because by construction, the response is equal to the sensitivity and s_{BD^*} has peaks at BD^*. This means that in the absence of normalization, the Best Delay measured from the noise delay curve is exactly the difference in neural delays leading to the binaural cell. This result is rather intuitive, and is a very common assumption in neurophysiology. As an example, large BDs computed from noise delay curves in the low-frequency cells of small mammals is thought of as reflecting correspondingly interaural neural delays converging to the cell. Here we show that the BD, as extracted from the noise delay curves of the cells do not necessarily reflect the interaural neural delays BD^*. In our model, this is by virtue of the introduction of normalization across cells.

Let us define the *observed* BD of the cell $BD_{obs}(BD^*)$, i.e. the value of ITD that makes the tone-delay curve maximal:

$$BD_{obs}(BD^*) = \text{argmax}_{ITD} r_{BD^*}(ITD) \tag{8.4}$$

This definition is to be contrasted with the estimated ITD $\widehat{ITD}$, corresponds to the position of the maximum of the response *across the population of cells*.

Along the arguments of the previous section, the observed BD is obtained by computing the partial derivative of r_{BD^*} with respect to ITD. Because the normalization term depends on the ITD, it will affect the partial derivative of the response, which may not peak anymore at the same ITDs as the sensitivity. This leads to a difference between *observed* BD and *neural* BDs.

This is shown in Fig. 8.1, **B**, for a population of $N = 2$ cells. The normalization term (black curve), is constant for all cells, but depends on the ITD in the input. For ITDs between the two BD^*, its value is strong because both cells respond to those ITDs strongly. The response of a cell close to the maximum of the normalization (here 0.75 ms) is attenuated strongly to keep the population response constant. On the opposite, the response away from this maximum is amplified. In our simplified example, this effect is extreme because the normalization curve almost vanishes for ITDs bigger than 1.5 ms.

As a result of the divisive normalization, the observed BD BD_{obs} for the red cell is equal to 1.5 ms, 50 % bigger than the neural BD. In fact, it can be shown that BD_{obs} is always shifted away from the value of the neural BD, in the direction corresponding to a decrease in normalization value.

In conclusion, each cell is the most active of the population for a delay corresponding to its *neural delay* BD^*, yet this delay may not be the maximum of the noise delay curve of the cell (the *observed BD*). This means that care must be taken in the interpretation

of BDs observed in noise-delay measurements. We now turn to the analysis of such a model for more realistic populations of cells comprising many more cells, and for populations responding at different CFs.

8.3 Results

In this section we are interested in the predictions of the normalized model with a realistic number of cells, and as we vary the CF of the population of cells. We simulated two populations consisting of N=100 neurons of the same CF, one at CF = 300 Hz and the other at 2000 Hz. The cells had neural BDs BD^* linearly distributed between -350 and 350 μs. These values roughly correspond to the range of axonal delay differences observed in the cat [Joris and Yin, 2007]. Each cell responded as Equation 8.2, the activity was normalized so that the population activity for a given ITD summed to one.

From this population of cells, observed BDs were recorded according to the definition Equation 8.4. Local maxima are computed over a range of ITD spanning -1/CF, 1/CF, that is two-periods of the stimulus. Indeed, because the sensitivity curves are periodic, it is expected that the curves exhibit maxima separated by the period 1/CF. We pick the lowest maximum, i.e. the closest to zero in absolute magnitude, that yields BD_{obs}.

8.3.1 Behavior of the model as a function of CF

The results of the model are shown on Fig. 8.2, for two CFs in the case of a cat. A cell's sensitivity (in Equation 8.1) depends on the CF of the population. In low frequencies (**A**), the curves are periodic, but the period (2 ms) is big, and hence over a wide range of ITDs the sensitivity curves do not have multiple peaks. Consequently, the normalization curve as a function of ITD has a unimodal behavior on this range of ITD. For negative and positive ITDs close to the maximal value, ITD $\pm$ 300μs, the normalization decreases as the absolute ITD increases. Therefore, we expect that the observed BDs are shifted to bigger absolute values by normalization.

On the other hand, in high frequencies (CF = 2000 Hz, **B**), each sensitivity curve has multiple peaks over the physiological range of ITDs (here only plotted within the π-limit). Consequently, the normalization curve is much flatter than in the low frequency case. Furthermore, the normalization rather increases with increasing absolute ITD. Therefore, we expect that the observed BDs are shifted to smaller absolute values by normalization (see Fig. 8.2 **A**, **C**, **E**).

We check that those predictions are met, by computing in the two cases the normalized responses of each cell. This is plotted in Fig. 8.2, **C** and **D**, for the two CF

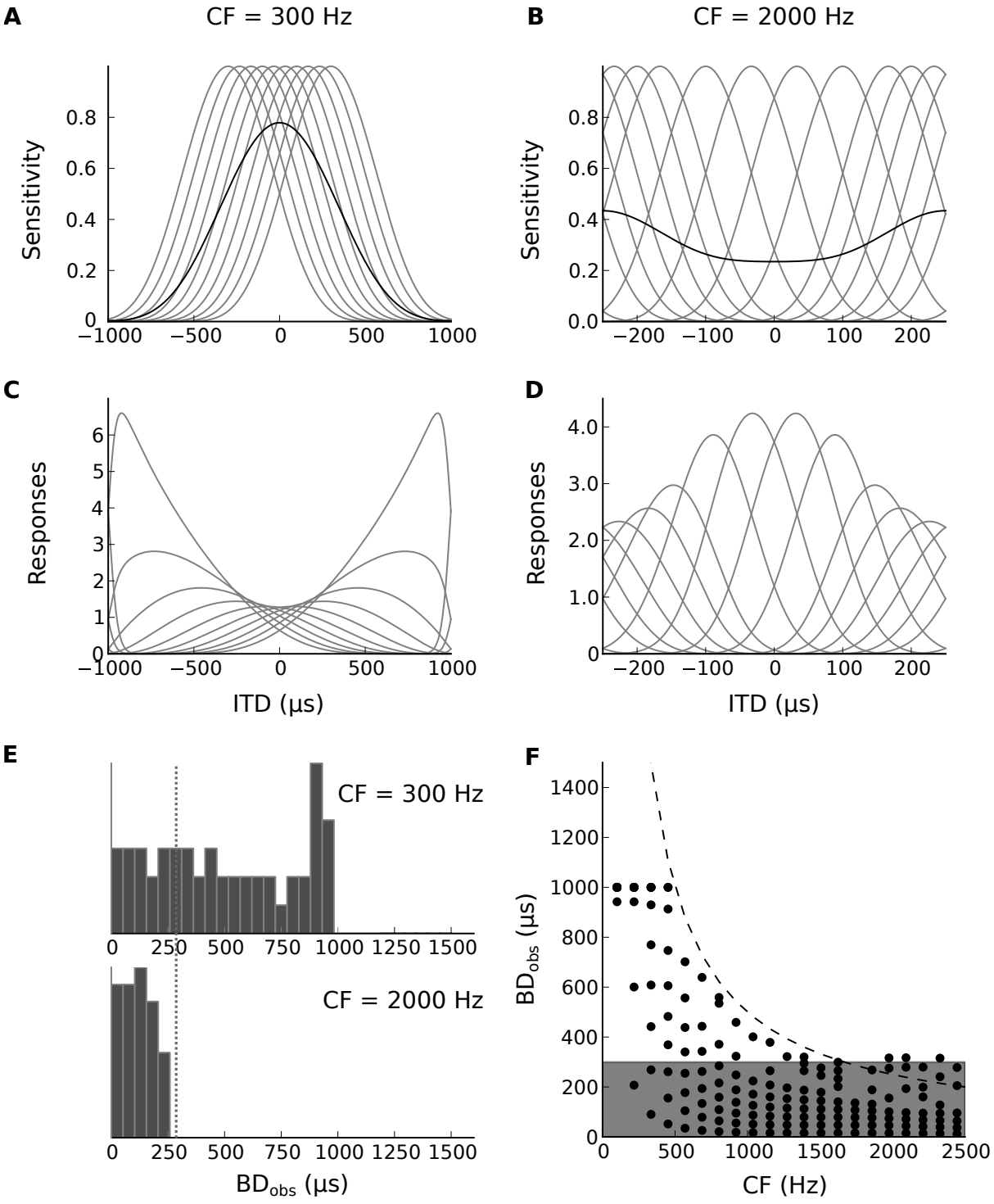

Figure 8.2: Behavior of the peak code with normalization model. **A** Sensitivity curves and global normalization as a function of ITD, for a population with $CF = 300$ Hz. **B** Sensitivity curves and global normalization as a function of ITD, for a population with $CF = 200$ Hz. **C** Response of the cells, computed from **A** with $CF = 300$ Hz. **D** Response of the cells, computed from **B** with $CF = 2000$ Hz. **E** Histogram of the observed BD distributions in low frequencies, high frequencies, and the distribution of neural BDs BD* **F** Observed BD for different CFs of the population.

cases. As a result of the normalization, the maxima of the response curves in the low frequency case (**C**) are shifted away from the physiological range. Furthermore, the range of response values is increased, and the maximal response for each curve can vary 6-fold in this model. This heterogeneity of responses is observed in electrophysiological recordings, where the response of MSO cells to stimuli presented at their best ITD varies within ten and several hundred hertz. Within the physiological range, it should be noted that the maxima of each curve is still attained at its neural BD*. In the high frequency case, the peaks of the responses (**D**) are not noticeably shifted away from the peaks without normalization (**B**). The curves are sitll heterogenous though, with the curves for extreme ITDs having on average a smaller response rate than those with BD* close to 0.

Computing the histograms of the observed BDs over cells with positive BD* for the two CFs (**E**), we observe that indeed, in the low frequencies observed BDs are shifted away from the physiological range, whereas in high frequencies the distribution is more concentrated around zero. This is also confirmed for more CFs, as reported on **F** for 30 CFs linearly spaced between 100 and 3000 Hz. In the low frequencies, the distribution of observed BDs is bimodal, with peaks well outside of the physiological range. As the CF increases, the BD gradually shifts back within the physiological range.

This in good agreement with measures of BDs in the cat, which exhibit large observed BDs in some low-CF neurons [Joris and Yin, 2007]. Yet, one should keep in mind that cells' neural BDs (BD*) are still in the physiological range. That is, those extreme BD values are not the result of postulating correspondingly long neural delays. In fact, the delays chosen here are in agreement with delays as derived from axonal reconstruction [Beckius et al., 1999].

8.3.2 Predictions for different species

In the previous section we have been working with a fixed maximal BD* corresponding roughly to the case of a cat. Yet, the results of the previous section crucially depend on this parameter. Indeed, the normalization term becomes flatter as the CF increases, because multiple periods of the sensitivity curves overlap. The CF above which this phenomenon occurs depends on the maximal ITD. Consequently, it is possible to obtain distributions of BDs for different species, by varying the maximal ITD.

The results of such simulations are reported in Fig. 8.3. Two head sizes are considered, in one case we use a maximal ITD of 300 μs, roughly corresponding to a cat. In the other case, the maximal ITD is 125 μs, corresponding to the case of the gerbil. On both panels of Fig. 8.3 the same trend of BD_{obs} distributions is broadly obtained.

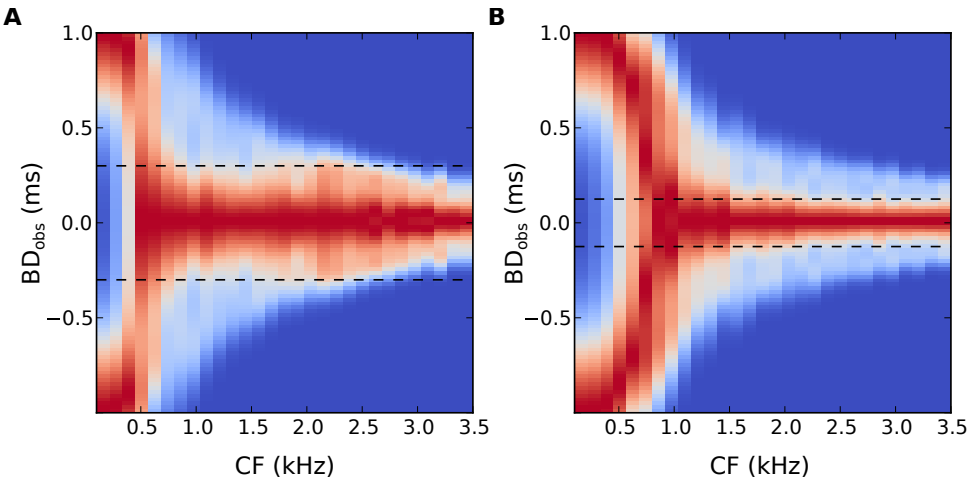

Figure 8.3: Result of the model for two species (maximal $|BD^*|$). Each panel represents the probability distributions of observed BDs at all CFs. Each column represents the distribution of BDs at this CF, color coded. Each curve is normalized to a maximum of 1 so that the modes appear clearer. **A** Case of the cat with a maximal ITD of 300 μs **B** The case of the gerbil with a maximal ITD of 125 μs

In low frequencies the observed BDs are shifted well outside of the physiological range, but not in high frequencies. The difference arises in the CF at which the BD distribution stops being bimodal with two modes at extreme BDs. This frequency seems to decrease with maximal ITD, as previously expected. Therefore, large BDs at low frequencies are predicted to occur over a broader range of frequencies for smaller mammals. This is consistent with published data that shows that in the gerbil, large BDs are still observed for quite large frequencies [Pecka et al., 2008]. In the cat, this effect seems rather restrained to much lower CF cells [Joris and Yin, 2007].

8.4 Discussion

8.4.1 On the normalization mechanism

The model presented here provides a theoretical proof that the observation of large BDs in low-frequencies is not inconsistent with an efficient peak code in the physiological range. It is based on realistic neural delays, but the question remains of whether a normalization mechanism of the kind that is postulated here is biologically plausible. Functionally, such normalization could arise from homeostatic constraints: the energy budget allocated to the population is constant, irregardless of the stimulus ITD. Unfortunately, the way such a mechanism could be implemented is not clear.

An interesting recent result on the physiology of binaural neurons in gerbil MSO provides a candidate mechanism for this normalization. The mammalian MSO receives contralateral glycinergic inhibitiong through the MNTB [Spangler et al., 1985] and ipsilateral from the AVCN through the LNTB [Cant and Hyson, 1992]. This inhibition is highly temporally precise [Grothe and Sanes, 1994], and it was long unclear what its precise role was. By blocking the inhibition in an *in vivo* model, [Brand et al., 2002, Pecka et al., 2008] showed that the observed BDs lie outside of the physiological range only when inhibition is present. Together with the insights about the hemispheric model [McAlpine et al., 2001], it was then argued that this precise inhibition was a prerequisite of a robust ITD encoding in mammals [Grothe et al., 2010, Pecka et al., 2008]. In our context, glycinergic inhibition is an interesting candidate for the implementation of the normalization mechanism. Indeed, the same shift in observed BD is predicted (as shown [Brand et al., 2002, Pecka et al., 2008]): when inhibition is blocked the BDs are inside the physiological range. Our model offers additional insights: only in the absence of inhibition does the observed BD reflect the difference in neural delays leading to the neuron under consideration. It must be noted that very recent results challenge the view that temporally-precise inhibition shapes the ITD sensitivity of MSO cells. [vanderHeijden et al., 2013] argues that such a mechanism requires that the inhibition has an effect that depends on the input phase at both ears. In their gerbil animal model, no influence of this sort is found, suggesting that inhibition is in fact not precise enough to significantly affect an MSO cell's ITD sensitivity. [vanderHeijden et al., 2013] finally suggest that the difference between their data and the previous data from [Brand et al., 2002, Pecka et al., 2008] may be due to their use of extracellular recordings or different anesthetization techniques.

Direct evidence for a population-wise normalization in MSO is not available in the literature, because recording from this structure is notoriously hard [Yin and Chan, 1990]. Such data might be more easily recorded in IC which may afford easier multi-electrode recordings. Yet, the function achieved by those neurons may also be different from MSO (e.g. effect of ILDs). Nonetheless, some insight can be found in fMRI experiments that provide coarse but population-wise measurements of brainstem activity. In such an experiment on human subjects [Thompson et al., 2006], the activities of the brainstem on each side is modulated by the ITD of the stimulus Because the summed activity of all the cells in *the two* lateral MSOs is normalized (instead of across one MSO), then it is each MSO's population activity is modulated by ITD.

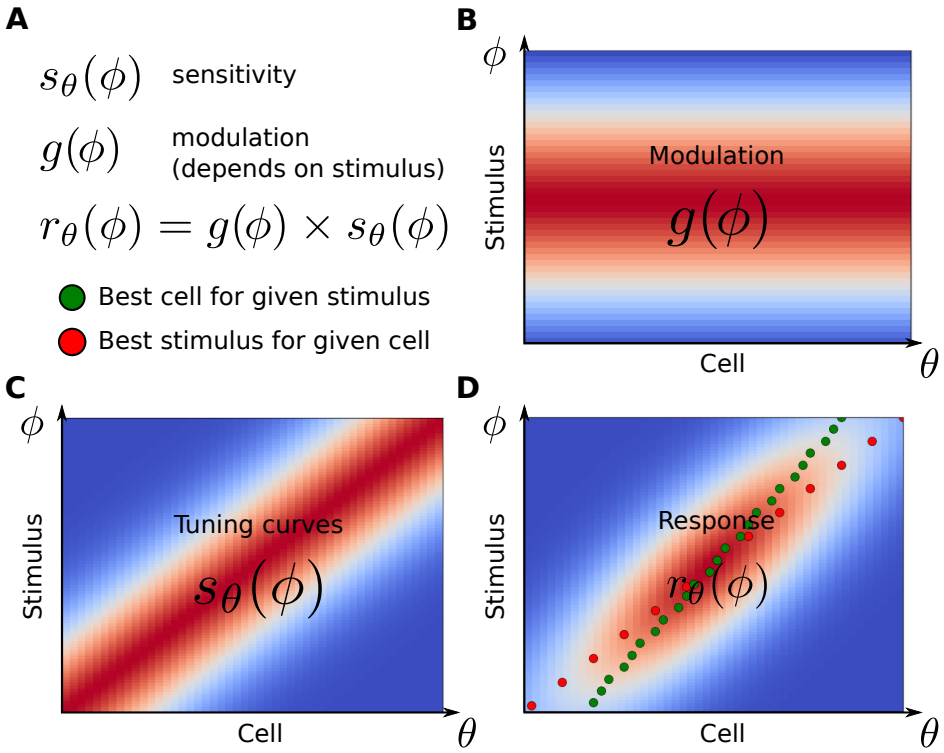

Figure 8.4: A more general situation. **A** For a given stimulus ϕ, the responds of all cells θ is modulated by the same value, the modulation **B** that does not depend on the cell. The tuning curves of the cells are ordered by best stimulus on **C**, and the response of all cells is represented on **D** alongside the maxima along the two dimensions (cell for given stimulus, stimulus for given cell). Cells are not most active at their best stimulus **D**.

8.4.2 Interpretation of sensitivity data

Because of experimental constraints, data from MSO is gathered for a small number of cells, presented with a very large number of stimuli. In a typical experiment, about a hundred cells are recorded in turn, possibly on different animals, and over the course of weeks. For each cell, data is aquired for many repetitions of a wide range of stimuli, e.g. with varying ITD. Therefore, the *sensitivity* of individual binaural cells *across stimuli* is very well documented, whereas the *population response* for any given ITD is not. A difficulty arises when making sense of this data in terms of coding. Indeed, observing sensitivity data is inherently different from what the system itself is doing. Indeed, the only accessible data for the neural system is the observation of *all the cells* presented with *one* ITD (for a single source). That is to say that the collective behavior of the cells is really relevant to the system. The framework developed here shows that care must be taken when interpreting single-cell data. Even in a peak code, a cell is not necessarily the most active of the population when the stimulus is its *best stimulus*.

The example peak code provided above treats the case where the cells are modulated by the summed activity. But in fact, we can show that this happens whenever the cells' responses are modulated by a function of the stimulus, that is constant for all cells. On Fig. 8.4, this behavior is shown in the case where neurons respond as the product of their *sensitivity* (which depends both on the stimulus dimension ϕ and the cell's best stimulus θ), and a global modulation term that does not depend on the cell's best stimulus. This modulation, in particular, could depend on other dimensions of the stimulus. This is shown on **A** of Fig. 8.4, where the equation for the schematic population of cells is shown. For each cell and possible stimulus dimension, **B** and **C** show the contribution of the sensitivities ($s_\theta(\phi)$, **C**) and the global modulation ($g(\theta)$, **B**) terms. On **D**, the response of each cell to each stimulus dimension is computed and reported, also, dots indicate the best cell for a given stimulus, that is the read-out cell of the population when presented with a given stimulus value, and the best stimulus for each cell (according to the color code in **A**).

Similar conclusions can be drawn from this schematic results: the most active cell of the population for a given input dimension ϕ is not necessarily attained when it is equal to its best stimulus. There is an inherent duality between responses across cells and stimuli. If the experimentalist typically has access to some data *across stimuli*, the nervous system (and theoreticians) really are only interested in the *response of the population* (that is, across cells) to any given input. Because the response of the cells can be modulated by other dimensions of the stimulus

This general feature is the crux of the argument of this chapter. In the context of MSO neurons, observed in vivo noise-delay curve recordings are highly heterogenous

(e.g. [Yin and Chan, 1990] in the cat). That is, the maximal spike frequency, measured for example as the magnitude of the peak of the noise-delay curve, is highly variable from neuron to neuron. Overall this fact suggests that it is more likely that a hidden global modulation exists than not.

Part III

Behavioral integration frequency-dependent ITDs

As the results of Chapter 1 show, a sound emitted by a source in space will reach the ears with different ITDs depending on its frequency content. Despite this observation, already predicted more than a century ago, the variation of ITD for a given position was ignored. This poses a logical problem, which can be formulated as follows. Assume that humans integrate ITD cues such that a given ITD is mapped to a position irregardless of the frequency content of the signal. A prediction of this is that the lateralization percept is the same for all signals that have the same ITD. Assuming that there is no ILD, free-field sound sources in high-frequencies are lateralized to more central positions than low-frequency ones. That is, subjects necessarily misinterpret the lateral position of sound sources in real environments, which this seems intuitively inconsistent with intuition.

In this part of the manuscript I am interested in studying the consequences of the variability of ITD across frequencies on the localization behavior of human subjects. First, in Chapter 9, I review classical and recent results on the psychoacoustics of sound localization. This is done with an emphasis on binaural cues, and more specifically the ITD. Then, in Chapter 10, I test the hypothesis that human subjects lateralize sounds with different frequency contents according to the ITD expected from a free field source, i.e. integrating the frequency-dependent nature of the cue. I designed and ran an experiment to test this prediction, in the form of a matching paradigm. Pilot results show that humans perceive sound sources as a function of both the ITD and the frequency content of the signal. Furthermore, the extent to which this happens is consistent with the acoustical data that was presented in the first part of this thesis.

Chapter 9

Introduction to the psychoacoustics of sound localization

Contents

9.1 Free-field localization abilities

Localization of an object through the visual modality is much more precise than through the auditory modality. Yet, the visual field is restrained in lateral extent, while sound source can be localized for all positions around the subject. This fact suggests that audition is primarily relevant for directing the gaze towards events not in the frontal field (i.e. that cannot be seen) [Perrott et al., 1990, Heffner and Heffner, 1992]. Localization sound sources in space is also essential for a number of hearing-related problems: e.g. to separate the sound signals emanating from different sources.

In this chapter, I introduce the most important results in behavioral studies of the localization abilities in humans. I first discuss the perception of the 3D position of a sound source (azimuth, elevation and distance) in free-field situations. When sounds are presented over headphones, the sound is heard inside the head, at a more or less lateral position. I then discuss lateralization experiments that provide understanding in the roles of binaural cues, with an emphasis on ITD. Finally, I discuss the impact of the frequency-dependence of ITD for a given position (shown in Chapter 1) on our understanding of the binaural system.

9.1.1 Perception of azimuth and elevation

In this section I review measures of human perception of space. Using a *threshold* measure, one can measure the ability of subjects to discriminate stimuli on the basis of a variable dimension (here, the position). Such a measure can be obatined in discrimination tasks, and quantifies the *acuity* of human perception of sound location. Another way to measure the system is by asking the subjects to perform absolute localization judgements. In this case, the *accuracy* of the human response is quantified. This can be acheived using, for example pointing tasks.

I first discuss the acuity of the system: the ability of humans to distinguish sounds based on their position. Second I study the accuracy of the system: the absolute localization abilities of humans. In a third part I discussed the perception of egocentric distance to the sound and the impact of reverberation on sound localization.

Threshold measures

Localization sound sources in space is essential for a number of hearing-related problems: e.g. to separate the sound signals emanating from different sources. This is generally known as the "cocktail party" problem. In many common situations (not restricted to fancy social gatherings), humans *hear* many concurrent sound sources,

emanating for example from different speakers around them. Despite those distractors, humans are able to only *listen* and understand a single selected sound source. It can be shown that discriminating the speakers positions is essential in solving the cocktail party problem [Hawley et al., 2004].

Discrimination abilities are usually assessed via *threshold measures*. An example measure is the Just-Noticeable Difference (JND), that is the smallest variation of the studied dimension that can be perceived by subjects. Those measures provide a good idea of the *precision* of the system: stimuli whose dimension differ by less than one JND are not distinguished.

In the context of binaural hearing, one can measure the JNDs in all attributes of the spatial position: azimuth, elevation and distance to the sound source. For angular measures this is measured in Minimum Audible Angles (MAA) [Mills, 1958], that is the smallest angle between two sound sources that can be resolved by a subject. Human JNDs can be as low as 1°in azimuth for low-frequency sounds around the midline [Mills, 1958, Blauert, 1997]. Yet, this performance is not uniform either in space and frequency. The MAA as measured around the most excentric positions (at azimuths $\pm 90°$) is bigger $\approx 8°$ [Blauert, 1997]. It does not appear, though, that the elevation of the sound source significantly degrades azimuth localization performance [Perrott and Saberi, 1990]. At some frequencies, between 1 and 2 kHz, the thresholds are higher [Mills, 1958, Perrott and Saberi, 1990], indicating that the acuity of human localization is not completely invariant on frequency. Similar studies can be conducted to assess MAAs in *elevation* of the source. Human elevation-localization abilities are in general coarser than in azimuth, of about 3.5°around the horizontal plane, and increasing as the sources is elevated from this plane [Perrott and Saberi, 1990, Blauert, 1997].

Absolute localization error

Human localization abilities are also the subject of a number of biases and complex errors that are not captured through threshold measures. For example, sources that are placed symmetrically in the front and back hemifield are often mistook for one another, this effect is usually known as the front/back ambiguity [Blauert, 1997]. More generally, if one assumes that the subject has a spherical head, with ears positioned on the diameter, then sources are theoretically indistinguishable when they are presented on cones with their axes aligned on the interaural axis: the so-called *cones of confusion*.

Ambiguity between front and back may be resolved in a variety of ways. In usual situations, vision can help confirming that a sound source is present in the frontal hemifield. Even in the absence of vision, movements of the subject (and not the sound source) also help alleviate the confusion [Wallach, 1940]. Indeed, the dynamic varia-

tion of cues is different for static sources in the front and in the back hemifields. By encouraging a subject to move its head or restraining the subject's heads [Wightman and Kistler, 1999], it can be shown that self-generated movements very efficiently allow front/back disambiguation.

9.1.2 Perception of other attributes, impact of reverberation

When a sound source is present in a real environment, it will reflect on the big enough surfaces, but also be scattered by objects in the environment that have smaller sizes. This represents a considerable issue for sound localization, and most common algorithms fall short of properly localizing sources in arbitrary echoic environments. It appears that humans are able suppress the reflections, and accurately perceive the location of the sound, almost irregardless of the environment [Hartmann, 1983]. Yet, it is not clear how this can be done, because listeners do not have a complete knowledge of their acoustic environment at all times.

Therefore, it is quite surprising that human localization abilities are essentially unaffected (if not facilitated) by the presence of arbitrary reflections [Hawley et al., 2004]. Artifical reflections can be introduced in binaural sounds, by presenting a sound together with the same sound with a very short delay. This simulates the additional path-length the sound has to cover when it reflects on an object before reaching the ear. Humans consistently report hearing only one sound when this delay is sufficiently short (between 3 to 5 ms, depending on the sound). This phenomenon belongs to the so-called precedence effect [Litovsky et al., 1999]. When the delay between the two sounds is increased, humans hear two distinct sound sources. It is thought that humans are able to suppress the response to the second sound, thus being quite insensitive to the clutter of information due to reflections.

Reflections are also important to judge the egocentric distance to a sound source. The relative distance between sound sources is judged using the difference in pressure level due to absorption of the acoustic energy by the air. This ability is greatly reduced when this relative level cue is removed, suggesting that distance perception via the auditory modality is coarse [Ashmead et al., 1990]. In fact, when the environment is anechoic [Nielsen, 1992], subjects are unable to judge the distance to a sound source. In all other reflective environments, humans are able to report the distance to the sound source. This is believed to rely on the processing of the ratio of the direct signal from the source and the reverberated energy from reflections [Hartmann, 1983].

9.2 Lateralization: ITD and ILD cues

The existence of two separate acoustical cues to location had also been proposed by Lord Rayleigh's work [Rayleigh, 1907], and then [Hartley and Fry, 1921]. Lord Rayleigh first realized that for most broadband and high frequencies sounds, varying the relative levels of the monaural signals (i.e. applying ILDs) resulted in lateralization of the sound source. He reasoned that because ILDs were absent in the lower frequencies, because of the physics of the head, some other cue was used in this range. By presenting subjects with tones at slightly mismatched frequencies (binaural beats), he was able to show that in low-frequencies humans are sensitive to time differences in the monaural inputs. Therefore, in low frequencies the cue to location is the ITD. This theory, postulating different roles for ITDs and ILDs in high and low frequencies, became known as the Duplex Theory. This view has since then been confirmed by many experiments, providing a good theory as to the way different binaural cues interact.

Yet, when presented over headphones, stimuli reproducing ITDs and ILDs are in general perceived in a virtual line between the ears, inside the ead. This kind of percept is generally known as *lateralization*. For reasons that are not completely clear to this day, reproducing the perception of sounds outside a listener's head (externalization) over headphones is hard, even when the full HRTF filters are reproduced. Nonetheless, the use of artificial stimuli, with only ITD or ILD imposed on them is a good way of probing the mechanisms by which the binaural system recovers the position of a sound source.

Using modern HRTF measurements, it is possible to quantify the information that are available to a subject when sounds are presented in free field. The cues fall in two general categories, *monaural* cues which we have left aside for most of the manuscript, and *binaural* cues (ITD and ILD). Here I report some fundamental results as to the perception of all those cues. Starting from monaural elevation cues, to ILD and finally a strong emphasis on ITD and its interplay with ILD.

9.2.1 Monaural cues and ILD

Monaural cues

The filtering imposed by a monaural HRTF filter shows amplitude variations, such as notches and peaks, which depend on the source position: these are termed *monaural cues*. Those cues are incomplete in that they inherently require the knowledge of the original signal to extract the position: i.e. the spectrum of the source signal and the contribution of the monaural filter are mixed in the signal at the ear. Therefore, it

is *a priori* impossible to extract the influence of the filtering of the head without substracting the original signal's contribution.

Despite this limitation, monaural cues are important in the cases where binaural disparities are absent. For example when a sound source is present in the median sagittal plane: the plane such that each point on the plane is at the exact same distance from both ears. In this plane, the interaural differences are absent (assuming that the head is symmetric), or in any case very small. Another example is when one ear is occluded, or the stimulus is presented at only one ear (e.g. over one headphone side). This arises in some common listening situations, but also for deaf people (as studied in [Wanrooij and Opstal, 2004]).

When narrow band sounds, such as tones, are presented in space, subjects with an occluded ear are unable to locate the sound source, despite having a consistent judgement about the position [Butler, 1971]. [Butler and Helwig, 1983]. In fact, monaural cues only influence lateralization for broadband signals, with energy above 4 kHz (see, e.g. [Musicant and Butler, 1985]). In the horizontal plane, [Oldfield and Parker, 1986] the lateraliaztion of broadband sources, presented monaurally is very coarse and much degraded with respect to when binaural information is available. In fact, when sounds are presented binaurally, monaural cues only have little influence on the perceived azimuth of the sound source [Macpherson and Middlebrooks, 2002].

In the medial sagittal plane, narrow bands of noise presented binaurally perceptually originate from different directions (front or back), depending on their center frequency [Butler and Helwig, 1983]. For low-frequency sounds in this plane, spectral features resulting from the head and torso reflections are are also important to resolve the elevation of the source [Algazi et al., 2001a].

Therefore, available results suggest that the perception of position through monaural cues is: limited to positions where binaural information is unabailable, and encoded via the complex pattern of monaural HRTF amplitude, in most cases above a few kHz.

ILD cues

Interaural level differences, ILDs, are also very important binaural cues to sound position, which were left aside for the most part of this thesis. When presented over headphones, ILDs also produce a lateralized percept.

The acuity of the system can be probed by similar measures as the JND presented in the introduction. Reports of JNDs in ILD on various sounds consistently report values between 1-3 decibels [Yost and Raymond H. Dye, 1988, Mills, 1958]. This JND varies with a number of factors, first with the reference point, that is around which lateral position the JND is measured. As this lateral position increases, the JND also

increases [Hartmann and Constan, 2002, Francart and Wouters, 2007]. The JND also depends on the nature of the stimulus, namely broader frequency bands lead to smaller JNDs [Buus, 1990]. As a function of the center frequency of the stimulus spectrum, the JND in ILD seems to be roughly constant over a broad range of frequencies [Yost and Raymond H. Dye, 1988], with a slight degradation around 1-2 kHZ. Recent data are consistent with this, even though some argue that the JND is reduced in higher frequency stimuli [Francart and Wouters, 2007, Hartmann and Constan, 2002]

9.2.2 Psychophysics of human ITD processing

Perception of timing differences at different frequencies

Threshold measures and limit of ITD processing In general, human JNDs for ITD on broadband sounds are of the order of 10-50 μs depending on the experimental setup. On a small number of subjects, [Mills, 1958] reports an ITD threshold as low as 10μs, and more recent studies put this measure at around 20 μs [Mossop and Culling, 1998, Akeroyd, 2006, Yost, 1974]. These threshold measures, however, are strongly dependent on the stimulus, e.g. on its center frequency [Dunai and Hartmann, 2011].

ITDs when presented in pure tones (i.e. IPDs) are not always lateralized. Rather, in high frequencies pure tones with arbitrary IPDs do not produce a sense of lateralization anymore. This upper limit of IPD sensitivity has long been measured, leading results with a high variability, albeit contrained to the $<$ 3kHz range. Some early data, [Wever, 1949] report lateralization for tones up to 3-5kHz (citing a unlocatable thesis), while most other (and more recent) studies report limits in the range 1-2 kHz [Perrott and Nelson, 1969, Yost, 1974, Zwislocki and Feldman, 1956, Dunai and Hartmann, 2011].

There are two hypotheses to account for this limit. The first one is physiological: this limit arises from a frequency limit in the coding of temporal fine structure. This argument is made first by J.C.R. Licklider in [Licklider et al., 1950], who in a very concise and intuitive paper describes that because of the fixed temporal resolution of the brain (defined as the EPSP time constant), ITDs in high frequencies are not resolvable. In this hypothesis, the behavioral IPD perception limit reflects the upper limit of time-locking in the auditory nerve (discussed in Chapter 5). In humans, the precise limit of phase-locking is unknown and the subject of much debate ([Joris and Verschooten, 2013]), because of its huge implications on, e.g. pitch perception.

Another hypothesis is that IPD cues are ambiguous above a certain frequency. A tone of period T delayed by τ is indistinguishable from the same tone delayed by $\tau + T$ Therefore, when multiple periods of the stimulus "fit" within the ecological range of ITD, then the ITD information is ambiguous. As an example, for pure tone sounds this

happens whenever $T < \text{ITD}_{\max}$, or $1/f < 2a/c$ using the Woodworth formula which amounts to around 1.7 kHz for a human subject. This hypothesis is still the subject of much discussion because it does not seem to match the precise frequency-dependent pattern of JNDs in ITD (as argued by [Hartmann et al., 2013]).

Envelope and FS ITDs When the stimulus is a high frequency pure tone (above a few kHz), it is known that humans cannot resolve ITDs, consequently such tones do not produce a sense of lateralization. Yet, it is still possible that for more complex stimuli, some temporal information is present and possibly processed by the binaural system. This is the case of amplitude-modulated stimuli, that is tones modulated by a lower-frequency envelope. The fine structure ITD, usually termed *ongoing* ITD is carried by too high frequency components to be resolved. Nonetheless, the envelope of the stimulus, a much lower frequency signal, also bears ITD information: the *envelope* ITD. The same argument can be made about the *onset* of the stimulus, because the transient has low-frequency components, it is possible that an ITD be detected in this part of the signal: it is the *onset* ITD.

The fact that humans are able to distinguish sounds with different onset or envelope ITDs has been shown by [Henning, 1974]. In this experiment, the authors use a (binaural) masking paradigm to show that the ITDs in complex stimuli with spectral energy constrained to high frequencies (above 3kHz) can still be distinguished. Furthermore, the threshold measures of sensitivity for envelope ITDs are very close to that for low-frequency sounds [McFadden and Pasanen, 1976]. Yet, the importance of the envelope ITD information as a cue to location is debated. For example, [Eberle et al., 2000] argue that imposing strong amplitude modulation does not improve the free-field localization accuray. Conversely, data in [Macpherson and Middlebrooks, 2002] suggest that when fine-structure ITDs are absent, then onset and envelope ITDs contribute to the lateralization percept. In conclusion, high-frequency envelope ITD cues only contributed to the lateralization when both ILD cues and fine-structure ITD cues (in low frequencies) are absent.

ITD dominance and trading experiments

It is generally accepted that ITD is a more reliable cue to sound location in low frequencies than ILD, because physically level cues are almost absent in this range. As shown in Chapter 1 this is because when the sound wavelength is smaller than the diameter of the sphere, then almost no shadowing occurs. This is reflected in behavior, as was postulated by Lord Rayleigh's Duplex theory [Rayleigh, 1907].

More recent studies corroborate this fact, for example [Wightman and Kistler, 1992] show that ITD better predicts the perceived location as soon as there are low frequencies (below 1.5 kHz) in the stimulus. Therefore, the location of any low-pass filtered stimulus is dominated by the ITD cue. Furthermore, when the stimulus is band limited to the range in which ITDs dominate, the preceived location of the sound is unaffected by ILDs up to 10 dB [Wightman and Kistler, 1997].

Those results can be generalized in the context of what is termed ITD/ILD trading experiments. In such experiments, listeners are presented with conflicting ITDs and ILDs for which they provide lateralization judgements. It is assumed that the judgement varies as a weighed sum of the two cues, and the relative weighing thereof is recorded. The *time-intensity* trade is then defined as the ITD necessary to center a virtual source with 1 dB of ILD (pointing in the opposite direction). In an early study [Harris, 1960], this ratio is measured at 25μs/dB for low-pass filtered sounds (and about 60μs/dB for high-pass stimuli). This measure is sensible to a number of factors, and displays a wide inter-subject variability. The most important of those factors is the *shift*: the position at which the subject has to match the ITD and ILD cues. Moreover, the presentation of stimuli with conflicting ITD and ILDs sometimes leads to the perception of two sepatate sources (one dominated by ITD and the other by ILD) which further complicates the interpretation of trading experiment data [Macpherson and Middlebrooks, 2002].

9.3 Discussion: Perceptual relevance of ITD variation with frequency

In Chapter 1, I showed that ITDs vary with frequency in a manner that is direction dependent. Additionally, there can be several JNDs between the low and high frequency values of ITD for a given position. Essentially, a high- and a low-frequency sound with the same ITD are *not* at the same position. These are good reasons to believe that the frequency variations of ITDs is perceptually relevant to human listeners. Yet, the ITD is most commonly thought of as being a constant value across frequencies for any given position. This is based on the widespread belief that reproducing the fine frequency variation of ITD does not make any difference when presenting spatialized sounds, for example in the context of spatial audio.

The perceptual salience of the frequency dependence of ITDs has been first addressed indirectly by the literature in trying to reduce the computational costs associated with the use of HRTFs. One mean to do so is to implement HRTFs as

their minimum-phase versions with accompanying pure delays [Kistler and Wightman, 1992b,Huopaniemi et al., 1999]. Such an operation distorts the natural frequency variations of ITDs and it was found that these distortions were perceptually noticeable under certain circumstances, rarely met for real HRTFs [Minnaar et al., 1999]. Distortions of the natural variations of the ITDs with frequency resulting from their modelization as minimimum-phase HRTFs and accompanying pure delays are thus thought to be almost imperceptible [Kulkarni et al., 1999, Plogsties et al., 2000, Toledo and Mller, 2008], suggesting that humans are not sensitive to the frequency variations of ITDs.

Other studies faced the problem more directly and tried to find which perceptual impact the frequency dependence of ITDs can have on several tasks. [Hartmann and Wittenberg, 1996] does not find a dependence of externalization on dispersion around the head, and thus on the frequency variations of ITDs. The conclusion that can be drawn from this result is that the frequency-dependence of ITDs does not help to externalize sound sources. Yet, given that no headphone reproduction technique conclusively leads to externalization, those results cannot merely be argued to rule out the importance of ITD variation. On the other hand, some studies hint at a potentail relevance of frequency-dependent ITDs. For example, vowels presented in separate spectral bands, they be grouped [Culling and Summerfield, 1995] have shown that listeners were unable to separate noise in different frequency bands on the basis of a constant ITD.

Other results furthermore showed that virtual sounds synthetized using real HRTFs (*i.e.* with frequency dependent ITDs) or HRTFs having a linear phase (*i.e.* with frequency-independent ITDs) were perceptually indistinguishable [Kulkarni et al., 1999], [Constan and Hartmann, 2003], and [Breebaart et al., 2010]. Nevertheless, the methodology consisting in perceptually comparing signals having frequency-independent ITDs to ones having frequency-dependent ITDs does not address directly the issue at hand here. This is because those measures assess the discriminability and/or identification of sounds with different ITDs. Indeed, consider that one position in space is perceptually encoded, but rather by a frequency-dependent ITD curve, spanning multiple ITD values The prediction of this is that the perceived location of sounds with the same ITD in different frequency bands varies. This would appear as *bias* in lateralization, which is out of the scope of discrimination studies [Hartmann and Wittenberg, 1996], [Kulkarni et al., 1999], [Constan and Hartmann, 2003], and [Breebaart et al., 2010].

Therefore, I argue that the results in the literature cannot completely rule out the perceptual salience of the variation of ITD across frequencies. I suggest that further experimental work based on a different methodological approach is needed to provide conclusive evidence as to the impact of the variation of ITD. Namely, absolute localiza-

tion judgements for sounds with different frequency contents and different ITD. Those, I hypothesize, should exhibit a *bias*: lower frequency sounds are perceived more centrally than high frequency sounds for the same ITD. Such an experiment is designed and performed in the next chapter (Chapter 10).

Chapter 10

Integration of ITD cues across frequencies

Contents

10.1 Introduction

Sounds presented binaurally over headphones with a non-zero ITD result in the subject hearing a lateralized sound. As it is clear from human acoustical data of the first part of this thesis (Chapter 1), the ITD for a given position in space depends on the frequency content of the stimulus. A consequence of this is that a given ITD signals a different position, depending on the frequency of the stimulus. This is reported, assuming a human subject with a spherical head on Fig. 10.1 (radius of 10.3 cm). For any position, the ITD is a decreasing function of the frequency of the stimulus (Fig. 10.1, **A**). Therefore, it is expected that a given ITD signals a more excentric position in high frequencies than in low frequencies. It is possible to quantify this effect by computing the position signalled by a given ITD across frequencies. The result of this operation is reported on Fig. 10.1, for ITDs between 0 and 800 μs. The extreme ITD of 800 μs, presented in a narrowband signal centered around 1kHz arises from a source at 70°, whereas when presented in a signal centered around 300 Hz the same ITD arises from a source at 60°. The shift in position occurs well below 1.5 kHZ for all ITDs (Fig. 10.1, **B**), within the range in which ITD is the dominant cue to azimuth [Wightman and Kistler, 1992]. For broadband sounds, the perceived laterality of a source is a monotonous function of the ITD in the input [] I hypothesize that, for band-limited sounds, the perceived lateral position of the sound source is the closest position that accounts for the ITD observed *in the frequency band of the stimulus*. This operation is performed for different ITDs on Fig. 10.1, **B**.

I designed an experiment to test this hypothesis: assuming that two sounds are presented, with energy within different ranges of frequencies, then they should be lateralized at the same position when presented with *different* ITDs. This difference we call a *bias*, and propose to measure in a matching experiment. At each trial, subjects are presented with two sounds, one of which has a fixed ITD and the other sound's ITD can be parameterized at will by the subject. He also is able to replay the sequence of sounds until he perceives the two sounds at the same position, at which time the ITDs in the stimuli are recorded, and the bias computed. Assuming that humans assign a constant ITD to each position in space, then this bias will be zero. On the other hand, if humans use the ITD variation across frequencies for a given position, then the bias should be a non zero value, of measurable magnitude.

10.2 Methods

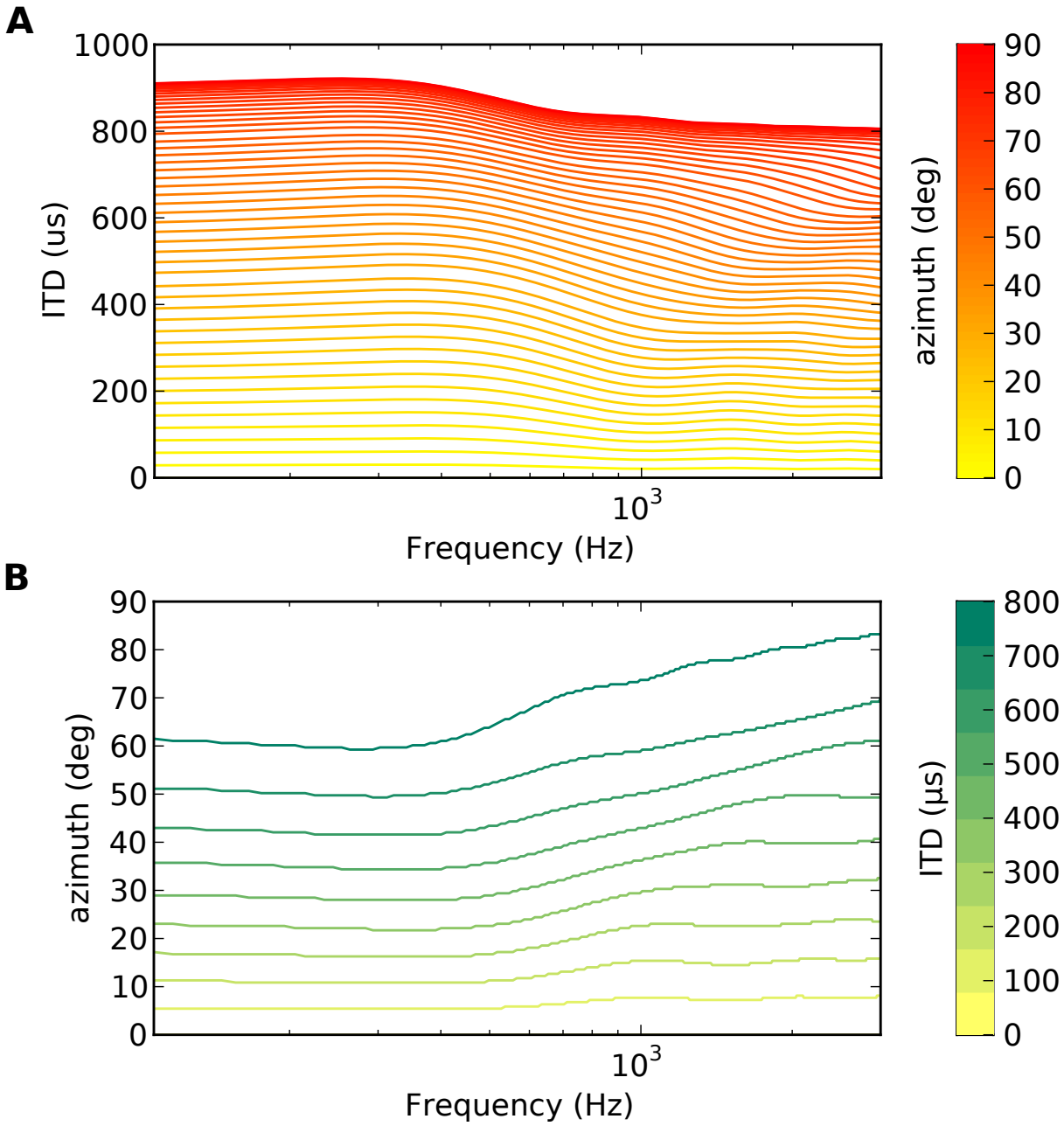

Figure 10.1: ITD as a function of position and frequency, Position as a function of ITD and frequency. **A** ITDs derived from a spherical head model wih head radius 10.3cm, as a function of frequency and position (line color). **B** For any ITD (line color) position as a function of frequency

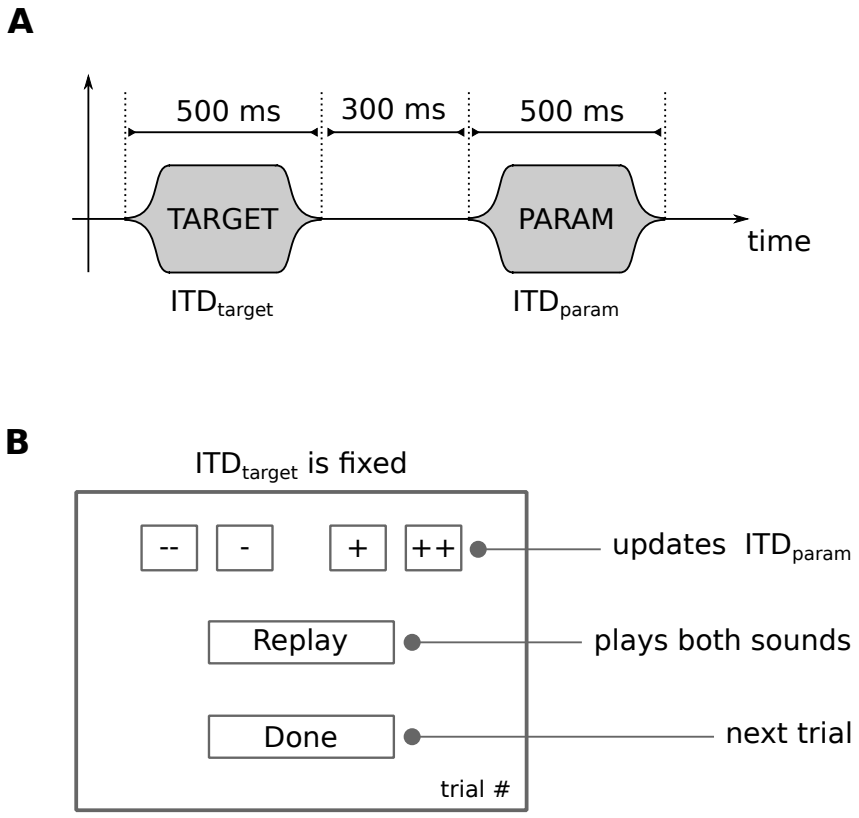

Figure 10.2: Depiction of the experimental paradigm. **A** Each stimulus presentation consisted in two sounds, the TARGET sound with a fixed ITD and the PARAM sound with an ITD that could be modified by the subject. Each sound is a bandpass noise stimulus, in all trials the TARGET and PARAM sounds had different non-overlapping spectra. **B** A schematic of the interface presented to the subject. Within each trial, the ITD of the PARAM sound could be varied in positive (+ signs) or negative (- signs) of two magnitudes (++/– and -/+). The subject could replay the sequence of sounds in **A** at will. He/she was instructed to click "Done" when the two sounds were in the same lateral position. At this point a new trial starts.

10.2.1 Experimental design

Paradigm

We use a matching paradigm to design the pilot experiment, as depicted in Fig. 10.2. Subjects are asked to match the perceived position of two sounds with different frequency contents presented in a sequence (**A** of Fig. 10.2). One of the sounds, the TARGET sound is presented with a constant ITD at each trial. Another sound, the PARAM sound has an ITD that can be adjusted by the subject. The subject increments or decrements the ITD of the PARAM sound, and can then replay the pair TARGET-REF at will (**B** of Fig. 10.2). When the subject responds that he/she perceives the sounds at the same position, the trial ends. At this point the ITD in the PARAM

sound is recorded, and the experiment goes on to the next trial. The main point of the experiment was to use two different noise having energy in different frequency bands, and measure the ITDs for which they are perveived in the same position. The difference between those ITDs is the ITD *bias*.

Parameter	Value
ITD_{LF} (norm.)	2.71, 2.33, 1.69
CF_{LF} (norm.)	0.6 ($\approx 320Hz$)
BW_{LF} (norm.)	0.4
ITD_{HF} (norm.)	2.3, 1.82, 1.21
CF_{HF} (norm.)	1.8 ($\approx 960Hz$)
CF_{HF} (norm.)	1.2
(++/- -) ITD step	200 μs
(+/-) ITD step	20 μs

Table 10.1: Parameters used in the experiment

Stimuli consisted of random phase white noise, further filtered using Butterworth filters to ensure no energy overlap. The main parameters of the experiment are reported on Tab. 10.1. TARGET sounds were generated with ITDs picked so as to maximize the expected bias. Therefore, they were parametrized in terms of *normalized* ITD (as in Chapter 1). For each subject, the ITD is adapted to the head size so that it should produce the same laterality percept in all subjects. This is done by multiplying the normalized ITD value by a/c where a is the head radius and c the speed of sound. The same applies to the center frequencies of the stimuli in the different conditions (reported in Tab. 10.1), which were converted into frequencies by multiplying by $c/(2\pi a)$.

The buttons controlling the ITD of the param sound incremented the ITD in either direction, the magnitude of the smallest possible step was set to approximately one JND, 20μs (for a head of radius $a = 10.3$cm) [Mossop and Culling, 1998, Akeroyd, 2006]. The ++ and - - buttons incremented the ITD with a magnitude of 10 JNDs. Note that the absolute ITD on the PARAM signal was constrained to be smaller than 1 ms.

Three conditions were distinguished :

1. *Frequency content* The TARGET sound can be presented either with low-frequency content, that is with $CF = 0.6, BW = 0.4$, or with a high-frequency content $CF = 1.8, BW = 1.2$. The REF sound is always presented with a different frequency content, high-CF when TARGET is low-CF and vice-versan.

2. *Position* The absolute value of the ITD as presented in the TARGET sound is computed from a spherical model ($a = 10.3$cm, Fig. 10.1).

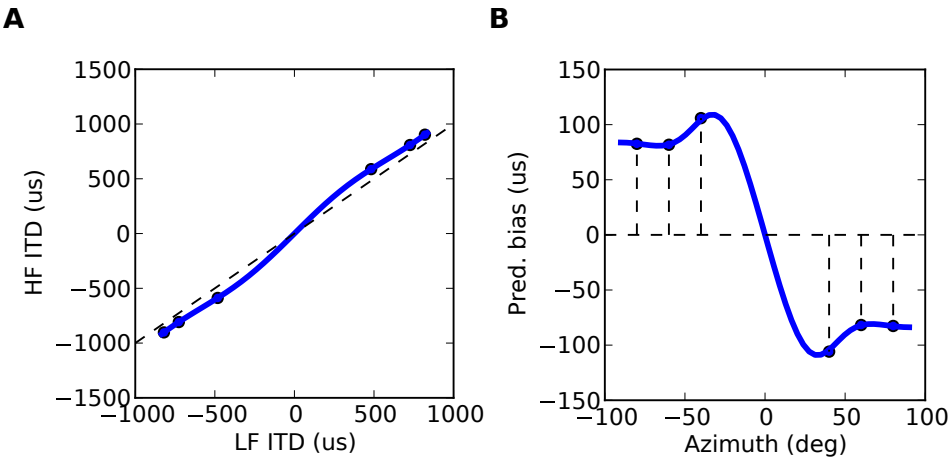

Figure 10.3: Theoretical predictions of ITD matching bias. **A** reports the low-frequency ITD (around 320 Hz) and the high frequency ITD (around 960 Hz) plotted against each other (different points are different positions). The main diagonal is reported (black dotted line) as well as the points chosen for the experiment. On **B** the predicted bias is shown, that is the difference between the LF ITD and the HF ITD. This signed bias peaks for positions around 40 degres, same as in **A** the azimuth points chosen for the experiment are reported.

3. *Side* The sign of the ITD, that is the side at which lateralization should occur in the TARGET sound, could be either left or right.

At the end of each trial, that is at the point where the subject perceives the two sounds at the same lateral position, the ITDs in the two sounds is recorded. I then compute the *normalized ITD bias*:

$$\text{ITD}_{\text{bias}} = |\text{ITD}_{LF} - \text{ITD}_{HF}|$$

We can make qualitative predictions under our hypothesis:

1. *Frequency content* The bias in ITD should be a positive nonzero value.

2. *Position* The magnitude of ITD_{bias} should be a function of position, because the difference between the LF and HF values of ITD is a function of position (see Fig. 10.1, **A**).

3. *Side* The ITD bias, defined as the difference in absolute LF and HF ITDs should not depend on position.

10.2.2 Theoretical model and predictions

HRTF data provides us with estimates of the ITD for a given position in all frequency bands. To make theoretical predictions, such HRTF data was generated on the basis of a spherical model, with a 10.3 cm diameter. ITDs in the frequency bands of the stimuli of the experiment are computed (they are reported on Tab. 10.1). On Fig. 10.3, the ITD found in the LF band and in the HF band are plotted against each other. Were they equal, then all the points would lie on the diagonal (black dotted lines). This is more striking on the difference of the ITDs in those two bands as a function of the azimuth of the sound source. Consistent with observations of Chapter 1, this bias is maximal for intermediate positions around 40-50°. The dots on **A** and **B** correspond to the points in azimuth that were chosen for the experiment so as to maximize the predicted bias. That is, they correspond to the azimuth positions in Tab. 10.1. We compute the difference between the high frequency and low frequency ITDs, to compute the absolute normalized bias, this is our prediction of what bias should be measured in human subjects.

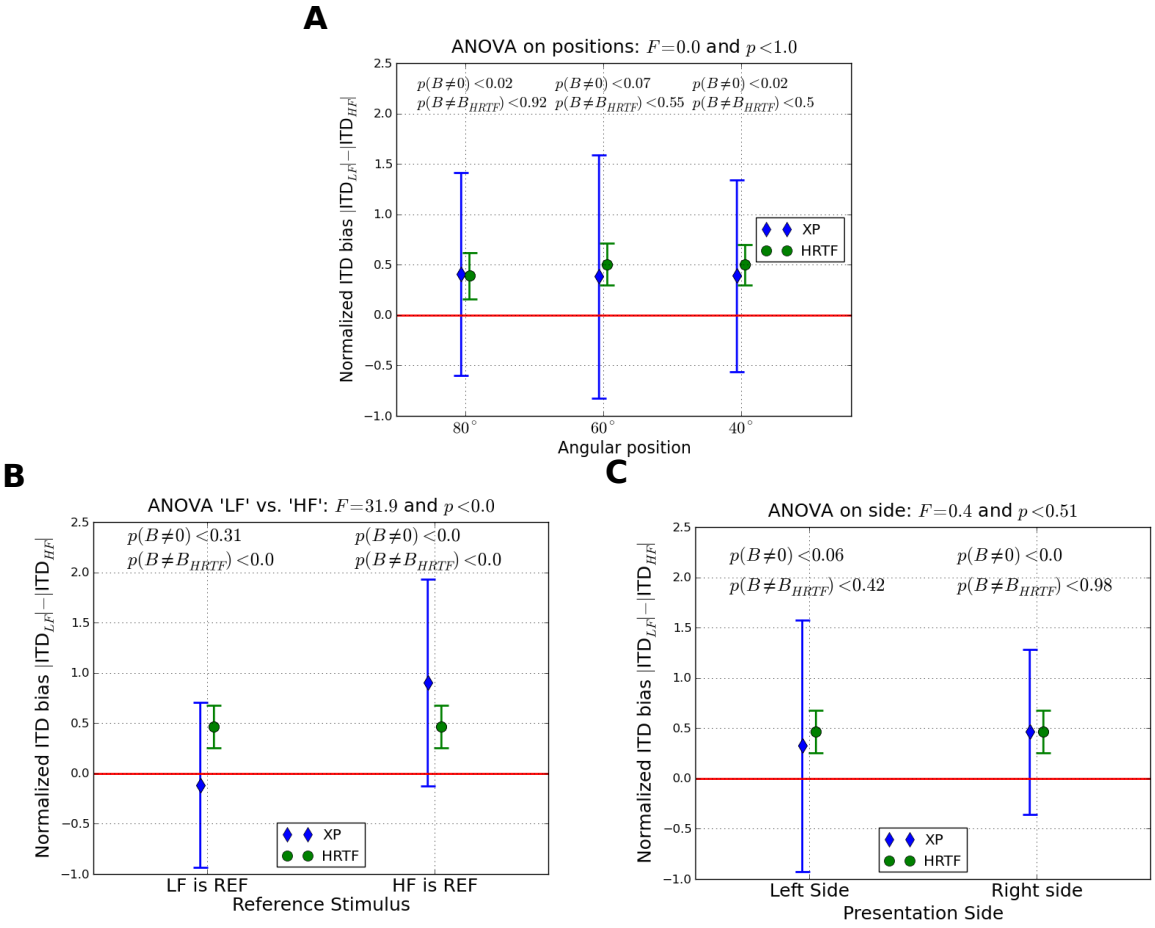

Figure 10.4: Results of the pilot experiment on N=9 subjects, measurement of the ITD bias at the end of the trials. **A** ANOVA on frequency content **B** ANOVA on position **C** ANOVA on side

10.3 Results

10.3.1 Analysis of results for the different positions

Data aggregated along the different position conditions is shown on Fig. 10.4. The magnitude of the predicted bias, as extracted from a large number of HRTF recordings (those of Chapter 1) is shown as green points plus or minus half a standard deviation for all positions. The normalized ITD bias is presented as blue dots, with standard deviation indicated as error bars. Becasue we use the standard deviations, those error bars are quite high, while the standard error would be 3 times smaller.

Results from an ANOVA performed on the ITD bias against the position show that there does not seem to be a significant impact of position on the bias. This is contrary to what is predicted by analysis of the HRTF data, even though the magnitude of the predicted change is quite small (some tenths of microseconds). In any event, for two

out of three positions (80°and 40°) the measured ITD bias is significantly different from zero ($p < 0.05$). A clear tendency is observed in the last (60°) position, even though we are unable to reject the null hypothesis in this case ($p = 0.07$). For all positions, it cannot be rejected either that the average ITD bias is equal to that predicted by the HRTF data. Furthermore, this tendency is confirmed by visual inspection of the results on Fig. 10.4, wherein it is observed that the mean of the bias is always within one standard deviation of the predicted bias.

Overall, our results suggest that there is indeed a measurable bias in ITD, that is sounds are perceived at the same location for different frequency content when presented at different ITDs. Furthermore, this seems relatively well accounted for by the acoustical data.

10.3.2 Other effects

Effect of side

REF and PARAM sounds were presented with either positive or negative ITDs, resulting in a lateralization percept either to the left or to the right of the listeners. We measured the ITD bias as normalized, i.e. the ITD biased is the absolute ITD difference between the low and high frequency ITDs in the data. Thus it is to be expected that this bias does not depend on the presentation side of the stimulus. As expected, the two conditions do not lead to significantly different measured biases.

Effect of presentation order

Stimuli were always presented so that the first sound is the TARGET sound. Therefore, in some trials the first sound was the HF sound, and in others the LF sound. Testing for the effect of presentation order revealed that there is a significant interaction (repeated measure ANOVA, $p < 0.05$). Visual inspection of the results Fig. 10.4 reveals that the ITD bias when the first presented sound contains low frequencies then the observed ITD bias is not significantly different from 0. When the first sound has a higher frequency content, then the ITD biais is significantly different from zero, albeit also bigger than the bias predicted using HRTF data.

This could be understood as an effect of masking of the second presented sound by the first, because of the relatively low interstimulus interval 300ms. The unidirectional nature of this masking effect can be explained as well considering cochlear filters have a broad lower frequency *tail*. That is, when the second stimulus comes in, lower-frequency cochlear filters may still be excited by the REF sound, biasing perception towards the lateral position indicated by REF.

10.4 Discussion

10.4.1 Conclusions of the study

Despite its limitations, our pilot experiment lays the basis for further investigation of the extent to which ITD variations across frequencies contributes to human lateralization. Our initial results suggest that sounds presented with different ITDs at different frequencies can be perceived in the same location, provided that the ITDs they bear are congruent with the ITDs of a position in space. Because the predicted effect is of a quite big magnitude, it is possible to observe it, even amongst a small set of listeners, with a relatively low number of repetitions.

Our results support the idea that humans use all the information about the position that is found in the frequency-dependent ITDs imposed on binaural sounds for a given position. This could seem in disagreement with the results of [Constan and Hartmann, 2003], who argue that humans are insensitive to the frequency-dependent ITD cues imposed by a spherical head. Yet, those results are based on *discrimination* experiments, where the subjects are asked to discriminate between sounds with constant and frequency-dependent ITDs. In particular, the sounds to discriminate have energy in the same bands, i.e. the spectra are not varied. The same argument applies to experiments by [Kistler and Wightman, 1992b, Hartmann and Wittenberg, 1996, Kulkarni et al., 1999]. In this study I used an absolute localization paradigm, to directly address the accuracy of perception of sounds with different spectra. Therefore directly addressing the existence of a bias in the perception of azimuth of sounds with different frequency contents.

10.4.2 Experiment limitations: ILD

There is an inherent flaw in this experiment, due to the fact that in natural conditions binaural stimuli contain both an ITD and an ILD. As described in the previous chapter, this may lead to a modification of the perceived lateralization. In our case, the absence of ILD means that the (inexistant) ILD cue points to more central positions. Because the ILD is expectedly bigger in the higher frequency sounds, then this *centering* effect should be greater on high frequency sounds. Therefore, to compensates for this effect a subject should match high-frequency stimuli with higher ITDs than the low frequency stimuli. Therefore, the absence of ILD in our paradigm should go *against* the effect that we were trying to observe.

Value of such a bias can nevertheless roughly be predicted from the literature. According to Harris *et al.* [Harris, 1960], 1 kHz low-pass click trains need a compensation

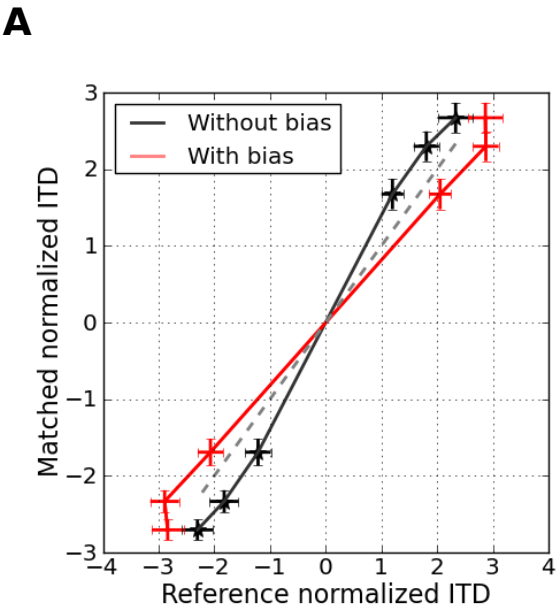

Figure 10.5: Expected results based on the means and standard deviations values as predicted in the previous figures text. **A** shows the expected results in the presence of a bias due to ILD.

of $\Delta T/\Delta I \simeq 25$ μs/dB to be perceptually recentered. From the spherical model, ILD differences of 8 dB, 10 dB and 5 dB are observed between 1 kHz and 1.5 kHz for azimuths of 40°, 60° and 80°. The additional ITD needed to compensate for the absence of ILD difference can be predicted. The results are shown in Fig. 10.5, **A**, where the ITD biases predicted as a function of position and frequency content of the source are shown. The magnitude of the bias predicted in that way is bigger than the effect, predicting that the measured ITD bias with the present analysis would be opposed to what our initial hypothesis implies.

Surprisingly enough, the results of our experiment show that human subjects are not sensitive to this bias, or at least that its magnitude is smaller than modeled in the results presented on Fig. 10.5. This is to be expected, because typical experiments used to measure the weighing ratio of ITD and ILD are measured using sources around the midline [Harris, 1960]. And it is to be expected that this ratio is different for different central positions, that is when the subject is asked to center two sounds at the same position off the midline. Furthermore, this effect is strongly dependent on a variety of parameters, e.g. the sound onset time, or amplitude modulation [Macpherson and Middlebrooks, 2002]. Therefore, it is to be expected that the magnitude of the ITD bias measured using this paradigm is an underestimation of the ITD bias predicted with the HRTF data.

In conclusion, a more thorough approach to this problem will necessarily include

conditions wherein the relative impact of the absence of ILDs is assessed on a per subject basis. Furthermore, our experiment should also include a condition for which the two sounds have equal ITDs, so as to provide a measure of the JND in ITD discrimination at different frequencies, and compare it with the magnitude of the bias.

Discussion

General conclusion

I have shown that, for a given source, binaural inputs have a temporal structure; they are equal up to a non-linear phase filter. This is because animal ears are not point receivers in space, rather they are embedded on a more or less spherical object: the head. Therefore, the cues available to an animal depend, in a predictable manner, on the frequency of the signal and the animal's head size. In signal processing, the problem of sound source localization is often reduced – for analytical ease – to that of localizing using inputs from two microphones in an anechoic environment. I argue that, as compared to this situation, the binaural structure available to animals is much richer. For example, it provides different cues to the location of the sound source in the fine structure and the envelope of the signals reaching the ears. Finally, I exposed how these cues can be quantified from the recording of HRTF filters, and how those filters can be modeled using a spherical head model, or more complex numerical techniques.

In a second part, I have shown that the richness of the ITD cues is reflected in the heterogeneous sensitivity of binaural neurons. More precisely, my results suggest that ITD-sensitive cells of the cat IC can be said to be tuned to a single position in space. Previously, binaural cells were considered tuned to constant ITD, in a way that varied with frequency [Yin and Kuwada, 1983]. My result simplifies this description: when presented with ecological stimuli, the response of the cells can be accurately characterized by the preferred position of the cell. Therefore, the complex pattern of ITD cues imposed by the "acoustical shape" of the animal's head are essential to the neural processing of sound source position. Following on this work, I showed that such cells can emerge from a combination of two properties: differential axonal delays from each side, and mismatches in the frequency tuning of the afferent monaural cells. Finally, I showed that the position of a sound source could be estimated from the activity of a population of neurons tuned to different ITDs with normalized avtivities, in a place code manner. I show how this model is consistent with the large low-CF, high-BD cells found in mammalian MSO, a puzzling anomaly of the field. This type

of code depends critically on the heterogeneous nature of the tuning of binaural cells, as opposed to the recently proposed hemispheric code for ITD [McAlpine et al., 2001]. Overall, my results suggest that heterogeneity of the tuning of binaural cells is finely tuned to the environment of the animal, and contributes to the efficiency of sound source localization.

In the third part, I provided evidence that human subjects are indeed sensitive to variations of ITD across the frequency spectrum for a given position. Therefore, the perceived sound position in a lateralization experiment can be predicted from the variation of ITD specific to the shape of human heads. This suggests that, unlike previously thought, the variation of ITD for a given position across stimuli with different spectra is indeed integrated by human listeners when making lateralization judgments.

On binaural timing cues

The notion of the spatial position of an object should not depend on other attributes of the object. For example, the position of an apple does not depend on its volume, color, or taste. In the very same way, the position of a sound source in space should be defined in a way that is *invariant* on frequency. As the results of the first part suggest, a notion of source azimuth that relies on a single broadband ITD is ill-defined in this respect: a sound source changes "position" as the signal it emits changes. Therefore, a spatial location in acoustical space must be defined with respect to the frequency-dependent cues associated with it.

It could be argued that this notion of spatial position is also ill-defined, as the precise pattern of variation of ITD for a given position depends on other parameters. These include the nature and geometry of the acoustical environment that is known to influence ITD (and ILD) cues [Gourevitch and Brette, 2012]. But they also include attributes of the sound source, e.g. the spatial extent of the sound source, its movement speed, or pointing direction. For most real sources (human speech), those attributes vary and yet we are still able to localize those sources. An interesting question then is to understand what structure in the binaural inputs is invariant on the other environmental parameters. HRTF measurements obtained under many environmental configurations may help understand the – probably complex – nature of this invariant structure. Another problem is that of dynamical cues, how do the cues vary with dynamically changing attributes of the stimuli. Those include movement speed, but also movement direction, which may also dynamically change the apparent size of the source (e.g. sources moving towards the subject).

Neurophysiology of sound localization

As the results of the second part suggest, much of the heterogeneity in a binaural cell's frequency-dependent ITD tuning can be accounted for by using a richer description of the sensory inputs. Because it reflects the complexity of the acoustical environment, this heterogeneity seems to be essential to the robust localization of sound sources. I argue that this is true in two different respects: for single-cell sensitivity to ITD to arise, but also for population-wise coding of auditory space.

First, heterogeneities in the auditory periphery leading to a binaural cell, i.e. axonal delay differences and cochlear wiring mismatches, are shown to be primordial for the sensitivity of single cells to ITD. Those are finely tuned so that cells receive a maximal number of coincident inputs when the sound emanates from their preferred-position, i.e. precisely reflecting the complexity of the acoustical environment. An important prediction of the model of binaural cells that is proposed in Chapter 6, and Chapter 7 is that binaural ITD-sensitive cells are in fact tuned to a *position in space*, invariant on the frequency content of the source signal. This prediction opens a new array of possible experiments that can be performed using free-field methods. This invariance property is remarkable, because it necessarily requires a fine adaptation of the binaural system to its environment.

Second, heterogeneous tuning to ITD cues is a constitutive component of peak code models (or place codes, as opposed to population codes). I provide arguments in favor of place codes, which are shown to be in agreement with the data in MSO. As I show in Chapter 6, a peak code is sufficient to explain the frequency-dependent sensitivity of binaural cells, provided that the complexity of ITD cues is taken into account. Nonetheless, the debate between the place and population codes evoked in Chapter 7, is not settled by this thesis. Conclusive evidence on this topic will necessarily come from simultaneous measurements of the whole population of binaural ITD sensitive cells. Such techniques are available, yet the laminar nature of MSO, and the strong neurophonic potential, make most of them unsuitable for these measures (e.g. multi electrode arrays). The development of two-photon imaging provide good leads for population-wise measurement paradigms in IC.

Behavioral studies of ITD processing

The psychoacoustical study presented in the present thesis, despite being preliminary, suggests that the traditional view of an ITD that is the same for all signals coming from the same source is incomplete. Rather, humans seem to attribute the same position to

sounds with different ITDs. In other words, the percept of space, precisely because it is invariant on frequency, relies on a frequency-dependent representation of ITD.

As discussed in Chapter 1, the variation of ITD occurs at two scales, and our psychophysics data only addresses the *global* variation of ITD (i.e. across cochlear channels). On the local scale, filtering by the head and body introduces two separate cues: envelope and fine structure ITDs. Furthermore, at the neurophysiological level those cues seem to be indeed integrated (Chapter 6). Therefore, an interesting development is to try and assess the independent contribution (if any) of these cues to human localization abilities. Results from Chapter 2 indicate that for a given position, envelope and fine structure ITDs in anechoic conditions are equal over a range of frequencies (below 600 Hz). Therefore, a prediction could be that the localization performance degrades in this band when only one of the two ITD cues is presented in the stimulus.

In silico sound localization

Finally, the results presented in this study suggest that the problem of sound source localization in biological system cannot be solved without taking into consideration its *embodied* aspect. For biological systems, in this sense, it provides an advantage to have complex shapes in the sense that the binaural structure is then much more informative. This, as I show, is by virtue of the presence of a head, body and their interaction with the environment. The results I present in a second part indicate that the binaural system's heterogeneity makes it especially suited to perform sound localization.

Moreover, theoretical requirements for spike-based integration of those cues were studied: time-invariant transduction into spikes, axonal delays and differences in CF wiring at both ears, and finally coincidence detection. Neuromorphic sensors, that mimic the transduction of physical signals into neural signals, have become available and provide an exciting use-case for these theoretical developments [Liu and Delbruck, 2010]. Recently, groups have developed *in silico* cochleas, using either analog [Chan et al., 2007], or digital [Domnguez-Morales et al., 2011] hardware. The analog-based AER-EAR cochlea [Chan et al., 2007] provide a real-time implementation of two 64-channels cochleas. The parallel filtering of the cochlea, otherwise computationally expensive (in digital hardware), is here provided in real-time and for low-power consumption.

If this hardware is implemented in, e.g. a spherical shape, the developments from this thesis can be used to design the specific arrays of coincidence detectors that are necessary for spectrum-invariant sound localization. Yet, designing an *in silico* array of coincidence detectors, with sufficiently long delays is a challenge. The reason is

that imposing long delays on fast electrical signals is complicated without using digital hardware. Nonetheless, I believe that this is a fruitful endeavor: first by virtue of the improvements that such a technology introduces (e.g. in terms of energy consumption). Furthermore, using such a platform in a robot could provide a way of testing the peak and hemispheric codes in a real-world context.

Bibliography

[,]

[Abbagnaro et al., 1975] Abbagnaro, L. A., Bauer, B. B., and Torick, E. L. (1975). Measurements of diffraction and interaural delay of a progressive sound wave caused by the human head. II. *The Journal of the Acoustical Society of America*, 58:693.

[Abeles, 1991] Abeles, M. (1991). *Corticonics: Neural Circuits of the Cerebral Cortex*. Cambridge University Press. undefined Corticonics: Neural Circuits of the Cerebral Cortex 807 NOT IN FILE.

[Ackerman et al., 1957] Ackerman, E., Adam, A., Berger, R. L., Campanella, A. J., Danner, P. A., Farwell, R. W., Frings, H. W., Oda, F., and Tu, L. (1957). Sound absorption at the surfaces of small laboratory animals. technical report 57-461. Technical report, Wright Air Development Center.

[Agmon-Snir et al., 1998] Agmon-Snir, H., Carr, C. E., and Rinzel, J. (1998). The role of dendrites in auditory coincidence detection. *Nature*, 393(6682):268–272.

[Aitkin and Schuck, 1985] Aitkin, L. and Schuck, D. (1985). Low frequency neurons in the lateral central nucleus of the cat inferior colliculus receive their input predominantly from the medial superior olive. *Hearing research*, 17(1):87–93. PMID: 3997685.

[Aitkin et al., 1985] Aitkin, L. M., Pettigrew, J. D., Calford, M. B., Phillips, S. C., and Wise, L. Z. (1985). Representation of stimulus azimuth by low-frequency neurons in inferior colliculus of the cat. *Journal of neurophysiology*, 53(1):4359.

[Akeroyd, 2006] Akeroyd, M. A. (2006). The psychoacoustics of binaural hearing: La psicoacstica de la audicin binaural. *International journal of audiology*, 45(S1):2533.

[Algazi et al., 2001a] Algazi, V. R., Avendano, C., and Duda, R. O. (2001a). Elevation localization and head-related transfer function analysis at low frequencies. *The Journal of the Acoustical Society of America*, 109(3):1110–1122.

[Algazi et al., 2001b] Algazi, V. R., Avendano, C., and Duda, R. O. (2001b). Estimation of a spherical-head model from anthropometry. *Journal of the Audio Engineering Society*, 49(6):472479.

[Algazi et al., 2002] Algazi, V. R., Duda, R. O., Duraiswami, R., Gumerov, N. A., and Tang, Z. (2002). Approximating the head-related transfer function using simple geometric models of the head and torso. *The Journal of the Acoustical Society of America*, 112(5):2053–2064.

[Algazi et al., 2001c] Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (2001c). The cipic hrtf database. In *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, page 99102.

[Anderson, 1973] Anderson, D. I. (1973). Quantitative model for the effects of stimulus frequency upon synchronization of auditory nerve discharges. *The Journal of the Acoustical Society of America*, 54(2):361–364.

[Anderson, 1950] Anderson, V. C. (1950). Sound scattering from a fluid sphere. *The Journal of the Acoustical Society of America*, 22(4):426–431.

[ARI, 2010] ARI, D. (2010). http://www.kfs.oeaw.ac.at/content/view/608/606/lang,8859-1/.

[Ashmead et al., 1990] Ashmead, D. H., LeRoy, D., and Odom, R. D. (1990). Perception of the relative distances of nearby sound sources. *Perception & psychophysics*, 47(4):326–331. PMID: 2345684.

[Attwell and Laughlin, 2001] Attwell, D. and Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow & Metabolism*, 21(10):11331145.

[Batra et al., 1997] Batra, R., Kuwada, S., and Fitzpatrick, D. C. (1997). Sensitivity to interaural temporal disparities of low- and high-frequency neurons in the superior olivary complex. i. heterogeneity of responses. *Journal of Neurophysiology*, 78(3):1222–1236.

[Beckius et al., 1999] Beckius, G. E., Batra, R., and Oliver, D. L. (1999). Axons from anteroventral cochlear nucleus that terminate in medial superior olive of cat: observations related to delay lines. *The Journal of neuroscience*, 19(8):31463161.

[Bernardini and Rushmeier, 2002] Bernardini, F. and Rushmeier, H. (2002). The 3D model acquisition pipeline. *Computer Graphics Forum*, 21(2):149–172.

[Blauert, 1997] Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. The MIT press.

[Boehnke and Phillips, 1999] Boehnke, S. E. and Phillips, D. P. (1999). Azimuthal tuning of human perceptual channels for sound location. *The Journal of the Acoustical Society of America*, 106(4):1948–1955.

[Bonham and Lewis, 1999] Bonham, B. H. and Lewis, E. R. (1999). Localization by interaural time difference (ITD): effects of interaural frequency mismatch. *The Journal of the Acoustical Society of America*, 106(1):281–290.

[Brand et al., 2002] Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (2002). Precise inhibition is essential for microsecond interaural time difference coding. *Nature*, 417(6888):543547.

[Breebaart et al., 2010] Breebaart, J., Nater, F., and Kohlrausch, A. (2010). Spectral and spatial parameter resolution requirements for parametric, filter-bank-based HRTF processing. *Journal of the Audio Engineering Society*, 58(3):126140.

[Brette, 2010] Brette, R. (2010). On the interpretation of sensitivity analyses of neural responses. *The Journal of the Acoustical Society of America*, 128:2965.

[Brette, 2012] Brette, R. (2012). Computing with neural synchrony. *PLoS Computational Biology*, 8(6):e1002561.

[Brette and Gerstner, 2005] Brette, R. and Gerstner, W. (2005). Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of neurophysiology*, 94(5):36373642.

[Brown and Duda, 1998a] Brown, C. and Duda, R. (1998a). A structural model for binaural sound synthesis. *IEEE Transactions on Speech and Audio Processing*, 6(5):476–488.

[Brown and Duda, 1998b] Brown, C. P. and Duda, R. O. (1998b). A structural model for binaural sound synthesis. *Speech and Audio Processing, IEEE Transactions on*, 6(5):476488.

[Bruneau, 2010] Bruneau, M. (2010). *Fundamentals of acoustics*. Wiley-ISTE.

[Brungart and Rabinowitz, 1999] Brungart, D. S. and Rabinowitz, W. M. (1999). Auditory localization of nearby sources. head-related transfer functions. *The Journal of the Acoustical Society of America*, 106(3):1465–1479.

[Burger et al., 2011] Burger, R. M., Fukui, I., Ohmori, H., and Rubel, E. W. (2011). Inhibition in the balance: binaurally coupled inhibitory feedback in sound localization circuitry. *Journal of neurophysiology*, 106(1):414.

[Burkhard and Sachs, 1975] Burkhard, M. D. and Sachs, R. M. (1975). Anthropometric manikin for acoustic research. *Journal of the Acoustical Society of America*, 58(1):214–222.

[Butler, 1971] Butler, R. A. (1971). The monaural localization of tonal stimuli. *Perception & Psychophysics*, 9(1):99–101.

[Butler and Helwig, 1983] Butler, R. A. and Helwig, C. C. (1983). The spatial attributes of stimulus frequency in the median sagittal plane and their role in sound localization. *American Journal of Otolaryngology*, 4(3):165173.

[Buus, 1990] Buus, S. (1990). Level discrimination of frozen and random noise. *The Journal of the Acoustical Society of America*, 87(6):2643–2654.

[Campenhausen and Wagner, 2006] Campenhausen, M. v. and Wagner, H. (2006). Influence of the facial ruff on the sound-receiving characteristics of the barn owls ears. *Journal of Comparative Physiology A*, 192(10):1073–1082.

[Cant and Casseday, 1986] Cant, N. B. and Casseday, J. H. (1986). Projections from the anteroventral cochlear nucleus to the lateral and medial superior olivary nuclei. *Journal of Comparative Neurology*, 247(4):457476.

[Cant and Hyson, 1992] Cant, N. B. and Hyson, R. L. (1992). Projections from the lateral nucleus of the trapezoid body to the medial superior olivary nucleus in the gerbil. *Hearing research*, 58(1):2634.

[Carney, 1993] Carney, L. H. (1993). A model for the responses of low-frequency auditory-nerve fibers in cat. *The Journal of the Acoustical Society of America*, 93(1):401–417.

[Carney et al., 2011] Carney, L. H., Sarkar, S., Abrams, K. S., and Idrobo, F. (2011). Sound-localization ability of the mongolian gerbil (meriones unguiculatus) in a task with a simplified response map. *Hearing research*, 275(1-2):89–95. PMID: 21147208.

[Carney and Yin, 1988] Carney, L. H. and Yin, T. C. (1988). Temporal coding of resonances by low-frequency auditory nerve fibers: single-fiber responses and a population model. *Journal of neurophysiology*, 60(5):16531677.

[Carr and MacLeod, 2010] Carr, C. E. and MacLeod, K. M. (2010). Microseconds matter. *PLoS Biol*, 8(6):e1000405.

[Carr et al., 2009] Carr, C. E., Soares, D., Smolders, J., and Simon, J. Z. (2009). Detection of interaural time differences in the alligator. *The Journal of Neuroscience*, 29(25):7978–7990. PMID: 19553438.

[Casseday and Neff, 1973] Casseday, J. H. and Neff, W. D. (1973). Localization of pure tones. *The Journal of the Acoustical Society of America*, 54(2):365–372.

[Casseday and Neff, 1975] Casseday, J. H. and Neff, W. D. (1975). Auditory localization: role of auditory pathways in brain stem of the cat. *Journal of Neurophysiology*, 38(4):842–858. PMID: 1159469.

[Chan et al., 2007] Chan, V., Liu, S.-C., and van Schaik, A. (2007). AER EAR: a matched silicon cochlea pair with address event representation interface. *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 54(1):4859.

[Christensen-Dalsgaard et al., 2011] Christensen-Dalsgaard, J., Tang, Y., and Carr, C. E. (2011). Binaural processing by the gecko auditory periphery. *Journal of Neurophysiology*, 105(5):19922004.

[Colburn, 1973] Colburn, H. S. (1973). Theory of binaural interaction based on auditory-nerve data. i. general strategy and preliminary results on interaural discrimination. *The Journal of the Acoustical Society of America*, 54(6):1458–1470.

[Colburn, 1996] Colburn, H. S. (1996). Computational models of binaural processing. In *Auditory computation*, page 332400. Springer.

[Colburn and Durlach, 1978] Colburn, H. S. and Durlach, N. I. (1978). Models of binaural interaction. *Handbook of perception*, 4:467518.

[Constan and Hartmann, 2003] Constan, Z. A. and Hartmann, W. M. (2003). On the detection of dispersion in the head-related transfer function. *The Journal of the Acoustical Society of America*, 114:998.

[Couchman et al., 2010] Couchman, K., Grothe, B., and Felmy, F. (2010). Medial superior olivary neurons receive surprisingly few excitatory and inhibitory inputs with balanced strength and short-term dynamics. *The Journal of Neuroscience*, 30(50):1711117121.

[Culling and Summerfield, 1995] Culling, J. F. and Summerfield, Q. (1995). Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay. *The Journal of the Acoustical Society of America*, 98:785.

[Database, 2002] Database, L. (2002). http://recherche.ircam.fr/equipes/salles/listen/index.html.

[Day and Semple, 2011] 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(4):19851999.

[De Mey et al., 2008] De Mey, F., Reijniers, J., Peremans, H., Otani, M., and Firzlaff, U. (2008). Simulated head related transfer function of the phyllostomid bat Phyllostomus discolor. *Journal of the Acoustical Society of America*, 124(4, Part 1):2123–2132.

[Dellepiane et al., 2008] Dellepiane, M., Pietroni, N., Tsingos, N., Asselot, M., and Scopigno, R. (2008). Reconstructing head models from photographs for individualized 3D-audio processing. *Computer Graphics Forum*, 27(7):1719–1727.

[Dingle et al., 2013] Dingle, R. N., Hall, S. E., and Phillips, D. P. (2013). The three-channel model of sound localization mechanisms: Interaural time differences. *The Journal of the Acoustical Society of America*, 133(1):417–424.

[Domnguez-Morales et al., 2011] Domnguez-Morales, M., Jimenez-Fernandez, A., Cerezuela-Escudero, E., Paz-Vicente, R., Linares-Barranco, A., and Jimenez, G. (2011). On the designing of spikes band-pass filters for FPGA. In Honkela, T., Duch, W., Girolami, M., and Kaski, S., editors, *Artificial Neural Networks and Machine Learning ICANN 2011*, number 6792 in Lecture Notes in Computer Science, pages 389–396. Springer Berlin Heidelberg.

[Doyle and Snowden, 2001] Doyle, M. C. and Snowden, R. J. (2001). Identification of visual stimuli is improved by accompanying auditory stimuli: The role of eye movements and sound location. *Perception*, 30(7):795810.

[Dreyer and Delgutte, 2006] Dreyer, A. and Delgutte, B. (2006). Phase locking of auditory-nerve fibers to the envelopes of high-frequency sounds: Implications for sound localization. *Journal of Neurophysiology*, 96(5):2327–2341. PMID: 16807349.

[Duda et al., 1999a] Duda, R. O., Avendano, C., and Algazi, V. R. (1999a). An adaptable ellipsoidal head model for the interaural time difference. In *Acoustics, Speech,*

and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on, volume 2, page 965968.

[Duda et al., 1999b] Duda, R. O., Avendano, C., and Algazi, V. R. (1999b). An adaptable ellipsoidal head model for the interaural time difference. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 965–968.

[Duda and Martens, 1998] Duda, R. O. and Martens, W. L. (1998). Range dependence of the response of a spherical head model. *The Journal of the Acoustical Society of America*, 104:3048.

[Dunai and Hartmann, 2011] Dunai, L. and Hartmann, W. M. (2011). Frequency dependence of the interaural time difference thresholds in human listeners. *The Journal of the Acoustical Society of America*, 129:2485.

[Durkovic et al., 2011] 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(6):EL392EL398.

[Dynes and Delgutte, 1992] Dynes, S. B. and Delgutte, B. (1992). Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea. *Hearing Research*, 58(1):79–90.

[Eberle et al., 2000] Eberle, G., McAnally, K. I., Martin, R. L., and Flanagan, P. (2000). Localization of amplitude-modulated high-frequency noise. *The Journal of the Acoustical Society of America*, 107:3568.

[Evans, 1972] Evans, E. F. (1972). The frequency response and other properties of single fibres in the guinea-pig cochlear nerve. *The Journal of Physiology*, 226(1):263–287. PMID: 5083170.

[Farina, 2000] Farina, A. (2000). Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *Audio Engineering Society Convention 108*.

[Fels and Vorlander, 2009] Fels, J. and Vorlander, M. (2009). Anthropometric parameters influencing head-related transfer functions. *Acta Acustica united with Acustica*, 95(2):331342.

[Feuillade, 2004] Feuillade, C. (2004). Animations for visualizing and teaching acoustic impulse scattering from spheres. *The Journal of the Acoustical Society of America*, 115(5):1893–1904.

[Fischer et al., 2008] Fischer, B. J., Christianson, G. B., and Pena, J. L. (2008). Cross-correlation in the auditory coincidence detectors of owls. *J. Neurosci.*, 28(32):8107–8115.

[Fischer and Pea, 2011] Fischer, B. J. and Pea, J. L. (2011). Owl's behavior and neural representation predicted by bayesian inference. *Nature neuroscience*, 14(8):10611066.

[Fitzpatrick et al., 1997a] Fitzpatrick, D. C., Batra, R., Stanford, T. R., and Kuwada, S. (1997a). A neuronal population code for sound localization. *Nature*, 388(6645):871–874. PMID: 9278047.

[Fitzpatrick et al., 1997b] Fitzpatrick, D. C., Batra, R., Stanford, T. R., and Kuwada, S. (1997b). A neuronal population code for sound localization. *Nature*, 388(6645):871–874.

[Fitzpatrick et al., 2000] Fitzpatrick, D. C., Kuwada, S., and Batra, R. (2000). Neural sensitivity to interaural time differences: Beyond the jeffress model. *The Journal of Neuroscience*, 20(4):1605–1615. PMID: 10662850.

[Fontaine et al., 2013] Fontaine, B., Benichoux, V., Joris, P. X., and Brette, R. (2013). Predicting spike timing in highly synchronous auditory neurons at different sound levels. *Journal of neurophysiology*, 110(7):1672–1688. PMID: 23864375.

[Fontaine and Brette, 2011] Fontaine, B. and Brette, R. (2011). Neural development of binaural tuning through hebbian learning predicts frequency-dependent best delays. *The Journal of Neuroscience*, 31(32):1169211696.

[Fontaine et al., 2011] Fontaine, B., Goodman, D. F., Benichoux, V., and Brette, R. (2011). Brian hears: online auditory processing using vectorization over channels. *Frontiers in neuroinformatics*, 5.

[Francart and Wouters, 2007] Francart, T. and Wouters, J. (2007). Perception of across-frequency interaural level differences. *The Journal of the Acoustical Society of America*, 122:2826.

[Gibson, 1986] Gibson, J. J. (1986). *The Ecological approach to visual perception*. Lawrence Erlbaum Associates, Hillsdale (N.J.).

[Goldberg and Brown, 1969] Goldberg, J. M. and Brown, P. B. (1969). Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *Journal of neurophysiology*.

[Goodman and Brette, 2010] Goodman, D. F. and Brette, R. (2010). Spike-timing-based computation in sound localization. *PLoS computational biology*, 6(11):e1000993.

[Goodman and Brette, 2009] Goodman, D. F. M. and Brette, R. (2009). The brian simulator. *Frontiers in Neuroscience*, 3(2):192–197. PMID: 20011141 PMCID: PMC2751620.

[Gourevitch and Brette, 2012] Gourevitch, B. and Brette, R. (2012). The impact of early reflections on binaural cues. *The Journal of the Acoustical Society of America*, 132(1):9–27.

[Grace et al., 2008] Grace, S. M., Quaranta, E., Shinn-Cunningham, B. G., and Voigt, H. F. (2008). Simulation of the binaural environmental transfer function for gerbils using a boundary element method. *Acta Acustica United With Acustica*, 94(2):310–320.

[Grenander and Miller, 1994] Grenander, U. and Miller, M. I. (1994). Representations of knowledge in complex systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(4):pp. 549–603.

[Grothe et al., 2010] Grothe, B., Pecka, M., and McAlpine, D. (2010). Mechanisms of sound localization in mammals. *Physiological Reviews*, 90(3):9831012.

[Grothe and Sanes, 1994] Grothe, B. and Sanes, D. H. (1994). Synaptic inhibition influences the temporal coding properties of medial superior olivary neurons: an in vitro study. *The Journal of neuroscience*, 14(3):17011709.

[Guinan et al., 1972] Guinan, J. J., Norris, B. E., and Guinan, S. S. (1972). Single auditory units in the superior olivary complex: II: locations of unit categories and tonotopic organization. *International Journal of Neuroscience*, 4(4):147166.

[Hammershoi and Moller, 1996] Hammershoi, D. and Moller, H. (1996). Sound transmission to and within the human ear canal. *Journal of the Acoustical Society of America*, 100(1):408–427.

[Hammershi et al., 1992] Hammershi, D., Mller, H., and Srensen, M. F. (1992). Head-related transfer functions: measurements on 40 human subjects. In *Audio Engineering Society Convention 92*.

[Harper and McAlpine, 2004] Harper, N. S. and McAlpine, D. (2004). Optimal neural population coding of an auditory spatial cue. *Nature*, 430(7000):682686.

[Harris, 1960] Harris, G. G. (1960). Binaural interactions of impulsive stimuli and pure tones. *The Journal of the Acoustical Society of America*, 32(6):685–692.

[Hartley and Fry, 1921] Hartley, R. V. L. and Fry, T. C. (1921). The binaural location of pure tones. *Physical Review*, 18(6):431–442.

[Hartmann, 1983] Hartmann, W. M. (1983). Localization of sound in rooms. *The Journal of the Acoustical Society of America*, 74(5):1380–1391.

[Hartmann and Constan, 2002] Hartmann, W. M. and Constan, Z. A. (2002). Interaural level differences and the level-meter model. *The Journal of the Acoustical Society of America*, 112:1037.

[Hartmann et al., 2013] Hartmann, W. M., Rakerd, B., and Macaulay, E. J. (2013). On the ecological interpretation of limits of interaural time difference sensitivity. In *ICA 2013 Montreal*, volume 19, page 050151, Montreal, Canada. ASA.

[Hartmann and Wittenberg, 1996] Hartmann, W. M. and Wittenberg, A. (1996). On the externalization of sound images. *The Journal of the Acoustical Society of America*, 99:3678.

[Hawley et al., 2004] Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *The Journal of the Acoustical Society of America*, 115(2):833–843.

[Heffner and Heffner, 1988a] Heffner, R. S. and Heffner, H. E. (1988a). Sound localization acuity in the cat: effect of azimuth, signal duration, and test procedure. *Hearing research*, 36(2-3):221–232. PMID: 3209494.

[Heffner and Heffner, 1988b] Heffner, R. S. and Heffner, H. E. (1988b). Sound localization and use of binaural cues by the gerbil (meriones unguiculatus). *Behavioral neuroscience*, 102(3):422.

[Heffner and Heffner, 1992] Heffner, R. S. and Heffner, H. E. (1992). Visual factors in sound localization in mammals. *Journal of Comparative Neurology*, 317(3):219232.

[Henning, 1974] Henning, G. B. (1974). Lateralization and the binaural masking-level difference. *The Journal of the Acoustical Society of America*, 55(6):1259–1262.

[Huopaniemi et al., 1999] Huopaniemi, J., Zacharov, N., and Karjalainen, M. (1999). Objective and subjective evaluation of head-related transfer function filter design. *Journal of the Audio Engineering Society*, 47(4):218239.

[Hutchinson et al., 1970] Hutchinson, N. A., Koles, Z. J., and Smith, R. S. (1970). Conduction velocity in myelinated nerve fibres of xenopus laevis. *The Journal of Physiology*, 208(2):279–289. PMID: 5500723 PMCID: PMC1348748.

[Imig et al., 2000] Imig, T. J., Bibikov, N. G., Poirier, P., and Samson, F. K. (2000). Directionality derived from pinna-cue spectral notches in cat dorsal cochlear nucleus. *Journal of neurophysiology*, 83(2):907925.

[Irino and Patterson, 1997] Irino, T. and Patterson, R. D. (1997). A time-domain, level-dependent auditory filter: The gammachirp. *The Journal of the Acoustical Society of America*, 101(1):412–419.

[Izhikevich, 2006] Izhikevich, E. M. (2006). Polychronization: computation with spikes. *Neural computation*, 18(2):245282.

[Jacobson et al., 2001] Jacobson, G., Poganiatz, I., and Nelken, I. (2001). Synthesizing spatially complex sound in virtual space: an accurate offline algorithm. *Journal of Neuroscience Methods*, 106(1):29–38.

[Jeffress, 1948] Jeffress, L. A. (1948). A place theory of sound localization. *Journal of comparative and physiological psychology*, 41(1):35.

[Jenkins and Masterton, 1982] Jenkins, W. M. and Masterton, R. B. (1982). Sound localization: effects of unilateral lesions in central auditory system. *Journal of neurophysiology*, 47(6):987–1016. PMID: 7108581.

[Jercog et al., 2010] Jercog, P. E., Svirskis, G., Kotak, V. C., Sanes, D. H., and Rinzel, J. (2010). Asymmetric excitatory synaptic dynamics underlie interaural time difference processing in the auditory system. *PLoS biology*, 8(6):e1000406.

[Jo et al., 2008] Jo, H., Park, Y., and Park, Y.-s. (2008). Optimization of spherical and spheroidal head model for head related transfer function customization: Magnitude comparison. In *International Conference on Control, Automation and Systems, 2008. ICCAS 2008*, pages 251–254.

[Jones et al., 2011] Jones, H., Koka, K., Thornton, J., and Tollin, D. (2011). Concurrent development of the head and pinnae and the acoustical cues to sound location in a precocious species, the chinchilla (chinchilla lanigera). *Journal of the Association for Research in Otolaryngology*, 12:127–140.

[Joris and Yin, 2007] Joris, P. and Yin, T. C. (2007). A matter of time: internal delays in binaural processing. *Trends in neurosciences*, 30(2):7078.

[Joris, 2003] Joris, P. X. (2003). Interaural time sensitivity dominated by cochlea-induced envelope patterns. *The Journal of neuroscience*, 23(15):63456350.

[Joris et al., 2011] Joris, P. X., Bergevin, C., Kalluri, R., Laughlin, M. M., Michelet, P., Heijden, M. v. d., and Shera, C. A. (2011). Frequency selectivity in old-world monkeys corroborates sharp cochlear tuning in humans. *Proceedings of the National Academy of Sciences*, 108(42):17516–17520. PMID: 21987783.

[Joris et al., 1994] Joris, P. X., Carney, L. H., Smith, P. H., and Yin, T. C. (1994). Enhancement of neural synchronization in the anteroventral cochlear nucleus. i. responses to tones at the characteristic frequency. *Journal of neurophysiology*, 71(3):10221036.

[Joris et al., 2004] Joris, P. X., Schreiner, C. E., and Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological Reviews*, 84(2):541577.

[Joris et al., 2006] Joris, P. X., Van de Sande, B., Louage, D. H., and van der Heijden, M. (2006). Binaural and cochlear disparities. *Proceedings of the National Academy of Sciences of the United States of America*, 103(34):12917–12922. PMID: 16908859 PMCID: PMC1568946.

[Joris and Verschooten, 2013] Joris, P. X. and Verschooten, E. (2013). On the limit of neural phase locking to fine structure in humans. In *Basic Aspects of Hearing*, page 101108. Springer.

[Kahana, 2000] Kahana, Y. (2000). *Numerical modelling of the head-related transfer function*. PhD thesis, University of Southampton.

[Karino et al., 2011] Karino, S., Smith, P. H., Yin, T. C. T., and Joris, P. X. (2011). Axonal branching patterns as sources of delay in the mammalian auditory brainstem: A re-examination. *The Journal of Neuroscience*, 31(8):3016–3031. PMID: 21414923.

[Katz, 1998] Katz, B. F. G. (1998). *Measurement and calculation of individual head-related transfer function using a boundary element model incuding the measurement and effect of skin and hair impedance*. PhD thesis, Pennsylvania State University.

[Katz, 2000] Katz, B. F. G. (2000). Acoustic absorption measurement of human hair and skin within the audible frequency range. *Journal of the Acoustical Society of America*, 108(5):2238–2242.

[Katz, 2001] Katz, B. F. G. (2001). Boundary element method calculation of individual head-related transfer function. ii. impedance effects and comparisons to real measurements. *Journal of the Acoustical Society of America*, 110(5):2449–2455.

[Keller et al., 1998] Keller, C. H., Hartung, K., and Takahashi, T. T. (1998). Head-related transfer functions of the barn owl: measurement and neural responses. *Hearing research*, 118(1-2):13–34. PMID: 9606058.

[Kiang, 1965] Kiang, N. Y.-S. (1965). DISCHARGE PATTERNS OF SINGLE FIBERS IN THE CAT'S AUDITORY NERVE. Technical report.

[Kim et al., 2010a] Kim, D. O., Bishop, B., and Kuwada, S. (2010a). 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(4):541–557.

[Kim et al., 2010b] Kim, D. O., Bishop, B., and Kuwada, S. (2010b). 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(4):541–557.

[Kistler and Wightman, 1992a] Kistler, D. J. and Wightman, F. L. (1992a). A model of head-related transfer functions based on principal component analysis and minimum-phase reconstruction. *Journal of the Acoustical Society of America*, 91(3):1637–1647.

[Kistler and Wightman, 1992b] Kistler, D. J. and Wightman, F. L. (1992b). A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *The Journal of the Acoustical Society of America*, 91(3):1637–1647.

[Knudsen et al., 1979] Knudsen, E. I., Blasdel, G. G., and Konishi, M. (1979). Sound localization by the barn owl (tyto alba) measured with the search coil technique. *Journal of comparative physiology*, 133(1):1–11.

[Knudsen and Konishi, 1978] Knudsen, E. I. and Konishi, M. (1978). A neural map of auditory space in the owl. *Science*, 200(4343):795–797. PMID: 644324.

[Koka et al., 2011a] Koka, K., Jones, H. G., Thornton, J. L., Lupo, J. E., and Tollin, D. J. (2011a). Sound pressure transformations by the head and pinnae of the adult chinchilla (chinchilla lanigera). *Hearing Research*, 272(12):135–147.

[Koka et al., 2011b] Koka, K., Jones, H. G., Thornton, J. L., Lupo, J. E., and Tollin, D. J. (2011b). Sound pressure transformations by the head and pinnae of the adult Chinchilla (Chinchilla lanigera). *Hearing Research*, 272(1-2):135–147.

[Koka et al., 2008a] Koka, K., Read, H. L., and Tollin, D. J. (2008a). The acoustical cues to sound location in the rat: measurements of directional transfer functions. *The Journal of the Acoustical Society of America*, 123:4297.

[Koka et al., 2008b] Koka, K., Read, H. L., and Tollin, D. J. (2008b). The acoustical cues to sound location in the rat: Measurements of directional transfer functions. *Journal of the Acoustical Society of America*, 123(6):4297–4309.

[Konishi, 1971] Konishi, M. (1971). Sound localization in the barn owl. *The Journal of the Acoustical Society of America*, 50:148.

[Konishi, 2003] Konishi, M. (2003). Coding of auditory space. *Annual review of neuroscience*, 26(1):3155.

[Koppl, 1997] Koppl, C. (1997). Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, tyto alba. *The Journal of neuroscience*, 17(9):33123321.

[Koppl and Carr, 2008] Koppl, C. and Carr, C. E. (2008). Maps of interaural time difference in the chickens brainstem nucleus laminaris. *Biological cybernetics*, 98(6):541559.

[Kuba et al., 2003] Kuba, H., Yamada, R., and Ohmori, H. (2003). Evaluation of the limiting acuity of coincidence detection in nucleus laminaris of the chicken. *The Journal of physiology*, 552(2):611620.

[Kuhn, 1977] Kuhn, G. F. (1977). Model for the interaural time differences in the azimuthal plane. *The Journal of the Acoustical Society of America*, 62(1):157–167.

[Kuhn, 1983] Kuhn, G. F. (1983). Physical acoustics and measurements pertaining to directional hearing. *The Journal of the Acoustical Society of America*, 73(S1):S1–S1.

[Kulkarni et al., 1999] Kulkarni, A., Isabelle, S. K., and Colburn, H. S. (1999). Sensitivity of human subjects to head-related transfer-function phase spectra. *The Journal of the Acoustical Society of America*, 105:2821.

[Kuwada et al., 2011] Kuwada, S., Bishop, B., Alex, C., Condit, D. W., and Kim, D. O. (2011). Spatial tuning to sound-source azimuth in the inferior colliculus of unanesthetized rabbit. *Journal of Neurophysiology*, 106(5):2698–2708. PMID: 21849611.

[Kuwada et al., 1984] Kuwada, S., Yin, T. C., Syka, J., Buunen, T. J., and Wickesberg, R. E. (1984). Binaural interaction in low-frequency neurons in inferior colliculus of the cat. IV. comparison of monaural and binaural response properties. *Journal of neurophysiology*, 51(6):13061325.

[Kppl and Carr, 2008] Kppl, C. and Carr, C. E. (2008). Maps of interaural time difference in the chickens brainstem nucleus laminaris. *Biological cybernetics*, 98(6):541559.

[Lafarge et al., 2010] Lafarge, F., Keriven, R., Brdif, M., and Vu, H.-H. (2010). Hybrid multi-view reconstruction by jump-diffusion. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, San Fransisco.

[Larcher and Jot, 1997] Larcher, V. and Jot, J.-M. (1997). Techniques dinterpolation de filtres audio-numriques, applicationa la reproduction spatiale des sons sur couteurs. In *Proc. CFA: Congres Fran\ccais dAcoustique*.

[Leberl and Thurgood, 2004] Leberl, F. and Thurgood, J. (2004). The promise of softcopy photogrammetry revisited. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, 35:759–763.

[Leibold, 2010] Leibold, C. (2010). Influence of inhibitory synaptic kinetics on the interaural time difference sensitivity in a linear model of binaural coincidence detection. *The Journal of the Acoustical Society of America*, 127(2):931–942.

[Lesica et al., 2010] Lesica, N. A., Lingner, A., and Grothe, B. (2010). Population coding of interaural time differences in gerbils and barn owls. *The Journal of Neuroscience*, 30(35):1169611702.

[Licklider, 1951] Licklider, J. C. R. (1951). A duplex theory of pitch perception. *Experientia*, 7(4):128–134.

[Licklider et al., 1950] Licklider, J. C. R., Webster, J. C., and Hedlun, J. M. (1950). On the frequency limits of binaural beats. *The Journal of the Acoustical Society of America*, 22:468.

[Lindemann, 1986] Lindemann, W. (1986). Extension of a binaural cross-correlation model by contralateral inhibition. i. simulation of lateralization for stationary signals. *The Journal of the Acoustical Society of America*, 80:1608.

[Litovsky et al., 1999] Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (1999). The precedence effect. *The Journal of the Acoustical Society of America*, 106:1633.

[Liu and Delbruck, 2010] Liu, S.-C. and Delbruck, T. (2010). Neuromorphic sensory systems. *Current opinion in neurobiology*, 20(3):288295.

[Lopez-Poveda, 2005] Lopez-Poveda, E. A. (2005). Spectral processing by the peripheral auditory system: facts and models. *International review of neurobiology*, 70:748.

[Lopez-Poveda and Meddis, 2001] Lopez-Poveda, E. A. and Meddis, R. (2001). A human nonlinear cochlear filterbank. *The Journal of the Acoustical Society of America*, 110:3107.

[Lorenzi et al., 2006] Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences*, 103(49):1886618869.

[Louage et al., 2005] Louage, D. H., van der Heijden, M., and Joris, P. X. (2005). Enhanced temporal response properties of anteroventral cochlear nucleus neurons to broadband noise. *The Journal of neuroscience*, 25(6):15601570.

[Lling et al., 2011] Lling, H., Siveke, I., Grothe, B., and Leibold, C. (2011). Frequency-invariant representation of interaural time differences in mammals. *PLoS Comput Biol*, 7(3):e1002013.

[MacDonald, 2008] MacDonald, J. A. (2008). A localization algorithm based on head-related transfer functions. *The Journal of the Acoustical Society of America*, 123:4290.

[MacLeod et al., 2006] MacLeod, K. M., Soares, D., and Carr, C. E. (2006). Interaural timing difference circuits in the auditory brainstem of the emu (dromaius novaehollandiae). *The Journal of comparative neurology*, 495(2):185–201. PMID: 16435285.

[Macpherson and Middlebrooks, 2002] Macpherson, E. A. and Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. *The Journal of the Acoustical Society of America*, 111:2219.

[Maki and Furukawa, 2005a] Maki, K. and Furukawa, S. (2005a). Acoustical cues for sound localization by the mongolian gerbil, meriones unguiculatus. *The Journal of the Acoustical Society of America*, 118:872.

[Maki and Furukawa, 2005b] Maki, K. and Furukawa, S. (2005b). Acoustical cues for sound localization by the Mongolian gerbil, Meriones unguiculatus. *Journal of the Acoustical Society of America*, 118(2):872–886.

[Maki and Furukawa, 2005c] Maki, K. and Furukawa, S. (2005c). Reducing individual differences in the external-ear transfer functions of the mongolian gerbil. *Journal of the Acoustical Society of America*, 118(4):2392–2404.

[Mardia, 1975] Mardia, K. V. (1975). Statistics of directional data. *Journal of the Royal Statistical Society. Series B (Methodological)*, page 349393.

[Marple Jr, 1999] Marple Jr, S. L. (1999). Estimating group delay and phase delay via discrete-time analytic cross-correlation. *Signal Processing, IEEE Transactions on*, 47(9):26042607.

[Masterton et al., 1967] Masterton, B., Jane, J. A., and Diamond, I. T. (1967). Role of brainstem auditory structures in sound localization. i. trapezoid body, superior olive, and lateral lemniscus. *Journal of neurophysiology*, 30(2):341–359. PMID: 4167210.

[Mathews et al., 2010] Mathews, P. J., Jercog, P. E., Rinzel, J., Scott, L. L., and Golding, N. L. (2010). Control of submillisecond synaptic timing in binaural coincidence detectors by kv1 channels. *Nature Neuroscience*, 13(5):601–609.

[Mayer, 2008] Mayer, H. (2008). Object extraction in photogrammetric computer vision. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(2):213–222.

[Mc Laughlin et al., 2007] Mc Laughlin, M., Van de Sande, B., van der Heijden, M., and Joris, P. X. (2007). Comparison of bandwidths in the inferior colliculus and the auditory nerve. i. measurement using a spectrally manipulated stimulus. *Journal of neurophysiology*, 98(5):25662579.

[McAlpine et al., 1996] McAlpine, D., Jiang, D., and Palmer, A. R. (1996). Interaural delay sensitivity and the classification of low best-frequency binaural responses in the inferior colliculus of the guinea pig. *Hearing research*, 97(1-2):136–152. PMID: 8844194.

[McAlpine et al., 2001] McAlpine, D., Jiang, D., and Palmer, A. R. (2001). A neural code for low-frequency sound localization in mammals. *Nature neuroscience*, 4(4):396–401. PMID: 11276230.

[McFadden and Pasanen, 1976] McFadden, D. and Pasanen, E. G. (1976). Lateralization at high frequencies based on interaural time differences. *The Journal of the Acoustical Society of America*, 59:634.

[Mehrgardt and Mellert, 1977] Mehrgardt, S. and Mellert, V. (1977). Transformation characteristics of external human ear. *Journal of the Acoustical Society of America*, 61(6):1567–1576.

[Michelet et al., 2012] Michelet, P., Kovaci, D., and Joris, P. X. (2012). Ongoing temporal coding of a stochastic stimulus as a function of intensity: time-intensity trading. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 32(28):9517–9527. PMID: 22787037.

[Mills, 1958] Mills, A. W. (1958). On the minimum audible angle. *The Journal of the Acoustical Society of America*, 30:237.

[Minnaar et al., 1999] Minnaar, P., Christensen, F., Moller, H., Olesen, S. K., and Plogsties, J. (1999). Audibility of all-pass components in binaural synthesis. In *Audio Engineering Society Convention 106*.

[Morse and Ingard, 1968] Morse, P. M. and Ingard, K. (1968). *Uno: Theoretical Acoustics*. McGraw-Hill Book Co., Inc., c.

[Mossop and Culling, 1998] Mossop, J. E. and Culling, J. F. (1998). Lateralization of large interaural delays. *The Journal of the Acoustical Society of America*, 104:1574.

[Mrsic-Flogel et al., 2001] Mrsic-Flogel, T. D., King, A. J., Jenison, R. L., and Schnupp, J. W. H. (2001). Listening through different ears alters spatial response fields in ferret primary auditory cortex. *Journal of Neurophysiology*, 86(2):1043–1046.

[Mrsic-Flogel et al., 2005] Mrsic-Flogel, T. D., King, A. J., and Schnupp, J. W. H. (2005). Encoding of virtual acoustic space stimuli by neurons in ferret primary auditory cortex. *Journal of Neurophysiology*, 93(6):3489–3503. PMID: 15659534.

[Muller, 2004] Muller, R. (2004). A numerical study of the role of the tragus in the big brown bat. *Journal of the Acoustical Society of America*, 116(6):3701–3712.

[Musicant and Butler, 1985] Musicant, A. D. and Butler, R. A. (1985). Influence of monaural spectral cues on binaural localization. *The Journal of the Acoustical Society of America*, 77(1):202–208.

[Narayan et al., 1998] Narayan, S. S., Temchin, A. N., Recio, A., and Ruggero, M. A. (1998). Frequency tuning of basilar membrane and auditory nerve fibers in the same cochleae. *Science (New York, N.Y.)*, 282(5395):1882–1884. PMID: 9836636 PMCID: PMC3578392.

[Nicol, 2010] Nicol, R. (2010). *Reprsentation et perception des espaces auditifs virtuels (representation and perception of virtual auditory spaces)*. PhD thesis, Habilitation thesis, Univ. du Maine, F-Le Mans.

[Nielsen, 1992] Nielsen, S. H. (1992). Auditory distance perception in different rooms. Audio Engineering Society.

[Oertel and Young, 2004] Oertel, D. and Young, E. D. (2004). What's a cerebellar circuit doing in the auditory system? *Trends in Neurosciences*, 27(2):104–110.

[Oldfield and Parker, 1986] Oldfield, S. R. and Parker, S. P. A. (1986). Acuity of sound localisation: a topography of auditory space. III. monaural hearing conditions. *Perception*, 15(1):6781.

[Oliver, 1987] Oliver, D. L. (1987). Projections to the inferior colliculus from the anteroventral cochlear nucleus in the cat: Possible substrates for binaural interaction. *The Journal of Comparative Neurology*, 264(1):2446.

[Oliver et al., 2003] Oliver, D. L., Beckius, G. E., Bishop, D. C., Loftus, W. C., and Batra, R. (2003). Topography of interaural temporal disparity coding in projections of medial superior olive to inferior colliculus. *The Journal of neuroscience*, 23(19):74387449.

[Oppenheim et al., 1999] Oppenheim, A. V., Schafer, R. W., and Buck, J. R. (1999). *Discrete-time signal processing*, volume 5. Prentice hall Upper Saddle River.

[Otani and Ise, 2006] Otani, M. and Ise, S. (2006). Fast calculation system specialized for head-related transfer function based on boundary element method. *Journal of the Acoustical Society of America*, 119(5, Part 1):2589–2598.

[Oxenham and Shera, 2003] Oxenham, A. J. and Shera, C. A. (2003). Estimates of human cochlear tuning at low levels using forward and simultaneous masking. *Journal of the Association for Research in Otolaryngology*, 4(4):541554.

[Palmer and Russell, 1986] Palmer, A. R. and Russell, I. J. (1986). Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. *Hearing research*, 24(1):115.

[Paolini et al., 2001] Paolini, A. G., FitzGerald, J. V., Burkitt, A. N., and Clark, G. M. (2001). Temporal processing from the auditory nerve to the medial nucleus of the trapezoid body in the rat. *Hearing Research*, 159(12):101–116.

[Patterson, 1971] Patterson, R. D. (1971). Effect of amplitude on auditory filter shape. *The Journal of the Acoustical Society of America*, 49(1A):81–81.

[Pecka et al., 2008] Pecka, M., Brand, A., Behrend, O., and Grothe, B. (2008). Interaural time difference processing in the mammalian medial superior olive. *The journal of neuroscience*, (27):69146925.

[Pena et al., 2001] Pena, J. L., Viete, S., Funabiki, K., Saberi, K., and Konishi, M. (2001). Cochlear and neural delays for coincidence detection in owls. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 21(23):9455–9459. PMID: 11717379.

[Pernaux, 2003] Pernaux, J. M. (2003). *Spatialisation du son par les techniques binaurales: application aux services de tlcommunications.* PhD thesis, Institut National Polytechnique de Grenoble.

[Perrott and Nelson, 1969] Perrott, D. R. and Nelson, M. A. (1969). Limits for the detection of binaural beats. *The Journal of the Acoustical Society of America*, 46(6B):1477–1481.

[Perrott and Saberi, 1990] Perrott, D. R. and Saberi, K. (1990). Minimum audible angle thresholds for sources varying in both elevation and azimuth. *The Journal of the Acoustical Society of America*, 87(4):1728–1731.

[Perrott et al., 1990] Perrott, D. R., Saberi, K., Brown, K., and Strybel, T. Z. (1990). Auditory psychomotor coordination and visual search performance. *Perception & Psychophysics*, 48(3):214226.

[Phillips and Brugge, 1985] Phillips, D. P. and Brugge, J. F. (1985). Progress in neurophysiology of sound localization. *Annual review of psychology*, 36(1):245274.

[Pickles, 1988] Pickles, J. O. (1988). *An introduction to the physiology of hearing*, volume 2. Academic press London.

[Plogsties et al., 2000] Plogsties, J., Minnaar, P., Olesen, S. K., Christensen, F., and Mller, H. (2000). Audibility of all-pass components in head-related transfer functions. In *Audio Engineering Society Convention 108*.

[Quaranta, 2003] Quaranta, E. (2003). Application of the boundary element method to computation of the head-related transfer function for gerbils. Master's thesis, Boston University, College of Engineering.

[Rathouz and Trussell, 1998] Rathouz, M. and Trussell, L. (1998). Characterization of outward currents in neurons of the avian nucleus magnocellularis. *Journal of Neurophysiology*, 80(6):28242835.

[Rayleigh, 1907] Rayleigh, L. (1907). XII. on our perception of sound direction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 13(74):214232.

[Rayleigh and Lodge, 1904] Rayleigh, L. and Lodge, A. (1904). On the acoustic shadow of a sphere. with an appendix, giving the values of legendre's functions from p $ _ ${$0$}$ $ to p $ _ ${$20$}$ $ at intervals of 5 degrees. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 203(359-371):87110.

[Rb́illat et al., 2011] Rb́illat, M., Hennequin, R., Corteel, E., and Katz, B. F. (2011). Identification of cascade of hammerstein models for the description of nonlinearities in vibrating devices. *Journal of Sound and Vibration*, 330(5):1018–1038.

[Rose et al., 1967] Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (1967). Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *Journal of neurophysiology*, 30(4):769793.

[Rose et al., 1966] Rose, J. E., Gross, N. B., Geisler, C. D., and Hind, J. E. (1966). Some neural mechanisms in the inferior colliculus of the cat which may be relevant to localization of a sound source. *Journal of neurophysiology*, 29(2):288–314. PMID: 5927463.

[Rossant and Brette, 2010] Rossant, C. and Brette, R. (2010). Coincidence detection in neurons with in vivo-like synaptic activity. In *Cinquime confrence plnire fran\ccaise de Neurosciences Computationnelles,'Neurocomp'10'*.

[Rossant et al., 2011] Rossant, C., Leijon, S., Magnusson, A. K., and Brette, R. (2011). Sensitivity of noisy neurons to coincident inputs. *The Journal of Neuroscience*, 31(47):1719317206.

[Roth et al., 1980a] Roth, G., Kochhaar, R., and Hind, J. (1980a). Inter-aural time differences - implications regarding the neurophysiology of sound localization. *Journal of the Acoustical Society of America*, 68(6):1643–1651.

[Roth et al., 1980b] Roth, G. L., Kochhar, R. K., and Hind, J. E. (1980b). Interaural time differences: Implications regarding the neurophysiology of sound localization. *The Journal of the Acoustical Society of America*, 68(6):1643–1651.

[Ruggero and Rich, 1983] Ruggero, M. A. and Rich, N. C. (1983). Chinchilla auditory-nerve responses to low-frequency tones. *The Journal of the Acoustical Society of America*, 73:2096.

[Ruggero and Temchin, 2005] Ruggero, M. A. and Temchin, A. N. (2005). Unexceptional sharpness of frequency tuning in the human cochlea. *Proceedings of the National Academy of Sciences of the United States of America*, 102(51):18614–18619. PMID: 16344475.

[Sachs and Abbas, 1974] Sachs, M. B. and Abbas, P. J. (1974). Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli. *The Journal of the Acoustical Society of America*, 56:1835.

[Sayers and Cherry, 1957] Sayers, B. M. and Cherry, E. C. (1957). Mechanism of binaural fusion in the hearing of speech. *The Journal of the Acoustical Society of America*, 29(9):973–987.

[Schneggenburger and Forsythe, 2006] Schneggenburger, R. and Forsythe, I. D. (2006). The calyx of held. *Cell and Tissue Research*, 326(2):311–337.

[Schnupp et al., 2001] Schnupp, J., Mrsic-Flogel, T., and King, A. (2001). Linear processing of spatial cues in primary auditory cortex. *Nature*, 414(6860):200–204.

[Schnupp et al., 2003] Schnupp, J. W. H., Booth, J., and King, A. J. (2003). Modeling individual differences in ferret external ear transfer functions. *Journal of the Acoustical Society of America*, 113(4):2021–2030.

[Schnupp and Carr, 2009] Schnupp, J. W. H. and Carr, C. E. (2009). On hearing with more than one ear: lessons from evolution. *Nature Neuroscience*, 12(6):692–697.

[Schroeder, 1977] Schroeder, M. R. (1977). New viewpoints in binaural interactions. *Psychophysics and physiology of hearing*, page 455467.

[Scott et al., 2005] Scott, L. L., Mathews, P. J., and Golding, N. L. (2005). Posthearing developmental refinement of temporal processing in principal neurons of the medial superior olive. *The Journal of neuroscience*, 25(35):78877895.

[Seidl, 2013] Seidl, A. H. (2013). Regulation of conduction time along axons. *Neuroscience*.

[Seidl et al., 2010] Seidl, A. H., Rubel, E. W., and Harris, D. M. (2010). Mechanisms for adjusting interaural time differences to achieve binaural coincidence detection. *The Journal of Neuroscience*, 30(1):70–80. PMID: 20053889.

[Seitz et al., 2006] Seitz, S., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms.

In *International Conference on Computer Vision and Pattern Recognition*, volume 1, pages 519 – 528.

[Shamma et al., 1989] Shamma, S. A., Shen, N., and Gopalaswamy, P. (1989). Stereausis: binaural processing without neural delays. *The Journal of the Acoustical Society of America*, 86:989.

[Shera et al., 2010] Shera, C. A., Jr, J. J. G., and Oxenham, A. J. (2010). Otoacoustic estimation of cochlear tuning: Validation in the chinchilla. *Journal of the Association for Research in Otolaryngology*, 11(3):343–365.

[Smith et al., 1993] Smith, P. H., Joris, P. X., and Yin, T. C. (1993). Projections of physiologically characterized spherical bushy cell axons from the cochlear nucleus of the cat: evidence for delay lines to the medial superior olive. *Journal of Comparative Neurology*, 331(2):245260.

[Spangler et al., 1985] Spangler, K. M., Warr, W. B., and Henkel, C. K. (1985). The projections of principal cells of the medial nucleus of the trapezoid body in the cat. *Journal of Comparative Neurology*, 238(3):249262.

[Spencer et al., 2012] Spencer, M. J., Grayden, D. B., Bruce, I. C., Meffin, H., and Burkitt, A. N. (2012). An investigation of dendritic delay in octopus cells of the mammalian cochlear nucleus. *Frontiers in Computational Neuroscience*, 6. PMID: 23125831 PMCID: PMC3486622.

[Spezio et al., 2000a] Spezio, M., Keller, C., Marrocco, R., and Takahashi, T. (2000a). Head-related transfer functions of the Rhesus monkey. *Hearing Research*, 144(1-2):73–88.

[Spezio et al., 2000b] Spezio, M. L., Keller, C. H., Marrocco, R. T., and Takahashi, T. T. (2000b). Head-related transfer functions of the rhesus monkey. *Hearing Research*, 144(12):73–88.

[Stecker et al., 2005] Stecker, G. C., Harrington, I. A., and Middlebrooks, J. C. (2005). Location coding by opponent neural populations in the auditory cortex. *PLoS biology*, 3(3):e78.

[Sterbing et al., 2003a] Sterbing, S. J., Hartung, K., and Hoffmann, K.-P. (2003a). Spatial tuning to virtual sounds in the inferior colliculus of the guinea pig. *Journal of neurophysiology*, 90(4):26482659.

[Sterbing et al., 2003b] Sterbing, S. J., Hartung, K., and Hoffmann, K.-P. (2003b). Spatial tuning to virtual sounds in the inferior colliculus of the guinea pig. *Journal of Neurophysiology*, 90(4):2648–2659.

[Stern and Trahiotis, 1992] Stern, R. M. and Trahiotis, C. (1992). The role of consistency of interaural timing over frequency in binaural lateralization. *Auditory physiology and perception*, page 547554.

[Stern and Trahiotis, 1995] Stern, R. M. and Trahiotis, C. (1995). Models of binaural interaction. *Handbook of perception and cognition*, 6:347387.

[Stern et al., 1988] Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (1988). Lateralization of complex binaural stimuli: A weighted-image model. *The Journal of the Acoustical Society of America*, 84:156.

[Stern Jr and Colburn, 1978] Stern Jr, R. M. and Colburn, H. S. (1978). Theory of binaural interaction based on auditory-nerve data. IV. a model for subjective lateral position. *The Journal of the Acoustical Society of America*, 64:127.

[Stewart, 1914] Stewart, G. W. (1914). Phase relations in the acoustic shadow of a rigid sphere; phase difference at the ears. *Physical Review*, 4(3):252.

[Strecha et al., 2008] Strecha, C., von Hansen, W., Van Gool, L., Fua, P., and Thoennessen, U. (2008). On benchmarking camera calibration and multi-view stereo for high resolution imagery. *IEEE Conference on Computer Vision and Pattern Recognition, Vols 1-12*, pages 2838–2845.

[Sullivan and Konishi, 1984] Sullivan, W. E. and Konishi, M. (1984). Segregation of stimulus phase and intensity coding in the cochlear nucleus of the barn owl. *The Journal of neuroscience*, 4(7):17871799.

[Sullivan and Konishi, 1986] Sullivan, W. E. and Konishi, M. (1986). Neural map of interaural phase difference in the owl's brainstem. *Proceedings of the National Academy of Sciences*, 83(21):8400–8404. PMID: 3022292.

[Sumner et al., 2003] Sumner, C. J., OMard, L. P., Lopez-Poveda, E. A., and Meddis, R. (2003). A nonlinear filter-bank model of the guinea-pig cochlear nerve: rate responses. *The Journal of the Acoustical Society of America*, 113:3264.

[Svirskis et al., 2004] Svirskis, G., Kotak, V., Sanes, D. H., and Rinzel, J. (2004). Sodium along with low-threshold potassium currents enhance coincidence detection of subthreshold noisy signals in MSO neurons. *Journal of neurophysiology*, 91(6):24652473.

[Tan and Carney, 2003] Tan, Q. and Carney, L. H. (2003). A phenomenological model for the responses of auditory-nerve fibers. II. nonlinear tuning with a frequency glide. *The Journal of the Acoustical Society of America*, 114:2007.

[Thompson et al., 2006] Thompson, S. K., von Kriegstein, K., Deane-Pratt, A., Marquardt, T., Deichmann, R., Griffiths, T. D., and McAlpine, D. (2006). Representation of interaural time delay in the human auditory midbrain. *Nature neuroscience*, 9(9):1096–1098. PMID: 16921369.

[Toledo and Mller, 2008] Toledo, D. and Mller, H. (2008). Audibility of high q-factor all-pass components in head-related transfer functions. In *Audio Engineering Society Convention 125*.

[Tollin, 2003] Tollin, D. J. (2003). The lateral superior olive: a functional role in sound source localization. *The neuroscientist*, 9(2):127143.

[Tollin and Koka, 2009a] Tollin, D. J. and Koka, K. (2009a). Postnatal development of sound pressure transformations by the head and pinnae of the cat: Binaural characteristics. *The Journal of the Acoustical Society of America*, 126(6):3125–3136.

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

[Tollin and Koka, 2009c] Tollin, D. J. and Koka, K. (2009c). Postnatal development of sound pressure transformations by the head and pinnae of the cat: Monaural characteristics. *The Journal of the Acoustical Society of America*, 125(2):980–994.

[Tollin and Koka, 2009d] Tollin, D. J. and Koka, K. (2009d). Postnatal development of sound pressure transformations by the head and pinnae of the cat: Monaural characteristics. *The Journal of the Acoustical Society of America*, 125(2):980–994.

[Tollin and Yin, 2005] Tollin, D. J. and Yin, T. C. (2005). Interaural phase and level difference sensitivity in low-frequency neurons in the lateral superior olive. *The Journal of neuroscience*, 25(46):1064810657.

[Trahiotis and Stern, 1994] Trahiotis, C. and Stern, R. M. (1994). Across-frequency interaction in lateralization of complex binaural stimuli. *The Journal of the Acoustical Society of America*, 96:3804.

[Treeby et al., 2007] Treeby, B. E., Paurobally, R. M., and Pan, J. (2007). The effect of impedance on interaural azimuth cues derived from a spherical head model. *The Journal of the Acoustical Society of America*, 121:2217.

[vanderHeijden et al., 2013] vanderHeijden, M., Lorteije, J., Plauka, A., Roberts, M., Golding, N., and Borst, J. (2013). Directional hearing by linear summation of binaural inputs at the medial superior olive. *Neuron*, 78(5):936–948.

[Wakeford and Robinson, 1974] Wakeford, O. and Robinson, D. (1974). Lateralization of tonal stimuli by cat. *Journal of the Acoustical Society of America*, 55(3):649–652.

[Wallach, 1940] Wallach, H. (1940). The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology*, 27(4):339.

[Wanrooij and Opstal, 2004] Wanrooij, M. M. V. and Opstal, A. J. V. (2004). Contribution of head shadow and pinna cues to chronic monaural sound localization. *The Journal of Neuroscience*, 24(17):4163–4171. PMID: 15115811.

[Warusfel, 2002] Warusfel, O. (2002). The LISTEN database: http://recherche.ircam.fr/equipes/salles/listen/ (date last viewed 07/22/2013).

[Wever, 1949] Wever, E. G. (1949). Theory of hearing.

[Wiener, 1947] Wiener, F. M. (1947). Sound diffraction by rigid spheres and circular cylinders. *The Journal of the Acoustical Society of America*, 19(3):444–451.

[Wightman and Kistler, 2005a] Wightman, F. and Kistler, D. (2005a). Measurement and validation of human HRTFs for use in hearing research. *Acta Acustica united with Acustica*, 91(3):429439.

[Wightman and Kistler, 2005b] Wightman, F. and Kistler, D. (2005b). Measurement and validation of human HRTFs for use in hearing research. *Acta Acustica united with Acustica*, 91(3):429–439.

[Wightman and Kistler, 1989] Wightman, F. L. and Kistler, D. J. (1989). Headphone simulation of free-field listening. i: Stimulus synthesis. *The Journal of the Acoustical Society of America*, 85:858.

[Wightman and Kistler, 1992] Wightman, F. L. and Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, 91:1648.

[Wightman and Kistler, 1997] Wightman, F. L. and Kistler, D. J. (1997). Monaural sound localization revisited. *The Journal of the Acoustical Society of America*, 101(2):1050–1063.

[Wightman and Kistler, 1999] Wightman, F. L. and Kistler, D. J. (1999). Resolution of frontback ambiguity in spatial hearing by listener and source movement. *The Journal of the Acoustical Society of America*, 105:2841.

[Wu and Oertel, 1984] Wu, S. H. and Oertel, D. (1984). Intracellular injection with horseradish peroxidase of physiologically characterized stellate and bushy cells in slices of mouse anteroventral cochlear nucleus. *The Journal of Neuroscience*, 4(6):1577–1588. PMID: 6726347.

[Xie and Zhang, 2010] Xie, B. and Zhang, T. (2010). The audibility of spectral detail of head-related transfer functions at high frequency. *Acta Acustica united with Acustica*, 96(2):328339.

[Xu and Middlebrooks, 2000] Xu, L. and Middlebrooks, J. C. (2000). Individual differences in external-ear transfer functions of cats. *The Journal of the Acoustical Society of America*, 107(3):1451–1459.

[Yin and Chan, 1990] Yin, T. C. and Chan, J. C. (1990). Interaural time sensitivity in medial superior olive of cat. *Journal of Neurophysiology*, 64(2):465488.

[Yin and Kuwada, 1983] Yin, T. C. and Kuwada, S. (1983). Binaural interaction in low-frequency neurons in inferior colliculus of the cat. III. effects of changing frequency. *Journal of Neurophysiology*, 50(4):1020–1042.

[Yost, 1974] Yost, W. A. (1974). Discriminations of interaural phase differences. *The Journal of the Acoustical Society of America*, 55(6):1299–1303.

[Yost and Raymond H. Dye, 1988] Yost, W. A. and Raymond H. Dye, J. (1988). Discrimination of interaural differences of level as a function of frequency. *The Journal of the Acoustical Society of America*, 83(5):1846–1851.

[Yu et al., 2009] Yu, T., Schwartz, A., Harris, J., Slaney, M., and Liu, S.-C. (2009). Periodicity detection and localization using spike timing from the AER EAR. In *Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on*, page 109112.

[Zhang et al., 2001] Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (2001). A phenomenological model for the responses of auditory-nerve fibers: I. nonlinear

tuning with compression and suppression. *The Journal of the Acoustical Society of America*, 109(2):648–670.

[Zhou et al., 2005] Zhou, Y., Carney, L. H., and Colburn, H. S. (2005). A model for interaural time difference sensitivity in the medial superior olive: Interaction of excitatory and inhibitory synaptic inputs, channel dynamics, and cellular morphology. *The Journal of Neuroscience*, 25(12):3046–3058. PMID: 15788761.

[Zotkin et al., 2003] Zotkin, D. N., Hwang, J., Duraiswami, R., and Davis, L. S. (2003). HRTF personalization using anthropometric measurements. In *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 157–160.

[Zou et al., 2012] Zou, H., Mller, H. J., and Shi, Z. (2012). Non-spatial sounds regulate eye movements and enhance visual search. *Journal of Vision*, 12(5):2. PMID: 22562709.

[Zwiers et al., 2001] Zwiers, M. P., Van Opstal, A. J., and Cruysberg, J. R. (2001). A spatial hearing deficit in early-blind humans. *J Neurosci*, 21(9):141145.

[Zwislocki and Feldman, 1956] Zwislocki, J. and Feldman, R. S. (1956). Just noticeable differences in dichotic phase. *The Journal of the Acoustical Society of America*, 28:860.