Fakultät für Elektrotechnik und Informationstechnik
Professur für Bioanaloge Informationsverarbeitung

A Model of the Neuronal Representation of Interaural Phase Differences

Jörg Encke

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender:
Prof. Dr. Thomas Hamacher

Prüfende der Dissertation:
1. Prof. Dr. Werner Hemmert
2. Assoc. Prof. Ville Pulkki Ph.D.,
   Aalto University

Die Dissertation wurde am 25. September 2018 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 17. Dezember 2018 angenommen.

Abstract

Most mammals, including humans, are able to localize sound sources in space by making use of the differences in the sound signals arriving at the two ears. One important physical cue is the interaural phase difference – the frequency specific difference in phase between the two ears. Traditionally, the neuronal processing of interaural phase differences has been explained by a systematic arrangement of delay lines, but recent results from electrophysiological measurements questioned this approach. It has instead been suggested, that phase differences are encoded in the relative activity between neuronal populations within the two brain hemispheres. This is also known as the two-channel hypothesis.

This thesis investigates the neuronal representation following this two-channel hypothesis by means of computational modeling. A detailed spiking-neuron-network model of the neuronal circuit responsible for encoding interaural phase differences in mammals was developed. Using this model, it is shown that the two-channel code is able to reliably encode interaural phase differences over a large frequency range, also for complex signals. To evaluate the two-channel hypothesis, this thesis derives a simplified, functional model of the neuronal responses due to interaural phase differences in tones. By evaluating this model with several psychoacoustic experiments, it is shown that the acuity and accuracy of interaural phase difference perception can be predicted directly from the model responses. These findings directly support the two-channel hypothesis, creating a direct link between neuronal representation and perception.

Zusammenfassung

Der Mensch sowie die meisten Säugetiere sind in der Lage, Schallquellen im Raum zu lokalisieren. Dafür werden die Unterschiede zwischen den an beiden Ohren anliegenden Schallsignalen genutzt. Ein wichtiges Merkmal sind die interauralen Phasenunterschiede – frequenzspezifische Phasendifferenzen zwischen den Signalen an beiden Ohren. Traditionell wurde die neuronale Verarbeitung von interauralen Phasendifferenzen durch eine systematische Anordnung von Laufzeitgliedern erklärt. Neuere Ergebnisse aus elektrophysiologischen Messungen stellten diesen Ansatz allerdings in Frage. Stattdessen wurde vorgeschlagen, dass die Phasendifferenzen in der relativen Aktivität neuronaler Populationen innerhalb der beiden Gehirnhälften kodiert werden. Dieser Ansatz wird auch als Zweikanalhypothese bezeichnet.

Diese Arbeit untersucht mittels numerischer Modelle die neuronale Repräsentation, unter Annahme dieser Zweikanalhypothese. Ein detailliertes Netzwerk-Modell des neuronalen Schaltkreises, welcher im Säugetier Phasendifferenzen kodiert, wird vorgeschlagen. Anhand dieses Modells wird gezeigt, dass der Zweikanalcode in der Lage ist, interaurale Phasendifferenzen über einen großen Frequenzbereich, auch bei komplexen Signalen, zuverlässig zu kodieren. Um die Zweikanalhypothese genauer zu evaluieren, leitet diese Arbeit ein vereinfachtes Modell der neuronalen Antworten aufgrund interauraler Phasenunterschiede in Tönen ab. Durch den Vergleich dieses Modells mit Ergebnissen mehrerer psychoakustischen Experimente, wird gezeigt, dass die Schärfe und die Genauigkeit der Wahrnehmung von interauraler Phasendifferenzen direkt anhand der Modellantworten vorhergesagt werden kann. Diese Ergebnisse befürworten die Zweikanalhypothese und stellen anhand des Modells eine direkte Verbindung zwischen der neuronalen Repräsentation und der Wahrnehmung her.

Acknowledgments

I would like to use this opportunity to acknowledge some of the people who were instrumental in the process of writing this thesis. Of course, this was foremost my supervisor, Prof. Werner Hemmert, who introduced me to the field of audition and enabled me to work on this topic. Without his continuous support, ideas and broad knowledge, I would never have been able to produce this work.

Secondly, I would like to thank my mentor Dr.-Ing Florian Völk who provided guidance throughout the whole process that lead to this work, who introduced me to the field of psychoacoustics, and supported the experiments that are detailed in this document.

Thanks also to all the other current and past members of the group for Bio-Inspired Information Processing for creating such an enjoyable working environment. I would further like to acknowledge everyone involved in the priority program "Ultrafast and temporally precise information processing" of the Deutsche Forschungsgemeinschaft (DFG) and all the other people with whom I was able to collaborate.

Of course, I would never have reached the point of starting to work on this thesis without the support of my parents and two sisters. A lot of tribute also goes to my partner Eva for all the support and motivation that she provided during these years.

Munich, 25. September, 2018

Contents

List of Publications

Peer-Reviewed Articles

Jörg Encke and Werner Hemmert. Extraction of inter-aural time differences using a spiking neuron network model of the medial superior olive. *Frontiers in Neuroscience*, 12, Mar 2018. doi: 10.3389/fnins.2018.00140.

Jörg Encke, Jasmin Kreh, Florian Völk, and Werner Hemmert. Conversion of sound into auditory nerve action potentials. *HNO*, Oct 2016. ISSN 1433-0458. doi: 10.1007/s00106-016-0258-z.

Jörg Encke, Diana Reimann, Werner Hemmert, and Völk Florian. On the role of interaural level differences in low-frequency pure-tone lateralization. *Acta Acustica united with Acustica*, 104, 2018.

Talks

Jörg Encke and Werner Hemmert. The relative timing of inhibitory and excitatory currents tunes the shift of best ITD in a detailed mammalian MSO network model. In *The Auditory Model Workshop, Oldenburg*, 2015a.

Jörg Encke and Werner Hemmert. A spiking neuron network model of ITD detection in the medial superior olive. In *Midterm Colloquium and Summer School DFG Priority Program 1608, Kaiserslautern*, 2015b.

Jörg Encke and Werner Hemmert. ITD predictions for complex sounds using a spiking neuron network model. In *39th MidWinter Meeting of the Association for Research in Otolaryngology (ARO), San Diego*, 2016a.

Jörg Encke and Werner Hemmert. Precise sound localization with spiking neuron networks. In *Bernstein Retreat 2016, Tutzing*, 2016b.

Jörg Encke, Swei Bai, and Werner Hemmert. A biophysical model of the electrically stimulated auditory nerve. In *Progress Report / Summer School DFG Priority Program 1608*, 2017.

Jörg Encke, Diana Reimann, Werner Hemmert, and Florian Völk. On the role of interaural level differences in low-frequency pure-tone lateralization. In *International Symposium on Hearing (ISH), Snekkersten*, 2018a.

Jörg Encke, Florian Völk, and Werner Hemmert. (Invited Talk) A model of the neuronal coding of temporal localization cues. In *44th Jahrestagung für Akustik (DAGA), München*, 2018b.

Jörg Encke, Florian Völk, and Werner Hemmert. A model of the neuronal processing of interaural time disparities. In *175th Meeting of the Acoustical Society of America (ASA) Minneapolis*, 2018c.

Conference Proceedings

Jörg Encke, Jasmin Kreh, Werner Hemmert, and Florian Völk. Evaluierung eines Verfahrens für psychoakustische Lateralisationsexperimente. In *Fortschritte der Akustik, DAGA 2018*, Berlin, Mar 2018a. Deutsche Gesellschaft für Akustik e.V.

Jörg Encke, Diana Reimann, Werner Hemmert, and Florian Völk. Ein Zweikanalmodell der neuronalen Verarbeitung interauraler Phasendifferenzen. In *Fortschritte der Akustik, DAGA 2018*, Berlin, Mar 2018b. Deutsche Gesellschaft für Akustik e.V.

Jasmin Kreh, Jörg Encke, Werner Hemmert, and Florian Völk. IPD, nicht ITD: Psychoakustische Hinweise auf phasennormierte Feinstrukturverarbeitung bei der Lateralisation. In *Fortschritte der Akustik, DAGA 2018*, Berlin, Mar 2018. Deutsche Gesellschaft für Akustik e.V.

Diana Reimann, Jörg Encke, Werner Hemmert, and Florian Völk. Lateralisation bei symmetrisch und asymmetrisch definierten interauralen Pegeldifferenzen. In *Fortschritte der Akustik, DAGA 2018*, Berlin, Mar 2018. Deutsche Gesellschaft für Akustik e.V.

Florian Völk, Jörg Encke, Jasmin Kreh, and Werner Hemmert. Relevance of headphone characteristics in binaural listening experiments: A case study. In *143rd Convention of the Audio Engineering Society*, Oct 2017.

Florian Völk, Jörg Encke, Jasmin Kreh, and Werner Hemmert. Pure-tone lateralization revisited. In *Proceedings of Meetings on Acoustics*, volume 31, page 050004. Acoustical Society of America, 2017. doi: 10.1121/2.0000784.

Posters

Siwei Bai, Jörg Encke, Robin Weiss, Friederike Schäfer, Frank Böhnke, Braun Katharina, and Werner Hemmert. A high-resolution cochlear model with reconstructed cochlear nerve fibres. In *40th MidWinter Meeting of the Association for Research in Otolaryngology (ARO), Baltimore*, 2018.

Jörg Encke and Werner Hemmert. The relative timing of inhibitory and excitatory currents tunes the shift of best ITD in a detailed mammalian MSO network model. In *Bernstein Conference, Heidelberg*, 2015a. doi: 10.12751/nncn.bc2015.0158.

Jörg Encke and Werner Hemmert. A spiking neuron network model of ITD detection in cochlear implant patients. In *Conference on Implantable Auditory Prostheses (CIAP), Lake Tahoe*, 2015b.

Jörg Encke, Christian Keine, Meenakshi Asokan, Rudolf Rübsamen, and Werner Hemmert. A model of acoustically evoked inhibition in spherical bushy cells. In *38th MidWinter Meeting of the Association for Research in Otolaryngology (ARO), Baltimore*, 2015.

Jörg Encke, Siwei Bai, Niclas Bähr, and Werner Hemmert. A biophysical model of the auditory nerve based on high-resolution microCT scans. In *Conference on Implantable Auditory Prostheses (CIAP) 2017, Lake Tahoe*, 2017a.

Jörg Encke, Siwei Bai, Niclas Bähr, and Werner Hemmert. Reconstruction of a high-resolution cochlear model for cochlear implant research. In *Bernstein Retreat 2017, Tutzing*, 2017b.

Jörg Encke, Siwei Bai, and Werner Hemmert. Neuronal activation pattern of auditory nerve fibres due to cochlear implant stimulation: A modeling study. In *40th MidWinter Meeting of the Association for Research in Otolaryngology (ARO), Baltimore*, 2018.

Helge Gleiss, Todd Jennings, Jörg Encke, Benedikt Grothe, and Michael Pecka. Dynamic range adaptation of monaural inputs for efficient population encoding of sound source locations under naturalistic conditions. In *ARO 39th MidWinter Meeting of the Association for Research in Otolaryngology (ARO), San Diego*, 2016.

Christian Keine, Bernhard Englitz, Jörg Encke, Werner Hemmert, and Rudolf Rübsamen. Functional properties of spherical bushy cells assessed by spectro-temporal receptive fields. In *39th MidWinter Meeting MidWinter Meeting of the Association for Research in Otolaryngology (ARO), San Diego*, 2016.

Christian Wirtz, Jörg Encke, Peter Schleich, Peter Nopp, and Werner Hemmert. Comparing different models for sound localization within normal hearing and cochlear implant listeners. In *Conference on Implantable Auditory Prostheses (CIAP), Lake Tahoe*, 2015.

CHAPTER 1

Introduction

The ability to determine the location of sound sources in space is one of the more inconspicuous and yet essential abilities of the auditory system. By extending the range of space that can be perceived by vision to a full 360° (Rayleigh, 1907) and to situations where vision is restricted such as in dense vegetation or in the dark, it poses an important skill to both hunter and pray. Spacial hearing further aids communication by helping to orient a talker of interest within multiple sources (Bronkhorst, 2015) – a task that has aptly been named the cocktail party problem (Cherry, 1953). It is thus unsurprising that the ability to localize sound sources has been developed, at least to some degree, in most higher animals (Grothe and Pecka, 2014).

However, hearing-impaired listeners often have trouble locating sound sources (Husler et al., 1983; Noble et al., 1994), also when using hearing aids (Van den Bogaert et al., 2006) or when aided with a cochlear implant (CI) (Seeber et al., 2004; van Hoesel and Tyler, 2003) – a neuroprosthesis that can restore hearing in case of severe to profound sensorineural hearing loss. One explanation for this limitation in aided listeners is that sound localization relies on a highly precise temporal coding, which might be distorted by the signal processing within the aiding device. In the case of CIs, where a hearing sensation is evoked by electrically stimulating the auditory system thus bypassing the sensory neurons, there might additionally be a degradation or even omission within the neuronal representation of the physical cues (van Hoesel and Tyler, 2003; Kerber and Seeber, 2012). To understand these limitations and to improve on the strategies used to encode the physical cues requires in-depth knowledge of the underlying system.

There is a long history in the research of spacial hearing in general as well as in the neuronal representation of localization cues (Rayleigh, 1907; Von Békésy and Wever, 1960; Jeffress, 1948). Both, the representation itself and also the way this representation is analyzed to create a hearing sensation are still disputed (McAlpine et al., 2001; Belliveau et al., 2014; Bernstein and Trahiotis, 2017). One way to investigate and validate the different hypotheses is by applying computational models of the different processes involved in sound localization such as the models proposed in this thesis. Ultimately such models try, in some form, to explain the connection between a sound event and the perception evoked by this event.

In his book, "Räumliches Hören" (English translation: Blauert, 1999), Jens Blauert

described the process that connects a property of a sound event to a psychoacoustic measurement by a series of processing blocks (see Fig. 1), where a sound event is received by a perceiving system (e.g., the subject of the experiment) which might result in the perception of an auditory event (following Fastl and Zwicker (2007) from here on called hearing sensation). The hearing sensation can be characterized by the describing system (again the subject of the experiment) on a psychophysical scale resulting in a description of some properties of the hearing sensation.

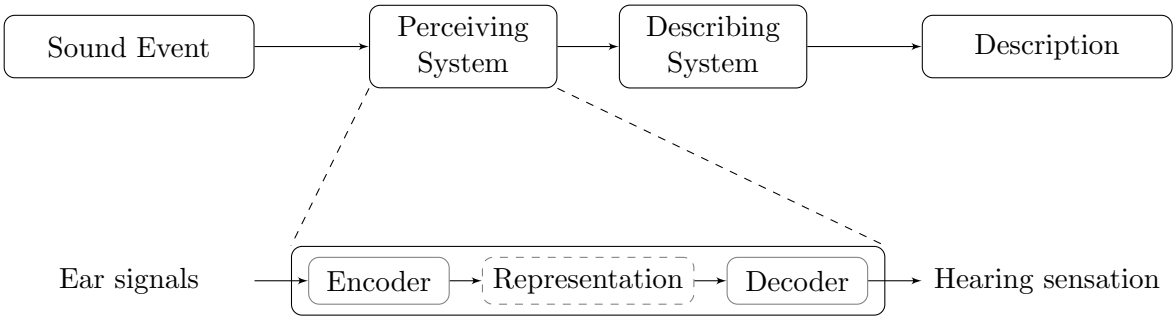

Figure 1.1: Top: Schematic representation of the process that leads from a sound event to a psychoacoustic measurement following Blauert (1999). Bottom: Breakdown of the stages within the perceiving system following an encoder/decoder model.

As noted by Blauert (1999), the hearing sensation itself, as well as the process that results in its creation, are not directly accessible in a psychoacoustic experiment. As a result, the subject is treated as a *black box* that takes a sound event as the input and gives a description of the hearing sensation as the output. Models are often used to better understand the internal working of this black box with the problem that multiple different models might capture the same input-output relationship. In sound localization, for example, multiple models based on very different assumptions about the internal representation successfully capture the outcome of psychoacoustic experiments (e.g. Stern and Shear, 1996; Dietz et al., 2009; Takanen et al., 2014). One way to constrain these models is by applying knowledge about the internal representation which can be obtained from additional measurement modalities. Such modalities include neuroimaging methods such as Electroencephalography (EEG) or Magnetoencephalography (MEG) as well as invasive electrophysiological measurements such as single- or multi-unit recordings of neuronal responses.

In this thesis, the internal structure of the perceiving system is described using an encoder/decoder model as shown in Fig. 1. In this model, an encoder describes a stage in the neuronal circuit that translates a property of the sound event into a neuronal representation of the same. This representation is then processed while ascending through the auditory pathway. At some point within this pathway, the neuronal representation is then interpreted by a decoder which might result in a

hearing sensation.

Following this framework, the decoder is only able to make use of information that is encoded in the neuronal representation so that any limitation within this encoding should also affect perception. As aforementioned, physiological measurements can be used to gain quantitative information about the neuronal representation and any model of the perceiving system should ultimately be consistent with physiology while being able to explain the resulting percept. The goal of this thesis was to investigate the neuronal representation of interaural phase differences (IPDs), which are one of the physical cues used to localize sound sources in space. This is done by developing a model of the assumed neuronal representation which is based on existing physiological data. The predictive power of this model is then evaluated based on psychoacoustic experiments with the goal of creating a direct link between the neuronal representation and behavioral data.

1.1 Structure and Overview

The content of this thesis is structured into six chapters. While each of the chapters builds on the previous one, they are written such that they can be read independently as each contains an individual introduction and discussion.

Chapter 2 is intended to introduce the unfamiliar reader to the necessary physical, neuronal and psychoacoustical principles of sound localization as well as models thereof. It focuses on the topics most important in the context of this thesis with more detailed introductions given in the corresponding chapters.

Chapter 3 proposes a novel biophysically-plausible model of the neuronal circuit of the medial superior olive (MSO) which is thought to be the encoding stage for temporal interaural disparities in the mammalian brainstem. The model accounts for the auditory periphery and responses are calculated using spiking-neuron models of two stages within the brainstem. Following this encoding stage, the temporal disparities are represented in the form of a two-channel code. Using a simple linear as well as an artificial neuronal network (ANN) decoder, it is shown that the modeled representation contains sufficient information to reliably encode both static and time-dependent time disparities over a broad frequency range.

Chapter 4 closely investigates the internal representation of IPDs following the two-channel hypothesis and a simplified, statistical model of this internal representation based on existing electrophysiological data is proposed. Results of a left-right discrimination task are used to determine the free model parameters, and it is shown that the model can replicate the results of this experiment. In a second psychoacoustic experiment, the appearance of multiple hearing sensations at large IPD values is investigated, and the implications are discussed in light of the two-channel model.

Chapter 5 uses the model introduced in the previous chapter to predict IPD detection thresholds at different frequencies and around difference reference IPDs which are

determined in a psychoacoustic experiment. It is further shown that the same model also predicts a high-frequency limit for the sensitivity to interaural phase differences which, in agreement with previous literature, is located at around 1.5 kHz

Chapter 6 introduces a concept of using the two-channel model to represent signals with multiple frequency components. Using a likelihood measure, it is shown that the results of the two-channel model can be represented in a manner that is very similar to the cross-correlogram used by delay-line models.

Chapter 7 investigates the influence of small low-frequency interaural level differences (ILDs) as they would occur in a natural environment on the perception of IPDs especially at higher frequencies where phase wrapping introduces ambiguities in the range of naturally occurring IPDs.

CHAPTER 2

Fundamentals and Models of Sound Localization

The process that connects a sound event to a hearing sensation as discussed in the Introduction (see Fig. 1) can be separated into three domains. Sound events are characterized by physical measurements and are, as such, described in the *physical domain*. Any information in the perceiving system is regarded as information about the neuronal representation. It is therefore characterized in the *neuronal domain*. The hearing sensation itself can only be assessed by obtaining a description using psychophysical measures which are attributed to the *perceptual domain*. This chapter is intended to present a short overview over the general process of sound localization as well as existing models of the same. The chapter is split into four sections, Sec. 2.1, 2.2 and 2.3 are structured following the three domains by discussing the physical, neuronal and perceptual background while Sec. 2.4 provides a short overview over existing models of the neuronal representation of interaural time disparities. Each of these chapters does not aspire to draw a complete picture but to provide a quick introduction into those principles most important in the scope of this thesis.

2.1 Physics

This section discusses the influence of sound source location on the physical properties of the signals at the two ears and introduces some definitions that will be used throughout the thesis. As this thesis exclusively discusses sound sources in the horizontal plane, this section will be limited to this spacial dimension and neglect any influence due to the elevation of a sound source. A more in-depth discussion can be found in the literature (e.g. Blauert, 1999).

A head centered coordinate system, as shown in Fig. 2.1, with the azimuthal angle being defined anti-clockwise, will be used throughout the thesis. The horizontal plane as shown in Fig. 2.1 is intersected by the median plane (MP). A location to the left of the MP (positive azimuth) will be denoted as *left* while any location to the right of the MP (negative azimuth) will be denoted as *right*. Following a popular notation, results associated with spatial locations to the left will usually be visualized using the

color blue while results associated with spatial locations to the right will be visualized using the color red.

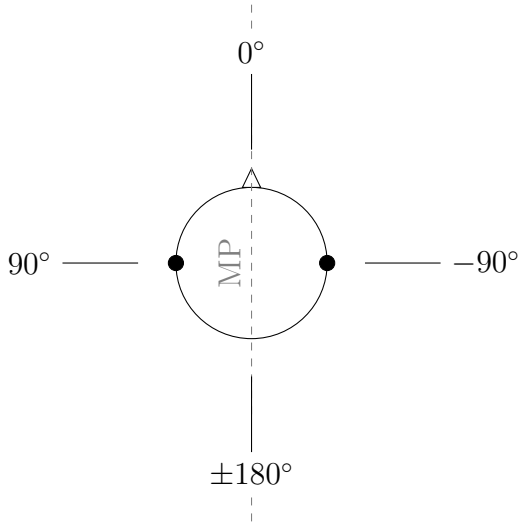

Figure 2.1: Schematic of the horizontal plane according to the coordinate system used throughout the thesis. The horizontal plane is intersected by the median plane (MP). Locations to the left of the MP will be denoted as 'left' while locations to the right of the median plane will be denoted as 'right'.

Information about the location of a sound source in the horizontal plane is conveyed by differences in the sound pressure waveform at the left and right eardrum (ear-signals) (e.g. Rayleigh, 1907; Von Békésy and Wever, 1960). Under natural conditions, these differences, which are also called interaural differences are introduced by the different locations of the two ears within the sound field. Following Fourier's theorem, the interaural differences of each frequency component can also be decomposed into a difference in amplitude (or level) and a difference in phase. In addition, both differences can also depend on time. As a consequence, the differences in the ear-signals due to a given sound source location can be defined by a frequency and time-dependent interaural level difference (ILD) $\Delta L(f,t)$ and a frequency and time-dependent interaural phase difference (IPD) $\Delta\varphi(f,t)$.

IPDs are mainly introduced by a difference in the effective distance the sound signal has to travel until reaching each ear. Any wavefront of a sound signal located to the left will reach the left ear earlier than the right ear which results in a difference of arrival time. For a defined frequency f, the resulting IPD can be calculated as

$$\Delta\varphi(f) = 2\pi f(t_{\mathrm{a,\,l}} - t_{\mathrm{a,\,r}}), \tag{2.1}$$

where $t_{\mathrm{a,\,l}}$ and $t_{\mathrm{a,\,r}}$ are the arrival time of a given wavefront at the left and right ear respectively. By using this definition, positive IPD values are associated with a sound

source to the right of the MP while negative IPD values are associated with a sound source to the left of the MP. To prevent any confusion between the azimuth and IPDs, IPDs will be stated in radian while azimuths will be stated in degree. As the difference in arrival time $\Delta t = (t_{\mathrm{a,\,l}} - t_{\mathrm{a,\,r}})$ for a specific position in space is nearly independent of frequency, many studies use the notation of interaural time difference (ITD) instead of IPD. The two measures can be used interchangeably, but it has to be noted that a frequency independent ITD results in frequency dependent IPDs $\Delta\varphi(f)$ which becomes relevant for signals with multiple frequency components.

ILDs are defined as the difference in sound pressure level (SPL) between the two ear-signals and will be calculated as

$$\Delta L = L_r - L_l, \tag{2.2}$$

where L_r is the SPL of the right and L_l the SPL of the left ear-signal so that a sound source location to the left will mostly result in a negative ILD while a sound source to the right will mostly result in a positive ILD. Level differences are mainly introduced by a shadowing effect of the head due to diffraction (Sivian and White, 1933). The effect of diffraction strongly depends on the sound frequency with near to no effect at lower frequencies and an increasing effect at higher frequencies. As a consequence, ILDs also show a strong frequency dependence which is in contrast to time differences which are, as a first approximation, independent of frequency (see Chapter 7 for more details).

Under natural conditions, ILDs and IPDs do not exist independently as a specific location of the sound source always results in both interaural differences. Using headphones instead of a sound source in space enables the independent presentation of the two ear-signals so that $\Delta L(f,t)$ and $\Delta\varphi(f,t)$ can be defined independently from each other[i].

2.2 Physiology

This section provides a short overview of the most important stages of the mammalian brainstem that are involved in sound localization. As in the previous section, the discussion will be limited to the two cues ILD and IPD that are of importance in the scope of the thesis. A complete overview can be found in the literature (e.g., reviewed in Grothe et al., 2010).

At each of the two ears, the ear-signals are converted independently into action potentials which are conducted by the auditory nerve fibers (ANFs) to the cochlear nucleus (CN) in their respective hemisphere. As discussed in the previous section, the information of both ear-signals and thus of both hemispheres is necessary to

[i] This only holds if the transfer functions associated with conduction of sound from the actuator to the tympanic membrane are assumed to be symmetric

convey information about the location of a sound source. The first stage at which this convergence of inputs of both hemispheres onto the same neuron takes place is within the superior olivary complex, a collection of nuclei located in the brainstem of mammals. Within the olivary complex, the most important nuclei in respect to sound localization are the lateral superior olive (LSO) and the medial superior olive (MSO).

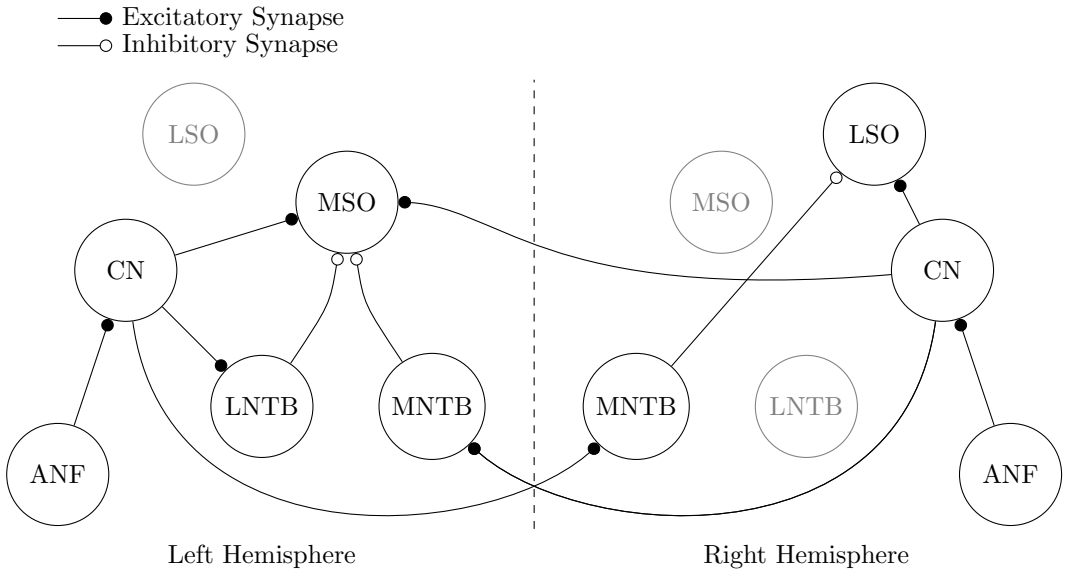

Figure 2.2: Schematic of the neuronal sound localization pathway in the brainstem. To improve readability, only the connections of the MSO in the left and the LSO in the right hemisphere are shown.

The principal neurons of the LSO receive excitatory inputs from spherical bushy cells (SBCs) located in the ipsilateral CN (Cant and Casseday, 1986) while receiving inhibitory inputs from neurons in the medial nucleus of the trapezoid body (MNTB) (Spangler et al., 1985; Friauf and Ostwald, 1988) which are innervated by globular bushy cells (GBCs) of the contralateral CN (Harrison and Warr, 1962) (see Fig. 2.2). This excitation, inhibition (EI) circuit can be interpreted as a subtraction of the contralateral input from the ipsilateral input which would be predestinated to analyze ILDs. Unsurprisingly, neurons in the LSO have been found to be sensitive to ILDs (Boudreau and Tsuchitani, 1968). As expected from the circuit, most LSO neurons of a given hemisphere respond strongest if the sound level of the ipsilateral ear-signal (excitatory input) is higher than the sound level of the contralateral ear-signal (inhibitory input) resulting in approximately sigmoidal ILD response functions (see Fig. 2.3a). Some studies have shown that neurons in the LSO are also sensitive to time differences due to the resulting shift in the stimulus envelope (Joris and Yin, 1995; Ashida et al., 2016) or when using clicks (Beiderbeck et al., 2018) but when following the definition of Sec. 2.1, a time difference in the envelopes of the two ear-signals

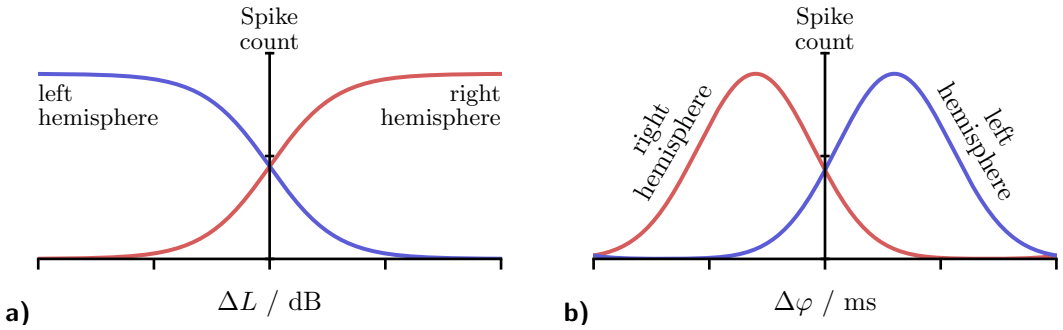

Figure 2.3: Schematic of typical responses of neurons in the LSO and MSO of both hemispheres. **a)** Typical sigmoid response of neurons in the LSO to stimuli with different ILDs. **b)** Typical bell-shaped response of MSO neurons to stimuli with different IPDs.

would be interpreted as a time-dependent ILD.

For frequencies below approximately 4 kHz, responses coded in the ANF are phase-locked to the ear-signals meaning that the probability of ANFs to fire an action potential is largest at a specific phase of the stimulus. This phase-locked firing of the ANFs allows the auditory system to detect IPDs by comparing the timing of the responses of the left and right ANFs. Principal neurons in the MSO receive phase-locked excitatory inputs from contra- and ipsilateral SBCs as well inhibitory inputs from the ipsilateral lateral nucleus of the trapezoid body (LNTB) and MNTB (Spangler et al., 1985) (see Fig. 2.2). This excitatory, excitatory (EE) circuit functions as a coincidence detector where the neuron responds strongest if the inputs from the two hemispheres arrive simultaneously (no time difference - zero IPD). The response decreases with increasing phase or time difference between the two inputs resulting in bell-shaped IPD-response functions. Neurons in the MSO have been found to follow this response pattern with the difference that most neurons do not show maximal response for zero IPD but for IPDs that correspond to contralateral leading ear-signals so that MSO neurons in the left hemisphere predominantly show maximal response for positive IPDs while MSO neurons in the right hemisphere predominantly respond strongest to negative IPDs (see Fig. 2.3b). While the origin for the shift in the maximum of the IPD-response function (also called the best-IPD) towards contralateral leading IPDs is still under discussion (Joris et al., 2006; Franken et al., 2015; Brand et al., 2002), most studies reported the best-IPDs of the majorities of IPD sensitive neurons to fall within the range between zero and π (McAlpine et al., 2001; Joris et al., 2006; Pecka et al., 2008)[ii].

To summarize, the brainstem circuitry seems to be divided into two pathways, IPDs are primarily analyzed in the MSO while ILDs seem to be analyzed in the LSO. As the information about both, IPDs and the ILDs, is only available after combining the inputs from the two ears in the MSO and LSO, the two nuclei will be considered as

[ii]The data in the two studies McAlpine et al. (2001) and Joris et al. (2006) was obtained in the inferior colliculus which is the next higher processing stage after the MSO.

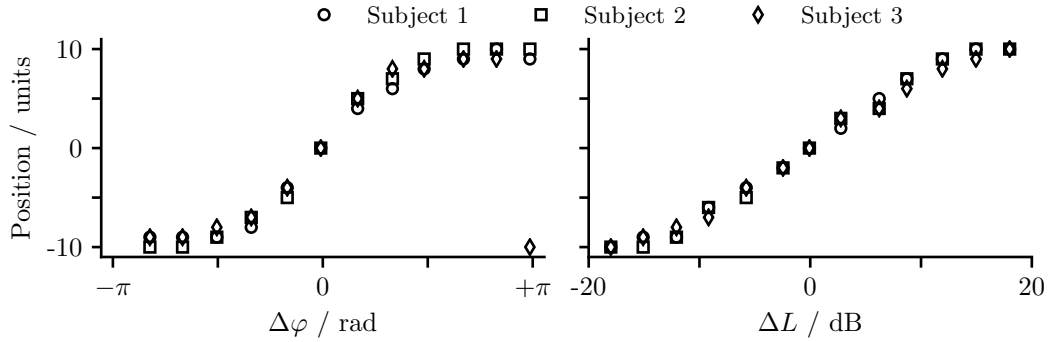

Figure 2.4: The perceived lateral position of the hearing sensation for a 200 Hz tone at different IPDs and ILDs as indicated on a scale of 21 values where -10 indicates fully left lateralized, and ten indicates fully right lateralized (Data for three subjects digitized from Yost (1981) Fig. 2 and Fig. 7)

the encoder for IPD and ILD.

2.3 Perception

The following section provides a concise overview of the effect of introducing IPDs and ILDs on the perceived location of a sound source. As all experiments presented in this thesis use stimuli presented over headphones, this section will also focus on this setting. A more in-depth discussion of the topic can be found e.g. in Blauert (1999).

Under natural conditions, listeners are able to assign a perceived spatial location, positioned mostly outside of the head, to a hearing sensation. In contrast to this, if the sound is presented via headphones, the hearing sensation is mostly perceived to be located inside the head (Sayers and Cherry, 1957; Plenge, 1974). Due to this difference, two terms are usually used to describe the two types of spatial perception. If the hearing sensation is perceived outside of the head, it is described as being *localized* while, it is described as being *lateralized* if it is perceived inside of the head (intracranial location). As it is hard to assign a physical location to an intracranial hearing sensation, the extent of lateralization is often stated in terms of the position of the hearing sensation on the interaural axis - an axis connecting the two ears (e.g. Sayers, 1964; Yost, 1981).

Introducing positive IPDs, for example by delaying the signal of the left ear, mostly results in a hearing sensation that is lateralized towards the right while introducing negative IPD mostly results in a hearing sensation lateralized towards the left (see Fig. 2.4). At larger IPDs close to $\pm\pi$, it is often not possible to unambiguously report the lateralization as these stimuli often result in multiple, spatially separated hearing sensation (see also Sec. 4.4). The sensitivity to changes in IPDs is often presented in terms of ITDs, where the smallest detectable change of ITDs is in the range of 10 µs

at 1000 Hz, but this value depends strongly on the stimulus frequency with higher thresholds at lower frequencies and a steep increase in thresholds towards higher frequencies (Zwislocki and Feldman, 1956). If the same values are presented in terms of IPDs, one instead finds a monotonic increase in IPD thresholds with frequency, the smallest values thus found at low frequencies. IPD thresholds will be discussed in detail in Chapter 5.

In agreement with the introduction of positive IPDs, introducing a positive ILD also results in a hearing sensation that is lateralized towards the right while a negative ILDs result in a lateralization in the opposite direction (see Fig. 2.4). Contrary to the findings with IPDs, ILD difference thresholds do not change significantly with frequency (Yost and Dye, 1988) reaching difference thresholds below 1 dB. As previously discussed in Sec. 2.1, introducing ITDs also introduces a time-dependent ILD due to the shift of the stimulus envelope. While the sensitivity to IPDs is limited to lower frequencies, listeners can make use of time shifts in the envelope at higher frequencies (Henning, 1974).

As aforementioned, ILDs and IPDs do not exist independently under natural conditions, and several studies have shown interactions between the two cues (e.g. Moushegian and Jeffress, 1959; Sayers, 1964; Domnitz and Colburn, 1977). The nature of this interaction for lower frequencies is also discussed in Chapter 7.

2.4 Models of the Neuronal Representation

This chapter gives a short overview over existing models for the neuronal representation of ITDs or IPDs. It focuses on two important concepts, the delay-line based and the rate models and, in favor of clarity, neglects some alternative concepts like the equalization-cancellation model (EC-model) (see Colburn and Durlach (1978) for an overview of EC-models). The chapter is intended to introduce the basic concepts without reviewing individual models as good overviews over existing models can be found in the literature (e.g. Stern et al., 2006; Dietz et al., 2018).

In 1948 Lloyd A. Jeffress published an influential paper (Jeffress, 1948) in which he hypothesized an ITD encoder based on a systematic arrangement of coincidence detecting neurons located along a delay line. These neurons respond strongest if the input of the two ears arrive simultaneously so that a given neuron responds strongest if the specific ITD is compensated by the axonal runtime of the neuron. Figure 2.5 shows a schematic of such an arrangement. In this example, neuron 1 will show maximum response if the left input is strongly delayed while neuron 10 will show the maximum response for a strong delay of the right input with the other neurons tuned to intermediate delays.

This kind of encoder creates a neuronal representation where the ITD is encoded by the specific delay of the strongest responding neuron. This representation is visualized in Fig. 2.6a which schematically shows a set of coincidence detector neurons, each

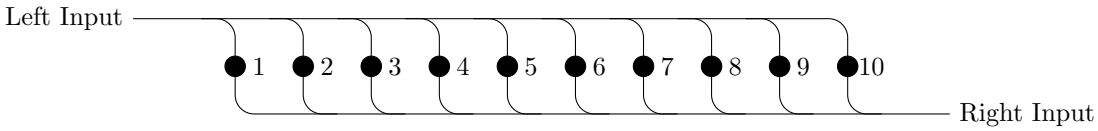

Figure 2.5: Schematic representation of a delay line circuit as found in the avian brain. The circuit is constructed by a systematic change of the axonal length from both ears into a set of coincidence detecting neurons represented by the black circles. The coincidence neurons respond maximally if the input from both ears arrives simultaneously.

narrowly tuned to a specific ITD value (color-coded in Fig. 2.6a). To decode the ITD from this representation, any decoder needs to know which delay is associated with each of the neurons or in other words, the neurons place within the delay line so that this kind of representation is often referred to as a place-code. Mathematically, this principle can be implemented in the form of a discreet cross-correlation of the inputs from both sides. Several models have been based on the idea of the cross-correlation model (e.g. Colburn, 1973; Stern and Colburn, 1985; Lindemann, 1986; Bernstein and Trahiotis, 2017) which, in this thesis, will be referred to as delay-line or Jeffress style models. Over time, these models developed to explain an impressive catalog of psychoacoustic results (see Stern et al. (2006) and Bernstein and Trahiotis (2017) for an overview). A neuronal delay line structure was also found in the avian nucleus laminaris (Carr and Konishi, 1988) but findings in the mammalian brainstem, as discussed in Sec. 2.1, are hard to reconcile with a delay-line mechanism. Specifically, most delay-line models presume neurons narrowly tuned to specific ITDs and also assume the majority of the neurons to be tuned to ITDs close to zero (Stern and Shear, 1996; Colburn, 1977). In the mammalian MSO, studies instead found rather broadly tuned neurons mostly responding strongest to non-zero ITDs within half a cycle of the neurons best-frequency (McAlpine et al., 2001; Joris et al., 2006). This half-cycle limit has also become known as the π-limit, and due to this and other properties which are easier expressed in terms of phase than in terms of time, this thesis will mainly use the measure of IPDs when referring to rate models. In addition to the frequency dependent tuning, a large proportion of the neurons was also tuned to IPDs outside of the respective natural range of small mammals (Yin and Kuwada, 1983; Goldberg and Brown, 1969). These observations lead to an alternate hypothesis in which the IPD is encoded in the neurons' firing rate instead of their tuning (McAlpine and Grothe, 2003). One way this could be implemented is by considering the summed activity in each of the two hemispheres. In this case, only two IPD-rate functions, one per hemisphere, are necessary to characterize the system which is why this model is also referred to as the two-channel model (TC-model). Some alternative names are opponent-channel (Magezi and Krumbholz, 2010), count-comparison (Colburn and Durlach, 1978) or hemifield model (Stecker et al., 2005). Figure 2.6b shows a schematic of this neuronal representation. Following the TC-model, specific IPD is encoded in

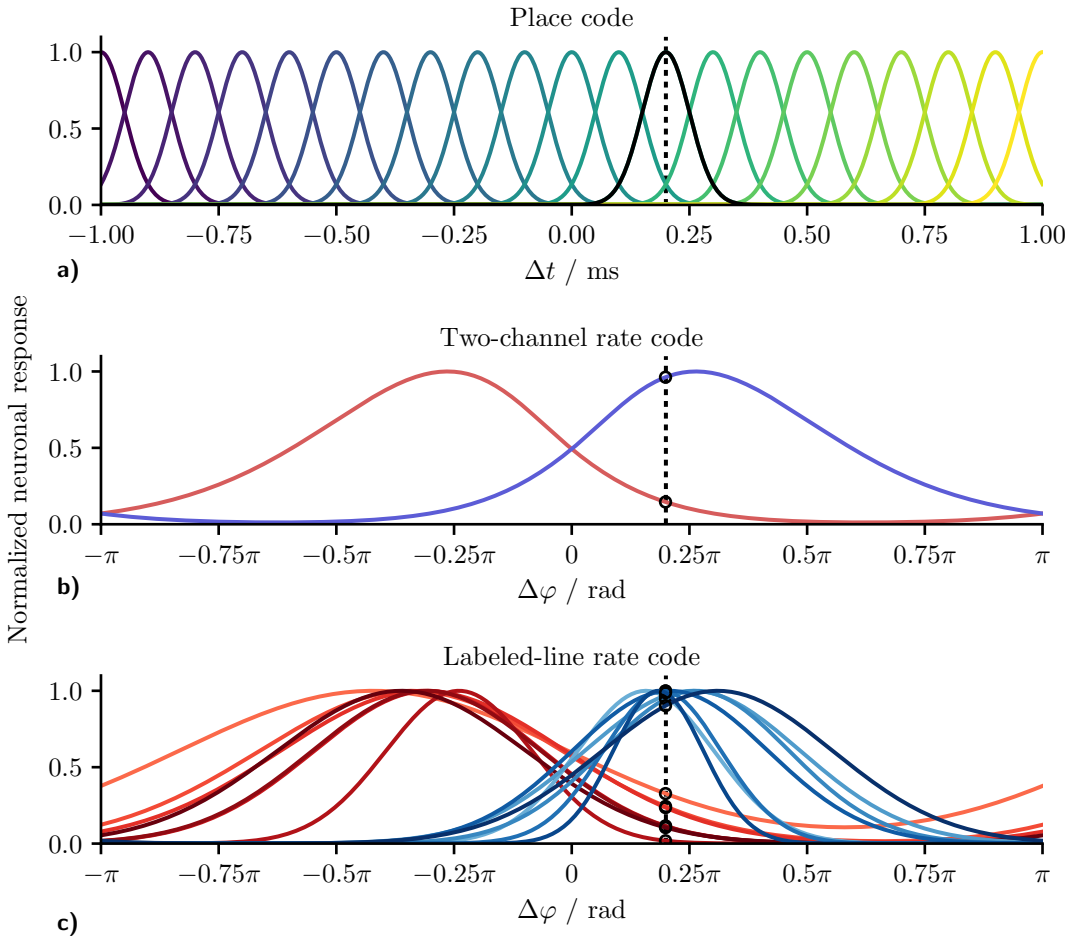

Figure 2.6: Comparison of three commonly assumed neuronal representations of interaural time or in case of the Hemispheric model phase differences. **a)** Schematic of the neuronal representation following the Jeffress style place-code. A set of coincidence detector neurons, each narrowly tuned to a specific ITD values (color coded), encode a specific ITD (dashed black line) in form of the specific delay of the neuron or the group of neurons with the strongest response (black curve) **b)** Neuronal representation of IPDs following the TC-model rate code . In this model, the IPD is encoded in the summed or averaged response of all neurons in each of the hemispheres. Given a certain IPD (dashed black line) two responses, one for each hemisphere, (black circles) are necessary to code for the specific IPD. **c)** Schematic of the labeled-line rate code. The graph shows an example of individual neuronal responses in the left (blue curves) and right (red curves) hemisphere. A specific IPD (dashed black line) is coded in the individual or sub-grouped responses (black circles) in both hemispheres.

the relative firing rates within the two hemispheres. Some previous models that make use of this type of encoding are the models proposed by Takanen et al. (2014) and, at least to some degree the model of Dietz et al. (2009) even though the latter uses both a mixture of TC and place code.

An extension of the two-channel rate code is the labeled-line rate code. This code also considers the variations in the IPD-rate functions of the individual or sub-grouped neurons. As a result, this representation uses multiple responses per hemisphere to code for an IPD (see Fig. 2.6c) and assumes knowledge about the individual IPD-response functions. An example for models employing this representation is the models proposed by Hancock (2004), and the implications of this type of code were further investigated by Goodman et al. (2013) and Belliveau et al. (2014).

CHAPTER 3

A Biophysical Model of the Medial Superior Olive

The content of this chapter was previously published as a peer-reviewed article[i]:

Encke, J. and Hemmert, W. Extraction of inter-aural time differences using a spiking neuron network model of the medial superior olive. *Frontiers in Neuroscience*, 12, Mar 2018. doi: 10.3389/fnins.2018.00140

It is reproduced here without content-related changes. Figures have been reformatted to fit the format of this document but are otherwise unchanged. Due to historical reasons, this chapter deviates from the convention of using interaural phase differences to address the effect of an interaural delay on the fine structure of a signal and instead uses interaural time differences.

3.1 Introduction

Our remarkable sound localization acuity relies on the ability of the auditory system to decode the arrival time and intensity difference between the ear canal signals into information about the direction of sound sources. In mammals, the primary nucleus to extract fine structure interaural time differences (ITDs) is the medial superior olive (MSO), while the interaural level differences (ILDs) are extracted primarily at the lateral superior olive (LSO) (Grothe et al., 2010). The MSO neurons detect fine-structure ITDs by acting as coincidence detectors receiving excitatory inputs from both hemispheres. The existence of such neurons was already hypothesized by Jeffress (1948), who proposed an array of coincident detectors to be arranged along a neural delay line. In this hypothesis, each neuron would respond maximally to a specific ITD (best-ITD) – generating a topographical mapping of time differences within the nucleus. Later, such a circuit was found in the nucleus laminaris of birds

like the barn owl (Carr and Konishi, 1988). However, more recent measurements of mammalian inferior colliculus (IC) and MSO neurons in gerbils (Brand et al., 2002) or guinea pigs (McAlpine et al., 2001) revealed broadly-tuned neurons, of which the majority had their best-ITDs at the border or even outside of the animals physiological range. This observation is inconsistent with place-code theory, which would require a vast amount of narrowly-tuned neurons with their best-ITDs distributed within the physiological range. One alternative ITD-coding mechanism is based on the comparison of firing rates between the nuclei in the two hemispheres. This mechanism has consequently been called the opponent-channel (Magezi and Krumbholz, 2010), count-comparison (Colburn and Durlach, 1978) or hemifield (Stecker et al., 2005) model. The opponent-coding model is in agreement with both observations, the wide tuning curves and the large best-ITDs (McAlpine and Grothe, 2003). There is also evidence that overall sound localization (Stecker et al., 2005; Briley et al., 2012) as well as specifically ITD-coding in the human auditory cortex is based on an opponent coding mechanism (Salminen et al., 2010). Lesion studies in cats showed that unilateral lesions at the level of the central auditory system (Jenkins and Masterton, 1982) as well as in cortical regions (Malhotra et al., 2004) mainly resulted in deficits localizing sounds from locations contralateral of the lesion. These results lead Jenkins and Masterton (1982) to conclude that each auditory-hemifield is represented solely in the respective contralateral hemisphere, which would contradict the opponent coding mechanism. One problem with applying this interpretation to ITD processing is that both studies used broad-band stimuli so that ITDs and ILDs, as well as spectral and monaural cues were available to localize the sound source this makes it difficult to draw conclusions about the representation of the individual cue. An alternative to the opponent-channel code, which uses the summed response of the neurons within each of the two hemispheres, is the population decoder that instead uses the individual response of each neuron for decoding. Based neuronal recordings of neurons in the IC, Goodman et al. (2013) and Day and Delgutte (2013) both proposed population decoders and showed that these decoders could outperform a two-channel decoder. On the other hand, Harper et al. (2014) used an optimal coding approach to show that ITDs in low-frequency signals would be best represented by a two-channel code. Additionally, results from psychoacoustic lateralization experiments using pure-tone adapter stimuli with fixed ITDs showed, that adaptation influences lateralization at ITDs not only close to that of the adapter but within the whole hemisphere (Phillips et al., 2006), which is more in line with an opponent-channel code.

The aforementioned remarkable sound localization ability has inspired numerous researchers to create computational binaural models. Most of the existing binaural models are phenomenological implementations of the delay-line principle proposed by Jeffress (1948), which have been tuned to successfully predict data from human psychoacoustics (Lindemann, 1986). Some more recent models were implemented following the opponent-coding mechanism (Dietz et al., 2011; Takanen et al., 2014;

Pulkki and Hirvonen, 2009). Even though these models closely follow the functionality of the neuronal sound localization pathway, they provide only a phenomenological description of the processing stages. On the other hand, several biophysical models of MSO neurons have been published as well (Zhou et al., 2005; Lehnert et al., 2014; Brughera et al., 2013, 1996), but there are only a few biophysical models covering the complete neuronal circuit. Wang et al. (2013) used a circuit containing a model of the auditory periphery as well as spiking models of the MSO and LSO and a simplified IC model to investigate the sensitivity of IC neurons to envelope ITDs in high-frequency sounds. Due to the focus on high-frequency sounds where ITDs are extracted from the envelope of the sound signal instead of its fine structure (Nuetzel and Hafter, 1976), Wang et al. (2013) did not include any source for a shift in best-ITD and also neglected inhibitory inputs to the MSO. Glackin et al. (2010) presented a spiking neural network (SNN) constructed from leaky integrate-and-fire models of the CN and MSO nuclei. In disagreement with newer physiological studies, the SNN was constructed as a Jeffress-type delay-line decoder. Glackin et al. (2010) trained the network to localize the sounds using spike-timing-dependent plasticity learning rules.

To our knowledge, none of the previous models combined an SNN approach with the concept of opponent-coding to investigate ITD sensitivity. Brughera et al. (2013) presented a single spiking neuron model of the MSO to investigated ITD sensitivity, but used a periodic Poisson-like process as an input to the MSO. This limits the model to simple pure-tone-like scenarios while also neglecting any non-linear processing of the auditory periphery. To that end, we present here a new binaural model based on biophysical spiking neuron models of the mammalian MSO circuit. We show that a simple linear hemisphere decoder applied to the output of the model is sufficient to encode ITDs in tones with a precision that matches human performance. Furthermore, we show how the model in conjunction with a simple artificial neural network can decode ITDs from broadband signals, including complex signals like speech.

3.2 Results

3.2.1 Model Structure

The primary mammalian MSO neurons receive excitatory inputs from spherical bushy cells (SBCs) as well as inhibitory inputs from the globular bushy cells (GBCs) of the cochlear nuclei in both hemispheres. Inhibitory inputs are being relayed via the trapezoid body (TB) (see Grothe et al. (2010) for an overview). Both SBCs and GBCs are directly excited by auditory nerve fibers (ANFs). GBCs in particular, but also SBCs have been found to enhance phase locking of the neuronal inputs (Joris et al., 1994; Dehmel et al., 2010). Our model consists of three stages, a model of the auditory periphery, a population of globular bushy cells and a population of MSO neurons (see Fig. 3.1). For simplicity, SBC as well as the TB nuclei, were reflected as direct relays of the ANF signals so that our MSO model receives direct excitatory input from the ANF

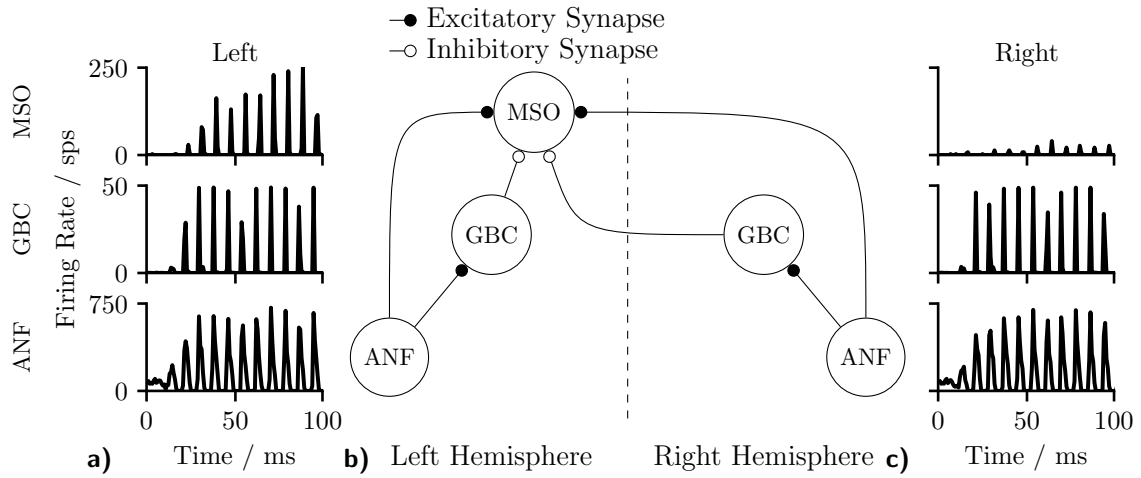

Figure 3.1: a) and **c)** Poststimulus time histograms (750 µs bin size) of the responses of the three model stages to a 100 ms long pure tone. **b)** The model network contains three stages. A model of the auditory periphery (ANF), A model of the globular bushy cells in the cochlear nucleus (GBC) and the model of the medial superior olive (MSO). (Graphic adapted from Encke and Hemmert (2018))

and inhibitory inputs from GBCs of both hemispheres (see Section 3.4 for details on the implementation). In practice, our model takes digitized binaural signals as input and processes them first through the peripheral hearing models of the left and right ears. The peripheral model consists of a middle-ear compensation filter, a non-linear model of the basilar membrane and a functional model of the neural transduction of the inner hair cell and auditory nerve fibers (Zilany et al., 2014). All ANFs were modeled as high spontaneous rate units. The spike timings of the peripheral hearing models were then used as input to the biophysical neuron models. As a consequence of the direct excitation by ANF fibers, the frequency responses of both MSO and GBCs resemble that of the ANFs from the peripheral hearing model (see Fig. S1 in the supplemental materials).

As an example of the output from the different model stages, Figs. 3.1a and 3.1c illustrate the outputs of ANFs, GBCs and the MSO of the two hemispheres for a left-leading (150 µs ITD) 125 Hz pure-tone input. The ANFs of both hemispheres show a phase-locked response to the input stimulus. This phase-locked response is sharpened by the population of GBC neurons. The MSO neurons of the two hemispheres respond with different firing rates depending on the delay between the signals delivered to the left and right ear.

Most MSO neurons of gerbils show bell-shaped ITD-rate functions with their maximum (best-ITD) located outside of the animals physiological range (Brand et al., 2002). There has been much debate about the origin of this shift ranging from intra-

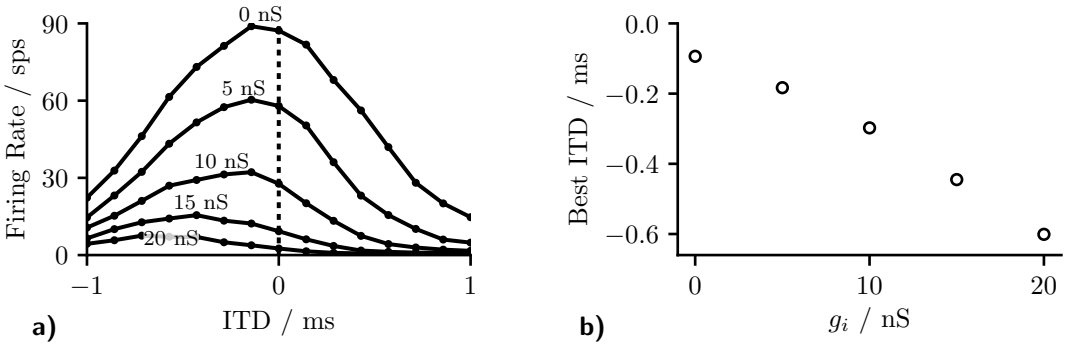

Figure 3.2: a) MSO ITD-rate functions (calculated for 15 ITDs in the range ±1 ms) for the right hemisphere of the model at different inhibitory conductivities $\hat{g}_{syn,i}$ **b)** Increased inhibition, reduces the overall firing rate and shifts the best-ITD towards more contra-lateral leading ITDs. Without inhibition the best-ITD equals the predefined shift of 100 µs. (Graphic adapted from Encke and Hemmert (2018))

cochlear delays (Joris et al., 2006) over asymmetric synaptic currents (Jercog et al., 2010) to effects of the recent stimulus history (Franken et al., 2015). Our model is based on the effect described by Brand et al. (2002) and Pecka et al. (2008), who showed that blocking of the inhibitory inputs results in a shift of the best-ITD towards zero. Measurements in gerbil brain slices have also shown that inhibitory inputs to the MSO precede the excitatory inputs in time (Roberts et al., 2013). Using conduction clamp measurements, Myoga et al. (2014) showed that the relative timing of inhibitory to excitatory inputs to the MSO can delay or advance the peak of the excitatory post-synaptic potential (EPSP) and consequently, affect the best ITD of the neurons. Our model is consistent with these findings. In agreement with Brand et al. (2002) and Pecka et al. (2008), the best-ITD shifted towards zero when simulating the effect of blocked inhibition by reducing the inhibitory synaptic strength (see Figs. 3.2a and b). Similarly, and in accordance with Myoga et al. (2014), we could shift the best-ITD of the MSO model by adjusting the delay of contra- and ipsilateral inhibitory inputs. For the model used in later evaluations, we optimized both arrival times to obtain a maximal shift of the best-ITD towards contra-leading ITDs. This optimization resulted in a delay of 0.6 ms for the contralateral inhibitory input and 0 ms for the ipsilateral input (both values relative to the timing of the excitatory input from the corresponding side). These values are in agreement with the timescales observed by both Myoga et al. (2014) and Roberts et al. (2013). The study by Pecka et al. (2008) showed a residual shift of the best-ITD even when the inhibitory inputs were blocked. This could be explained by fundamental physics as the axons connecting inputs from the contralateral hemisphere to the MSO have to span over a larger distance than the ones for ipsilateral inputs. We considered this observation by adding a constant delay of 100 µs to the contralateral excitatory and inhibitory inputs, which resulted in an

additional shift of the best-ITD towards negative values (see Fig. 3.2b).

3.2.2 Decoding ITD information from the neuronal responses.

The opponent-coding theory is based on two populations of neurons, both firing maximally when the sound source is on the opposite side of the midline (Stecker et al., 2005). Figure 3.3a shows firing rates of the MSO model in both hemispheres to a stimulation with varying ITDs. The left MSO responds strongest when the stimulus was right-leading (positive ITD), while the right hemisphere responds strongest to a left-leading ITD (negative value). Consequently, a change in ITD from zero results in an increased firing of one MSO and a reduced firing of the other. A very basic decoder for the opponent-channel code can be constructed by subtracting the firing rates of the left MSO (R_L) from the right MSO (R_R). Around zero ITD, the calculated firing rate difference $\Delta R = R_R - R_L$ shows an almost linear response to ITD changes (see Fig. 3.3b). Due to the subtraction, this approach increases the slope around zero by a factor of two and consequently maximizes the sensitivity in this region. However, this approach is applicable only for ITDs for which the linear approximation is valid. The linear ITD region depends primarily on the location of the best-ITD (see Fig. 3.3c in the two hemispheres. When calculating ITD-rate functions for neurons with different best-frequency, the best ITD decreases with increasing sound frequency (see Fig. 3.3c). The best-ITD is maximally 470 µs at 125 Hz and decreases to 110 µs at 1.4 kHz. The same trend of decreasing best-ITDs with increasing frequency has been found in *in-vivo* recordings of MSO neurons (Brand et al., 2002; Pecka et al., 2008) as well as in the IC (McAlpine et al., 2001). As aforementioned, this model mainly uses phase-locked inhibition to shift the best-ITD. This method relies on the slopes of the inhibitory post-synaptic potentials (IPSPs) of each phase (Myoga et al., 2014). At higher frequencies, the summation of individual IPSPs reduces the effectiveness in shifting the best-ITD (Roberts et al., 2013; Myoga et al., 2014), which is also seen in the model results. Experimental studies have shown that MSO and IC neurons exhibit a variety of different best-ITDs (Bremen and Joris, 2013; McAlpine et al., 2001), while in this model, all neurons with the same best frequency also show the same best-ITD. As this study does not use a population decoder but relies on the mean activity within each hemisphere, the single ITD-rate function can also be interpreted as the mean ITD-rate function of a single hemisphere.

The sensitivity of the linear-decoder to ITD changes is proportional to the slope of the ΔR function around zero ITD– a steeper slope results in larger changes. As the slope of the ΔR function around zero is twice the slope of a single hemisphere response, maximizing the slope of the single hemisphere will also result in a maximal slope of ΔR. Pecka et al. (2008) and McAlpine et al. (2001) both reported the maximal slope of single neuron responses to be located at or close to mid-line. In this model, responses to frequencies up to 700 Hz followed these findings (see Fig. 3.3e). At higher frequencies, the location of the largest slope started to shifted away from midline as the

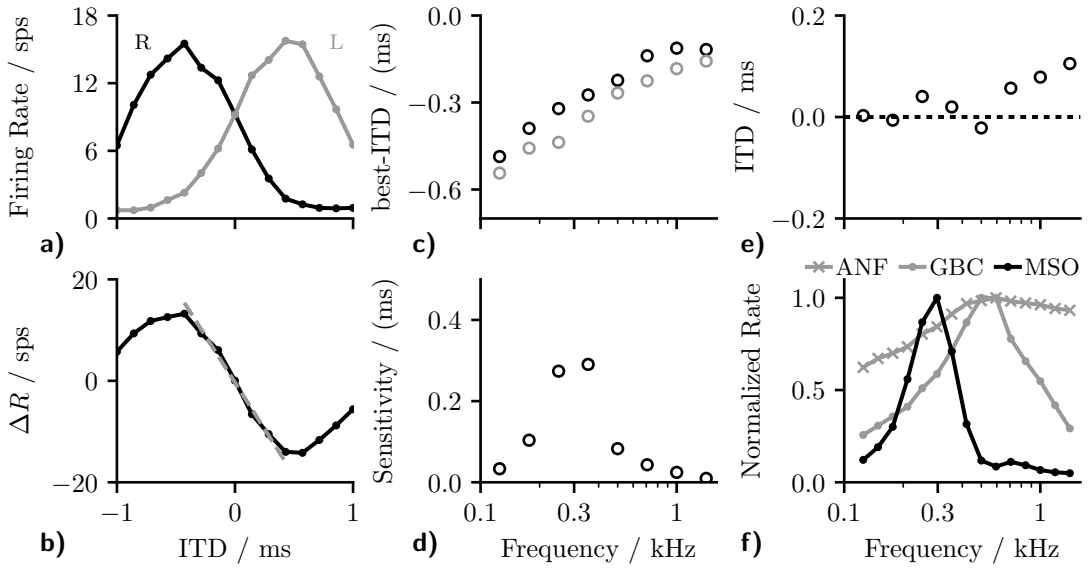

Figure 3.3: a) The model demonstrates ITD-rate functions for the left and right MSO that are effectively mirrored around zero so that a shift in ITD from the center line leads to an inverse response of the two channels forming the basis for the opponent-channel code (Stecker et al., 2005). **b)** A simple difference computation between the ITD-rate functions of the two hemispheres resulted in a nearly linear relationship around zero ITD (dashed line) **c)** The shift in the best-ITD decreases with increasing best-frequency of the MSO neuron, reaching the predefined shift of 100 µs at about 1.4 kHz. **d)** The slope of the rate difference curve around midline can be seen as a sensitivity to changes in ITD. This value changes with the best frequency. The sensitivity peaks at 300 Hz from where it decreases again towards higher frequencies. **e)** The single hemisphere responses showed the largest change at or close to the midline and therefore maximizes the sensitivity of the linear-decoder. **f)** Frequency-dependent normalized firing rates of the three neuron populations in the model in response to a 100 ms pure tone at 50 dB. (Graphic adapted from Encke and Hemmert (2018))

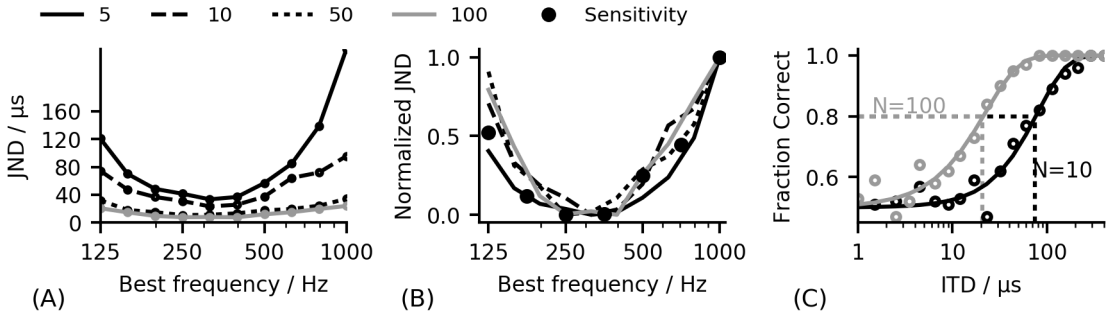

Figure 3.4: a) Frequency-dependent JNDs calculated for subsets of five, ten, fifty and one hundred randomly chosen neurons among a population of 500 neurons. **b)** When normalized to a region between 0 and 1, all JND curves overlap. **c)** The JND values were calculated by fitting weibull-functions to the fraction correct values (see Sec. 3.4).

best-ITD decreased faster than the width of the ITD-tuning function, which shifted the location of the largest slope towards positive ITDs. A second influencing factor on the sensitivity is the maximum firing rate of the MSO response – a higher rate of the single hemisphere responses will also result in a larger slope at midline. Figure 3.3f shows the frequency dependent normalized firing rates of all three neuron populations in the model. The firing rate of the MSO model is of course strongly influenced by the balance between the excitatory inputs from the ANFs and the inhibitory inputs from the GBCs but it is additionally modulated by changes of the spiking thresholds. MSO neurons have been found to exhibit subthreshold resonance (Remme et al., 2014; Mikiel-Hunter et al., 2016) which introduces frequency dependent thresholds. The MSO model used in this study exhibited a resonance frequency at about 260 Hz (see Fig. S2 in supplemental material) which is in agreement with the resonance frequencies found in electrophysiological studies (Remme et al., 2014; Mikiel-Hunter et al., 2016). The reduced spiking threshold around 260 Hz in combination with the dynamics of the synaptic inputs results in a peak in MSO response seen in Fig. 3.3f, which also corresponds to the peak in sensitivity shown in Figure 3.3d.

While applying the linear-decoder does not directly result in an ITD estimate, it can be used to predict ITDs. The link between ITD and ΔR also allows for a direct comparison of the laterality of two signals with different ITDs without the necessity to map the MSO model response to the absolute ITD estimates. This highlights the difference between an absolute localization task, which requires the mapping of the auditory perception to a spatial measure and a relative comparison task where the relative location of one perception in comparison to a second perception is reported. In psychoacoustical experiments, the sensitivity to ITDs is often assessed by determining the JND which describe the smallest change in ITD a subject can use to detect a change in lateralization between the two otherwise identical stimuli (Klumpp, 1956). Using the same method, we calculated JNDs for our network model using the linear-decoder

(see Sec. 3.4). In our model, the performance depends critically on the number of neurons composing the population, as the intrinsic stochasticity of the neuronal system loses its impact on the average firing rate when the population increases. To determine the influence of the population size on the performance of our model, JNDs were calculated separately for subsets of five, ten, fifty and one hundred randomly chosen neurons among a population of 500 neurons. Figure 3.4c shows exemplary psychometric curves derived for a population of ten and one hundred neurons. Figure 3.4a shows the result of the JND experiment for different pure tone stimuli. As expected, the predicted JND decreases when increasing the size of the population. The decrease in JND can be described by a $1/\sqrt{N}$ dependency, where N is the population size. The dependence is in line with the reduced effect of noise due to a larger population of neurons. If the JND thresholds are determined mainly by the noise of the system, they should also be reflected in the sensitivity described by the slope of the ΔR function. Figure 3.4b shows the JND curve as well as the inverse of the slope of ΔR with all values normalized to lie between 0 and 1. As expected, there is a good agreement between the normalized JND curves and the inverse of the slope, which confirms the aforementioned assumption that the detection threshold of the linear-decoder depends mainly on the slope of the rate-difference function around zero ITD.

One problem of such a linear-decoder is that the firing rates of the two MSO models depends not solely on the ITD, but also on other characteristics of the inputs to the MSO model. As the firing rate of the peripheral hearing model depend strongly on the sound pressure level, so will the output of the MSO model. To demonstrate such dependency, Fig. 3.5c shows how the predicted sensitivity of our model varies with both frequency and level of the pure tone input. The ANFs also exhibit strong spike-rate adaptation (Smith, 1977) which, consequently affects the MSO response (Fig. 3.5A). These variations could be compensated by normalizing the ITD-ΔRfunctions (overlay in Fig. 3.5b) but this is not possible in practice as it would require *a priori* knowledge about the maximum firing rate of the ITD-ΔRfunction at each point in time. A much more practical approach is to compensate such non-linear dependencies using the information that is already encoded in the ANF firing rates.

The MSO exhibits a distinct tonotopic organization along its dorsoventral axis. As the neuronal populations along the axis differ in their characteristic frequency (Guinan et al., 1972), consequently, a given ITD decoder can specialize on decoding of ITDs within a specific frequency range. In addition, the non-linear and time-dependent output of the peripheral hearing process can be compensated by using direct knowledge about the firing rates of the ANF. However, implementing such corrections would require designing a complex multi-dimensional correction function. artificial neuronal network (ANN) have been proven to be quite successful in learning the behavior of highly nonlinear systems (Almeida, 2002), hence, they provide an appealing alternative to tedious manual construction of a correction function.

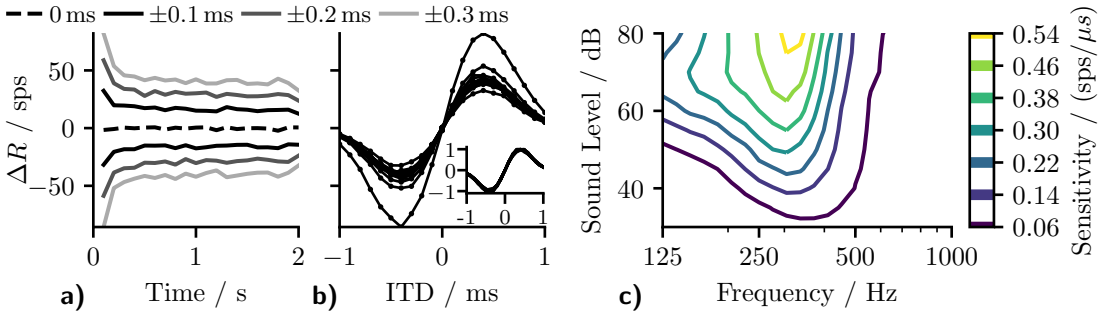

Figure 3.5: a) The ΔR function (calculated for 100 ms bins) of 500 MSO neurons in response to a two second long 250 Hz pure tone with ITDs ranging from 0 ms to ±0.3 ms show a strong influence of the peripheral hearing model's adaptation on the MSO model output. **b)** The same data as in a) but shown in form of ITD-ΔRfunctions. Every function corresponds to one point in time evaluated for 20 ITDs in the range from −1 ms to 1 ms. The difference in the slope of these functions illustrates that adaptation influences the sensitivity of the linear-decoder. Normalization with respect to the maximal rate of each function could compensate for this influence (overlay) **c)** The model displays strong variations in sensitivity with sound frequency and sound level. (Graphic adapted from Encke and Hemmert (2018))

3.2.3 Artificial neuronal network predictor.

We used a small multi-layer perceptron (MLP) to predict ITD values from the output of the SNN model by means of non-linear regression analysis. The regression is based on the average firing rates across the neuronal populations and predictions are calculated separately for each frequency band and time frame. The MLP was implemented using seven input nodes, one hidden layer with twenty nodes and two output nodes (for details see Sec. 3.4). One of the MLP output nodes was used for the prediction task, while the second output was used to classify the reliability of the prediction based on the firing rates. This was deemed necessary to omit predictions for parts of the input signal, which did not contain enough energy in the given frequency band to enable robust predictions based on sufficient spiking activity.

The inputs to the MLP were designed to consist of the firing rates from the MSO of the left and right hemisphere and the characteristic frequency of the neuron population (see Fig. 3.6 for a schematic of the networks in- and outputs). As one of the main tasks of the predictor was to compensate for the influence of variations in the peripheral hearing model output, the MLP was also provided with a monolateral input of the ANF firing rate. All firing rates were provided as an average value computed over a predefined time period of 30 ms. This duration was chosen as it offered reasonably high temporal resolution and ensured that several periods of the phase locked input were included. In addition to the rates within the given time frame, we also provided firing rates of the previous time frame which reduced the noise in the predictions by

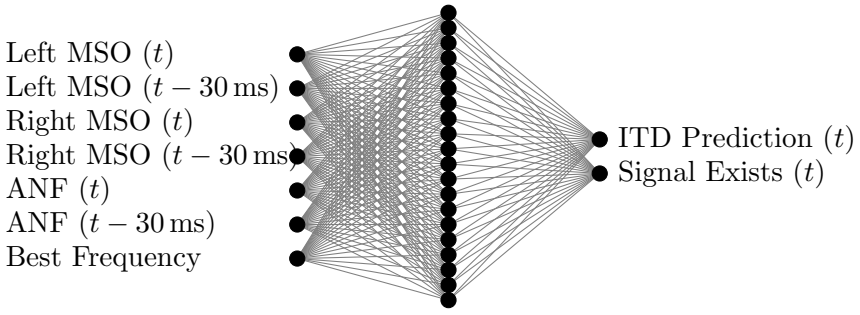

Figure 3.6: The ANN predictor was implemented using seven input nodes, one hidden layer with twenty nodes and two output nodes. Inputs denoted with (t) are firing rates within a given time period (typically 30 ms) for which the ITD should be predicted, while the ones denoted with (t-30 ms) are firing rates of the preceding time frame. (Graphic adapted from Encke and Hemmert (2018))

effectively doubling of the time span that the network can employ in its predictions. The MLP was trained on 300 ms long pure tones (see Sec. 3.4) covering the frequency range from 125 Hz to 1000 Hz so that predictions can be obtained for any stimuli within that range. For the following experiments we calculated predictions for 13 logarithmically spaced frequencies between 125 Hz to 1000 Hz.

Figures 3.7a - 3.7c compares the results of the ANN-predictor with those of the linear-decoder for an amplitude-modulated tone with 400 Hz carrier frequency and a modulation rate of 2 Hz. Since amplitude modulation is encoded in the firing rate of the ANF, it is also exhibited in the output of the linear-decoder (Fig. 3.7a and 3.7c). On the other hand, the predictions from the ANN (Fig. 3.7b) showed only minor deviations at the on- and offsets of each modulation cycle while largely compensating the strong onset response introduced by ANF's adaptation. Figure 3.7b shows only such predictions that the ANN classified to be reliable. In case of the amplitude-modulated signal, the frequency bands for which the ANN could predict ITDs are dependent on the phase of the modulation.

Omitting unreliable predictions enables the calculation of a general prediction across frequency bands. In case of the linear-decoder, zero output can correspond to two conditions – zero ITD and no signal. The employed method of omitting unreliable estimates is especially important for applying the ANN predictor to more complex signals that have several frequency components because the omission enables the ANN to predict ITDs without prior knowledge about the signal's frequency content. Figures 3.7d - 3.7f show examples of the ANN-predictor applied to a linear chirp. To demonstrate the ability of the predictor to follow changes both in frequency as well as in ITD, an additional phase shift was applied to the left ear signal. This phase shift was chosen to be proportional to an ITD-value that varied linearly from 300 µs to −300 µs. By calculating a cross-frequency prediction for every time frame, the

ANN-predictor was able to follow the change in frequency as well in ITD (Fig. 3.7c) deviating from the true value only in the last two time frames. As a final example, we show the ANN-predictor applied to a speech signal with a static ITD of 200 µs (Figs. 3.7d - 3.7f). Again, the across-frequency estimation, combined with the omission of unreliable predictions allows the ANN-predictor to offer an accurate estimate of the ITD for the whole signal.

3.3 Discussion and Conclusion

In this study, we presented a novel binaural model and used it to detect ITDs in arbitrary sound signals. In contrast to previous binaural models that used a phenomenological modeling approach (Pulkki and Hirvonen, 2009; Dietz et al., 2011), this study used biophysical neuron models based on the current knowledge about the function of the mammalian medial superior olive. Some previous studies implemented similar SNN but either used a simplified auditory periphery and thus limited the application of the model to pure tones (Brughera et al., 2013), based their model on topologies that disagree with newer physiological studies (Glackin et al., 2010) or focused on ITDs in the stimulus envelope (Wang et al., 2013). Using two different extraction methods, we found that applying the opponent-coding mechanism to the output of the model enabled a robust extraction of ITDs even in complex signals.

3.3.1 Sensitivity of the linear-decoder

We have shown that a simple linear-decoder can detect ITDs from the outputs of the left and right MSO models with a sensitivity that reflects human performance in a discrimination task and depended on sound frequency. The sensitivity was mainly determined by the maximum firing rate of the MSO at a given frequency. Our MSO neurons showed a peak at approximately 300 Hz. To our knowledge, no such systematic variation of firing rate with frequency has been described, but Yin and Chan (1990) noted a similar characteristic in the response of high-frequency MSO neurons. They recorded the response of neurons that phase-locked to the envelopes of amplitude-modulated tones and also showed a peak in the response at a modulation rate of 300 Hz. In the model, the responses were influenced by the subthreshold resonance of MSO neurons, which is due to the dynamics of the low threshold potassium current (Mikiel-Hunter et al., 2016). This resonance would also explain the results by Yin and Chan (1990). An explanation why no similar result in the response of low-frequency MSO neurons has been described is that these measurements are limited to responses derived at the neurons' best-frequency so that any systematic variation between neurons with different best-frequencies could be masked by variations in the overall response rate between neurons.

It should be noted that the sensitivity of the linear-decoder cannot be directly compared to results from psychoacoustical experiments, as the model only accounts for

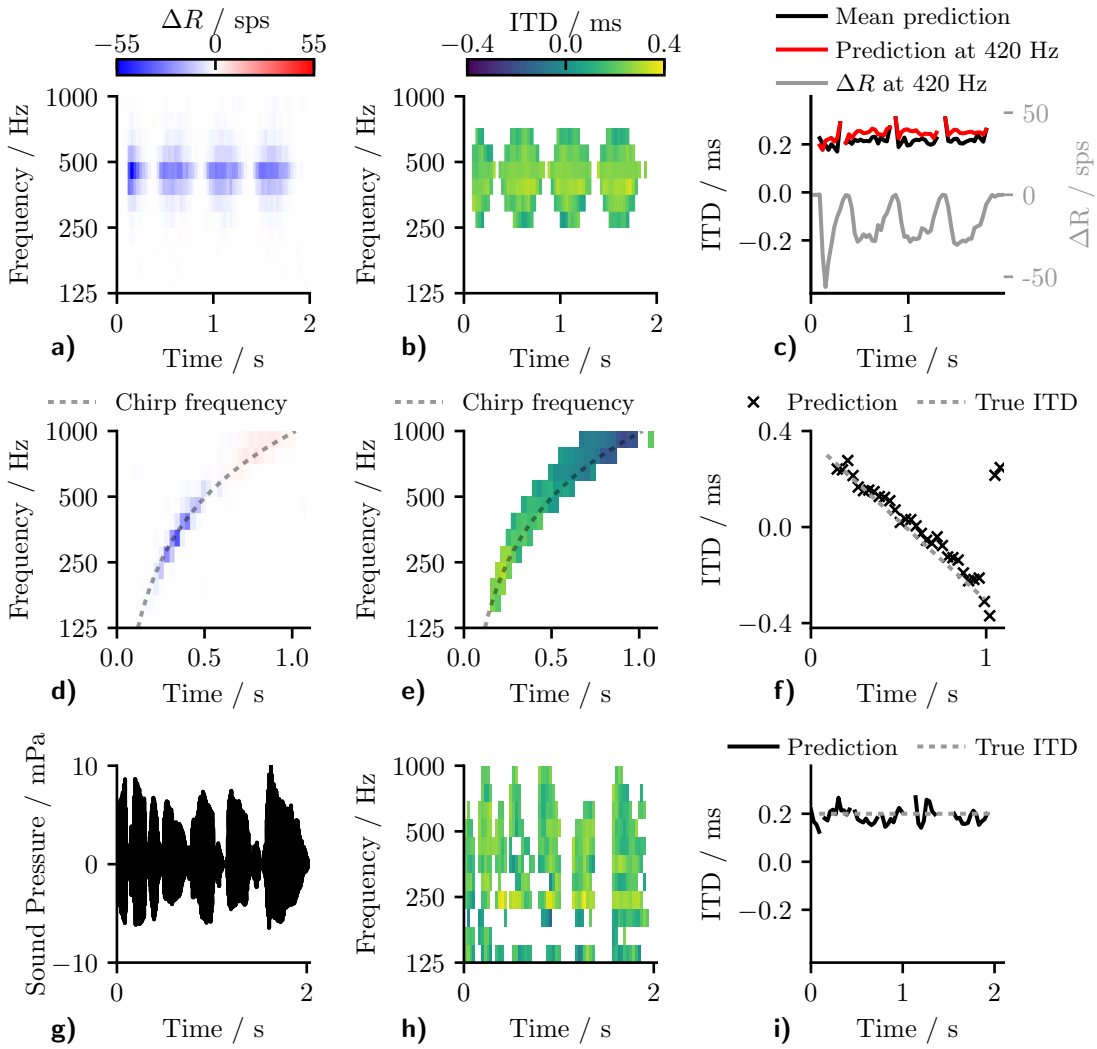

Figure 3.7: a) Results of the linear-decoder for an amplitude-modulated tone with 400 Hz carrier frequency and a modulation rate of 2 Hz presented with an ITD of 200 ms. ΔR showed strong modulation with the modulation frequency of the sound as well as an influence of ANF adaptation **b)** Results of the ANN-predictor predictor for the same signal as in a). The ANN was able to correct for the variations conveyed by the ANF inputs and to provide a stable prediction within the frequency bands from 250 Hz to 595 Hz. **c)** The output of both, the ANN-predictor and the linear decoder for the amplitude modulated signal over time. Red: ANN predictions for the 420 Hz channel which was the closest to the stimulation frequency. Black: Mean over all predictions that were classified to contain a useful signal. Gray: Result of the linear-decoder in the 420 Hz frequency band. **d), e)** Same plots as in a) and b) but for a linear, one-second long chirp ranging from 125 Hz to 1 kHz where the ITD changed from −0.4 ms to 0.4 ms. **f)** The ANN-predictor was able to follow the change in frequency as well as in ITD, deviating from the true value only at the end of the signal. **g) - i)** The ANN-predictor applied to a speech signal (German sentence "Britta gewann drei schwere Steine") taken from the OLSA sentence test (Wagener et al., 1999). (Graphic adapted from Encke and Hemmert (2018))

the lowest stages of the neuronal ITD-detection circuit in gerbils. In other words, it was not the goal of this study to replicate any psychophysical data *per se*, but rather to investigate the performance of the model on its own. Nevertheless, the model could be easily tuned to replicate psychoacoustic threshold data by adjusting the size of the neuronal population to fit human or animal data.

3.3.2 Influence of missing SBCs on the output of the model.

In the presented model network, MSO neurons received direct excitatory input from ANFs, while in the physiological case, they receive excitatory inputs from SBCs. SBCs have been found to increase the precision of phase-locking in comparison to ANFs (Dehmel et al., 2010; Künzel et al., 2011). The improvement shown in this study is rather small when compared to the large improvement that has been shown for GBCs (Joris et al., 1994). In spite of this Improvement, the precision is not much higher than that of the ANF model used in this study, and thus no further improvement in phase locking seemed necessary. A second function of SBCs could arise from non-monotonic rate-level functions due to an inhibitory sideband (Künzel et al., 2011; Keine and Rübsamen, 2015). Including a model that would reproduce the non-monotonic rate-level functions may also change the output of the MSO model, specifically, the behavior shown in Fig. 3.5c. This change in the MSO rate-level function may also be compensated by the ANN, so that the additional feature would not change the message of this paper, leading to the decision to neglect the influence of SBCs. It was also suggested that the slow GABA-ergic inhibition on the level of the SBC may support sound localization of complex sounds by acting as a gain control mechanism (Keine et al., 2016, 2017), this would be interesting to investigate in the context of the presented model but is outside of the scope of this paper.

3.3.3 Performance of the ANN-predictor

The model output showed a strong dependence on both frequency and level of the input signals. Previous models that employed the opponent-coding principle constructed the output of their models to be self normalizing (Takanen et al., 2014; Pulkki and Hirvonen, 2009) or directly extracted the phase from the left and the right input signals using gammatone filters (Dietz et al., 2011). While both methods are valid in view of a phenomenological modeling approach, they can not be easily applied to a neuronal network as presented in this study. We instead showed that a multilayer perceptron could be trained to compensate for frequency and level dependencies and to predict ITD values from the firing rate outputs of the spiking neuron network. By using an ANN to compensate for variability of the MSO output, this study neither makes any assumption about the exact location in the ascending auditory pathway, at which this compensation takes place, nor speculates about the exact mechanism underlying this compensation. We rather show that a very basic ANN containing only twenty hidden

nodes in one layer is able to perform the compensation. The ANN-predictor was also shown to provide accurate ITD predictions for complex signals and for time-variant ITDs, even though it was trained on pure tones only. This suggests that the necessary compensation is independent of context. Psychoacoustic studies have shown that sound localization performance depends on the duration (Tobias, 1959) and bandwidth (Trahiotis and Stern, 1989) of the stimulus indicating an integration of information across frequency and time. In this study, the ANN predicted ITDs independently for each frequency and time frame. While integration over the frequency bands was implemented by calculating the mean prediction across all frequencies, no integration over time apart from the calculation of 30 ms averages was performed. Hence, the prediction capability is expected to further improve if the output of the model would also be integrated over time.

While the goal of this study was to evaluate the models' performance on the detection of ITDs, the prime interest of our binaural hearing lies in estimating the direction of a sound source instead of the ITD value. Since low-frequency ITDs between the ear canal signals provide a salient cue about sound source direction, reliable prediction of the ITDs indicates that the azimuthal sound direction may also be accurately predicted. To that end, the ANN could also be trained to directly predict azimuthal angles instead of ITDs.

3.4 Methods

3.4.1 Topology of the Model

Both MSO and GBC neurons were modeled using single-compartment, Hodgkin-Huxley-type models simulated in python using the package Brian (Goodman, 2009). MSO as well as GBCs received direct excitatory input from ANF fibers, which were modeled using the model of Zilany et al. (2014), implemented in the python library *cochlea* (Rudnicki et al., 2015). Each population of neurons (ANF, GBC, MSO) always consisted of 500 independent neurons in each hemisphere. The frequency channel of the neuron population was set by selecting the appropriate critical frequency of the peripheral hearing model.

3.4.2 Spiking models

While this study does not discuss the effect of single ionic currents, it makes use of Hodgkin-Huxley-type models, as simpler neuron models like the leaky integrate-and-fire neurons neglect the influence of ion channel dynamics. Especially the shift of best-ITD towards contralateral-leading ITDs has been shown to be influenced by both low-threshold potassium (Myoga et al., 2014) and hyperpolarizing ionic currents (Baumann et al., 2013), both of which are included in this model.

MSO neurons were simulated using single-compartment, Hodgkin-Huxley-type models. The dynamic of their membrane potential V_m is given by the following equation:

$$\frac{\mathrm{d}V_\mathrm{m}}{\mathrm{d}t} = -\frac{1}{C_\mathrm{m}}(I_\mathrm{leak} + I_\mathrm{Na} + I_\mathrm{K} + I_\mathrm{h} + I_\mathrm{syn,e} + I_\mathrm{syn,i}), \tag{3.1}$$

where C_m is the membrane capacitance, I_{leak} is the leakage current, I_Na, I_K, I_h are the sodium, potassium and hyperpolarizing ionic currents and $I_\mathrm{syn,e}$, $I_\mathrm{syn,i}$ are the excitatory and inhibitory synaptic currents respectively. All ionic currents were defined as follows:

$$I_x = \hat{g}_x a^m b^n (V_m - E_x), \tag{3.2}$$

where $\hat{g}_x$ and E_x are the maximal conductivity and Nernst potential for the respective ion species x. The gating variables a^m and b^n determine the channel kinetics. Equations for these variables can be found in the original publication: The sodium dynamics were implemented according to Rothman and Manis (2003) and were corrected for a body temperature of 37 °C ($k = 3^{(T-22)/10}$). To gain realistic spike shapes as well as a spiking threshold, the activation kinetics had to be sped up by a factor of four. Potassium currents were modeled with the equations for the low threshold channels given by Khurana et al. (2011) with the steady-state inactivation z_∞ set to 0.4. The hyperpolarizing currents were modeled using the equations for dorsal MSO neurons from Baumann et al. (2013). We used a membrane capacity of 70 pF (Couchman et al., 2010) and the ionic conductivities were adjusted to fit the steady state and peak membrane resistances to values measured by Scott et al. (2005). Use of these values resulted in spiking thresholds close to the data published by Couchman et al. (2010). All parameters are summarized in Table 3.1. GBCs and their synaptic inputs were modeled using the neuron model with 40 non-depressing ANF inputs as proposed by Rudniki and Hemmert (2017).

Symbol	Value	Symbol	Value
C_m	70 pF	E_i	−70 mV
E_{rest}	−55.8 mV	$\hat{g}_{Na}$	3.9 µS
E_{Na}	56.2 mV	$\hat{g}_K$	650 nS
E_K	−90 mV	$\hat{g}_h$	520 nS
E_h	−35 mV	$\hat{g}_{leak}$	13 nS
E_e	0 mV		

Table 3.1: Parameters for the MSO model

3.4.3 Synaptic MSO inputs

Each MSO neuron received six excitatory inputs from ANFs of each hemisphere. The excitatory post-synaptic currents (EPSCs) were modeled as an alpha function:

$$I_{syn,e} = \frac{t \cdot e^{1-t/\tau}}{\tau_e}(V_m - E_e). \tag{3.3}$$

Inhibition was provided via three GBC inputs per hemisphere. The inhibitory post-synaptic currents (IPSCs) were modeled using a bi-exponential function:

$$g_i = \hat{g}_i \frac{\tau_2 \cdot (e^{-t/\tau_{i,1}} - e^{-t/\tau_{i,2}})}{\tau_{i,2} - \tau_{i,1}} \cdot (V_m - E_i). \tag{3.4}$$

Both, excitatory and inhibitory timeconstants were fitted to recordings by Couchman et al. (2010) yielding values of $\tau_e = 0.17\,\text{ms}$ and $\tau_{i,1} = 0.14\,\text{ms}$, $\tau_{i,2} = 1.6\,\text{ms}$.

3.4.4 Sound Signals and Data Analysis

All sound signals were generated in Python at a sampling rate of 100 kHz as this sampling rate is required by the peripheral hearing model (Zilany et al., 2014). In the case of the speech signal, the sound was up-sampled from 44.2 kHz to 100 kHz. Each sound signal was gated using a 20 ms long raised-cosine function and 20 ms of silence was attached to the beginning and the end of the signal. The stimuli were presented at a sound pressure level of 50 dB if not stated otherwise. ITDs were defined as the difference in the arrival times between the left and the right ears, with positive values corresponding to right leading sounds. To archive sub-sample ITDs, we generated the corresponding delays between the two signals by applying a fast Fourier-transform (FFT), adding the equivalent phase angles, which resulted from the delays, and reverse FFT back to time domain signal.

ITD-rate functions were fitted using a modified Gaussian function as shown in (3.5) were Δt is the ITD value, $R_{\max}$ the maximum firing rate, W defines the width of the curve and B the location of the maximum (best-ITD).

$$R(\tau) = R_{\max} \cdot \mathrm{e}^{\frac{-(\Delta t - B)^2}{W^2}} + R_{\text{offset}} \tag{3.5}$$

Spiking data were analyzed using the Thorns toolbox for python. Firing rates were always given as the average response of the whole population. To compensate for the intracochlear delay of the inner ear model, we only considered action potentials arriving 25 ms after signal onset and up to 25 ms after the end of the signal.

3.4.5 Calculation of Just Noticeable Differences

JNDs for our model were calculated by presenting two stimuli with ITDs located symmetrically around zero – i.e. $-\Delta t/2$ and $\Delta t/2$. The difference between the two

ITD was denoted ΔITD . We calculated independently, the difference in firing rate at both hemispheres (ΔR) for each of the presented signals. The two values were then compared to each other. If the ΔR value for the negative ITD signal was larger than the one for the positive ITD signal, the trial was considered as a correct prediction. Each ΔITD was presented 100 times and the fraction of correct trials was calculated. To calculate the JND, we presented 20 logarithmic arranged ΔITD in the range from 2 µs to 800 µs. The resulting fraction correct values were then fitted with a weibull function. The JND was defined as the ITD at which 75% correct predictions were achieved.

3.4.6 The Artificial Neural Network Predictor

The ANN network was implemented using the Theano package for Python. The ANN layout was that of a classic multilayer perceptron containing an input layer with seven nodes, one hidden layers with twenty nodes and an output layer with two nodes (see Fig. 3.6). Both the hidden and the output layer consisted of non-linear nodes with a $\tanh(x)$ activation function.

The predictor was designed to make predictions for every 30 ms section of the signal. For this, average firing rates for both MSO hemispheres and for the ANF of one hemisphere were calculated in bins of 30 ms. The model firing rates of MSO and ANF as well as the best frequency of these neurons were given as the ANN inputs. To provide some history which can be used to compensate for on- and offsets, the predictor was also provided with the firing rate in the previous 30 ms bin. Using this information, the ANN gave a prediction of the ITD value in the current bin and a classification whether the presented bin actually contained a signal (signal exists).

The network was trained on the MSO model output from 2000 different 300 ms long sine tones which were padded by 60 ms of quiet. For each tone the level, frequency as well as ITD were randomly chosen to lie between 30 dB to 70 dB, 125 Hz to 1000 Hz and ±500 µs respectively. The target data for the training set consisted of the ITD value of the corresponding input signal, as well as the classification whether the current time frame contained a signal or not. The target for the classification was set to -1 for the two time bins at the start and at the end of each signal as those contained silence. It was set to 1 for all other bins. The set of training signals was then split into three subsets, a training set containing 80% of the data, a validation and test set both containing 10% of the signals. The ANN was trained on the training set, until the improvement on the mean squared error function for the validation set stayed consistently below 0.01%.

CHAPTER 4

A Two-Channel Model of the Neuronal Representation of IPDs

4.1 Introduction

In order to localize sound sources in the horizontal plane, humans and other mammals make use of interaural time differences (ITDs) and interaural level differences (ILDs). Traditionally, the neuronal processing of ITD cues conveyed by the fine structure of a sound, in the following referred to as interaural phase differences (IPDs), has been explained using the delay-line model of Jeffress (1948), where a cascade of neurons, each tuned to a different specific time difference, encodes the IPD. This neuronal structure was also implemented in several models, which were able to explain a large variety of binaural phenomena (e. g., Lindemann, 1986; Colburn, 1977; Stern and Shear, 1996).

In the last two decades, the application of the neuronal delay-line code to mammals has been challenged by electrophysiological measurements in several species (Yin and Kuwada, 1983; Brand et al., 2002; McAlpine et al., 2001). Instead of a cascade of neurons narrowly tuned to IPDs within the physiologically-occurring range, most studies found rather broadly tuned neurons, several or sometimes most of which tuned to IPDs outside the physiologically-occurring range, and only a few neurons tuned to IPDs around midline (McAlpine et al., 2001). Based on a recent review by Grothe et al. (2010), these findings are more consistent with a neuronal representation where the IPD is encoded by the relative activity in the two brain hemispheres, as it was first proposed by Von Békésy and Wever (1960). This hemisphere code is also supported by psychoacoustic findings (Phillips and Hall, 2005; Phillips et al., 2006) and functional imaging studies in humans (Salminen et al., 2010; Magezi and Krumbholz, 2010; Salminen et al., 2018).

While there is ample physiological evidence for a hemisphere code, only a few models successfully applied the hemisphere code to explain the outcome of psychoacoustic experiments (e.g. Takanen et al., 2014). Dietz et al. (2009) presented a binaural model that employed the hemisphere code in combination with a Jeffress style place-code to explain several psychoacoustic experiments and noted that some of the experi-

ments could not be explained when relying solely on the hemisphere code. Using biophysical models Encke and Hemmert (2018) and Brughera et al. (2013) showed that the hemisphere code is sufficient to explain ITD-difference thresholds in the same magnitude as observed in humans but both studies could not account for the observed frequency-dependence of ITD-difference thresholds. Hancock (2004) on the other hand presented a model consistent with the mammalian physiology but relied on a neuronal representation that combined the hemisphere and place-code approach.

This chapter develops an empirical model of the neuronal representation of IPDs assuming a hemisphere code. The model is constrained solely by biophysical data (section 4.2) resulting in a condensed formulation of the complex physiological processes. This formulation allows a direct investigation of implications of the neuronal code on perception with only a fraction of the parameters necessary for more complex neurophysiological or empirical models. The output of the formulated model of the neuronal code predicts to what extent a positive IPD can be distinguished from negative IPDs. As this prediction is solely based on the neuronal code, which also restricts perception in that only the information conveyed by this code is accessible for forming hearing sensations, any limitation of the code is expected to be reflected in perception. In order to evaluate the model predictions and this hypothesis, sections 4.3 and 4.4 investigate the human IPD-lateralization performance in two psychoacoustic experiments, which are compared to model predictions in section 4.5.

4.2 The Model

An ITD is defined as the delay Δt between the signals at the two eardrums (Blauert, 1999, pp. 140f). For a pure-tone impulse of frequency f, the effect of an ITD can be decomposed into a time shift of the stimulus envelope and a phase shift of one ear signal with respect to the other. For low-frequency sounds, the latter was reported to provide, compared to envelope cues, the dominant cue for lateralization (Bernstein and Trahiotis, 1985; Smith et al., 2002). In order to separate effects of envelope and phase shifts, a pure-tone IPD

$$\Delta\varphi = \varphi_{0,r} - \varphi_{0,l} = 2\pi f \Delta t \tag{4.1}$$

is defined here as the difference between the starting phases φ_0 of the tones that constitute the right (r) and left (l) ear signals. This nomenclature shall emphasize that introducing an IPD, as used here, is meant to only affect the phase of the ear signals, while keeping their envelopes interaurally identical. For pure tones, IPDs are limited to the range $\Delta\varphi \in [-\pi, \pi]$, and the ear signals are identical for IPDs of $-\pi$ and π. As a consequence, neuronal and psychoacoustical results must be comparable for pure-tone stimuli with IPDs of $\pm\pi$.

In order to distinguish between an anatomical brain hemisphere and its model representation, each brain hemisphere will be referred to as a channel of the model.

The proposed model was largely based on the repeated finding that the average neuronal responses of ITD-processing nuclei at several stages of the auditory system are described well by Gaussian functions (Stange et al., 2013; Yin and Kuwada, 1983; Belliveau et al., 2014). Therefore, no specific location in the neuronal auditory pathway at which the hemisphere code is implemented was assumed when defining the average channel response as

$$\bar{R}_{\mathrm{C}}(\Delta\varphi, \Delta\varphi_{\mathrm{best}}, w_{\mathrm{C}}) = \sum_{i=-\infty}^{\infty} e^{-\frac{(\Delta\varphi-(2\pi i+\Delta\varphi_{\mathrm{best}}))^2}{2{w_{\mathrm{C}}}^2}} \qquad i \in \mathbb{N}_0, \tag{4.2}$$

where $\Delta\varphi_{\mathrm{best}}$ represents the IPD with maximum neuronal response (best-IPD) and w_{C} determines the width of the neuronal response function. Thereby, in order to implement the 2π periodicity of $\Delta\varphi$, each neuronal response is defined by a sum of Gaussian functions centered at $2\pi i + \Delta\varphi_{\mathrm{best}}$, so that $\bar{R}_{\mathrm{C}}(\Delta\varphi) = \bar{R}_{\mathrm{C}}(\Delta\varphi + 2\pi i)$ for all $i \in \mathbb{N}_0$. For applications where $\Delta\varphi \in [-\pi, \pi]$, it is sufficient to confine i to $[-1, 1]$, as the influence of larger i becomes negligible. Equation (4.2) gives a normalized, unit-less value, the modeled average neuronal channel response. While several other equations, like a rectified cosine or interaural correlation (Hancock, 2004) could be used to model the same relationship, Eq. 4.2 was used because its parameters can be easily interpreted.

Psychoacoustic experiments suggest that the sensitivity to ITD differences is highest around $\Delta t = 0$ (Hafter and De Maio, 1975; Domnitz and Colburn, 1977). As the slope of $\bar{R}_{\mathrm{C}}$ over IPD may be considered to model a sensitivity towards IPD differences, it appears reasonable to assume the maximum slope of $\bar{R}_{\mathrm{C}}$ at $\Delta\varphi = 0$. This assumption is also supported by neuronal recordings (McAlpine et al., 2001; Stecker et al., 2005), which found the maximum slope of single-neuron responses at or close to $\Delta\varphi = 0$. It follows from Eq. (4.2) that the maximum slope at $\Delta\varphi = 0$ is reached if $w_{\mathrm{C}} = |\Delta\varphi_{\mathrm{best}}|$.

Electrophysiological studies have shown that neurons in the left hemisphere normally respond strongest to positive IPDs and vice versa (McAlpine et al., 2001; Joris et al., 2006). Following these findings and further assuming the responses of the channels to be symmetric with respect to $\Delta\varphi = 0$, Eq. (4.2) can be specified to the IPD-dependent mean channel responses

$$\bar{R}_{\mathrm{C,l}}(\Delta\varphi) = \bar{R}_{\mathrm{C}}(\Delta\varphi, \Delta\varphi_{\mathrm{best}}) \tag{4.3}$$

of the left channel and

$$\bar{R}_{\mathrm{C,r}}(\Delta\varphi) = \bar{R}_{\mathrm{C}}(\Delta\varphi, -\Delta\varphi_{\mathrm{best}}) \tag{4.4}$$

of the right channel. Figure 4.1a shows those channel responses as functions of $\Delta\varphi$ for the parameter $\Delta\varphi_{\mathrm{best}} = 0.3\pi$.

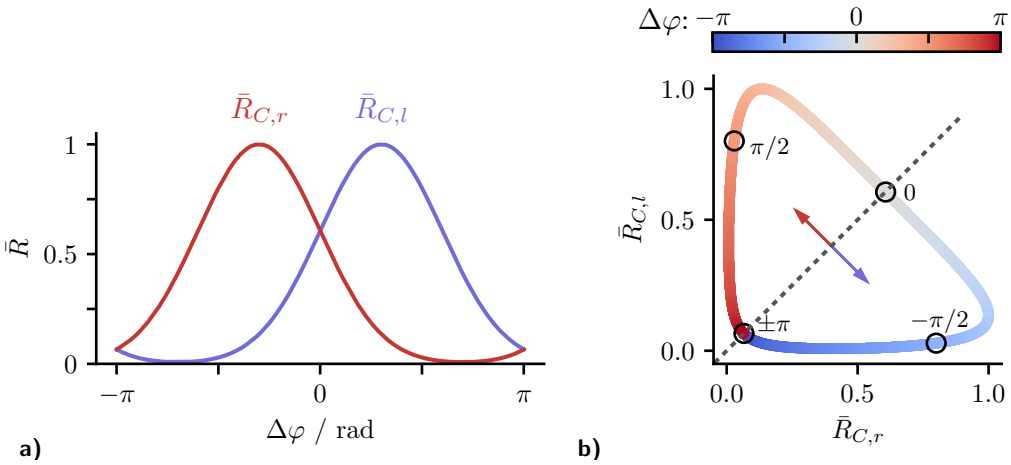

Figure 4.1: Modeled average neuronal responses $\bar{R}_{\mathrm{C,l}}\,(\Delta\varphi)$ and $\bar{R}_{\mathrm{C,r}}\,(\Delta\varphi)$ of the left and right model channels (hemispheres) according to Eq. (4.3) and Eq. (4.4), for the parameter $\Delta\varphi_{\mathrm{best}} = 0.3\pi$. **a)** Responses as functions of the IPD $\Delta\varphi$ for the left (blue) and right (red) channels. **b)** IPD-response plane: For all IPDs (color coded), the respectively corresponding responses, combined here to the response vector $\bar{\mathbf{R}}(\Delta\varphi)$, describe a closed contour in the IPD-response plane ($\bar{R}_{\mathrm{C,l}}$-$\bar{R}_{\mathrm{C,r}}$ plane).

In the model, the responses of the two channels are combined to the response vector

$$\bar{\mathbf{R}}(\Delta\varphi) = \left(\bar{R}_{\mathrm{C,r}}(\Delta\varphi) \quad \bar{R}_{\mathrm{C,l}}(\Delta\varphi)\right)^T, \tag{4.5}$$

which defines the center of the bivariate normal distribution of observable responses

$$\mathbf{R}_{obs} \sim \mathcal{N}_2\left(\bar{\mathbf{R}}(\Delta\varphi),\ \begin{pmatrix} {\sigma_{\mathrm{C}}}^2 & 0 \\ 0 & {\sigma_{\mathrm{C}}}^2 \end{pmatrix}\right), \tag{4.6}$$

where $\mathcal{N}_2\,(\mu,\,\Sigma)$ denotes a bivariate normal distribution with expectation vector μ and covariance matrix Σ. ${\sigma_{\mathrm{C}}}^2$ is the variance of the responses in each channel.

With changing IPD, the average response vector $\bar{\mathbf{R}}$ describes a closed contour in the $\bar{R}_{\mathrm{C,l}}$-$\bar{R}_{\mathrm{C,r}}$ plane (IPD-response plane). Each IPD can be assigned a specific $\bar{\mathbf{R}}$ on this contour (cf., Figure 4.1b, which, following Eq. (4.6), defines a distribution of observable responses $\mathbf{R}_{obs}$. Any decoder of this neuronal code would have to solve the inverse problem: Finding the $\bar{\mathbf{R}}$, or IPD, which resulted in the observed $\mathbf{R}_{obs}$. The most simple mapping task between $\mathbf{R}_{obs}$ and IPD is a basic left-right task, which only attributes an observed $\mathbf{R}_{obs}$ to a positive or negative IPD. It becomes clear in Fig. 4.1b that all positive IPDs result in $\bar{\mathbf{R}}$ above the equal-response diagonal ($\bar{R}_{\mathrm{C,l}} = \bar{R}_{\mathrm{C,r}}$), while negative IPDs result in $\bar{\mathbf{R}}$ below the equal-response diagonal.

In the model, the probability for a given IPD to result in a $\mathbf{R}_{obs}$ above the diagonal is calculated from Eq. (4.6). The probability is determined by the shortest distance

$$d_{\mathrm{diag}}(\Delta\varphi) = \sqrt{\left[\bar{R}_{\mathrm{C,l}}(\Delta\varphi) - \bar{R}_{\mathrm{C,r}}(\Delta\varphi)\right]^2/2} \tag{4.7}$$

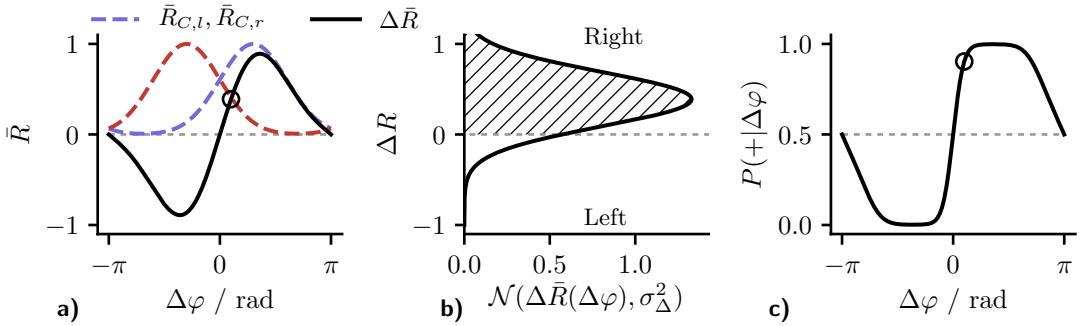

Figure 4.2: Overview of the proposed model. **a)** Modeled average neuronal channel responses $\bar{R}_{\mathrm{C,l}}$ and $\bar{R}_{\mathrm{C,r}}$ (dashed) and their difference ΔR (solid), with the parameter $\Delta\varphi_{\mathrm{best}} = 0.3\pi$. **b)** Probability distribution of the responses for the IPD indicated by the circle in a) with the parameter $\sigma_\Delta = 0.25\pi$. ΔR is normally distributed, and integrating the respective part of the distribution function results in the probability that a stimulus with a given IPD will result in a positive observed rate difference (hatched area). **c)** Probability of presenting the model with a given IPD to result in a positive observed rate difference. The hatched area in b) is the circled value.

of $\bar{\mathbf{R}}(\Delta\varphi)$ from the diagonal. Equation (4.7) only depends on the difference

$$\Delta R(\Delta\varphi) = \bar{R}_{\mathrm{C,l}}(\Delta\varphi) - \bar{R}_{\mathrm{C,r}}(\Delta\varphi) \tag{4.8}$$

between the average responses of the right and left channels. ΔR is normally distributed with the standard deviation $\sigma_\Delta = \sqrt{2{\sigma_{\mathrm{C}}}^2}$.

As a $\bar{\mathbf{R}}$ above the diagonal results in a positive rate difference and vice versa, it is possible to formulate the probability of a given IPD to result in a $\mathbf{R}_{obs}$ above the diagonal based on ΔR. This probability is

$$P(+|\Delta\varphi) = \int_0^\infty \mathcal{N}\left(\Delta R(\Delta\varphi),\, {\sigma_\Delta}^2\right) \mathrm{d}\Delta R, \tag{4.9}$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with expectation value μ and variance σ^2. According to Eq. (4.9), the probability $P(+|\Delta\varphi)$ for observing a positive rate difference is calculated by integrating the positive portion of the distribution centered at ΔR.

Two parameters are necessary for solving Eq. (4.9): The standard deviation σ_Δ and the best-IPD $\Delta\varphi_{\mathrm{best}}$, which determines the channel-response vector $\bar{\mathbf{R}}(\Delta\varphi)$. Electrophysiological studies in the brainstem of gerbils found $\Delta\varphi_{\mathrm{best}}$ between approximately 0.2π and 0.4π (Pecka et al., 2008).

Figure 4.2 gives an overview of the proposed model structure, assuming the realistic parameters $\Delta\varphi_{\mathrm{best}} = 0.3\pi$ and $\sigma_\Delta = 0.25\pi$. Figure 4.2a) shows, similar to Fig. 4.1, the basis of the model: The physiologically-motivated average channel responses (dashed)

and their difference (solid). Figure 4.2b depicts, as an example, the probability distribution for the IPD circled in Fig. 4.2a). Figure 4.2c shows the probability of the model to observe a positive rate difference given Eq. (4.9). The probability for the circled example case is given by the dashed area in Fig. 4.2b. The resulting IPD-dependent contour is symmetric with respect to (0, 0.5) and intersects the chance level of 0.5 twice, at $\Delta\varphi = 0$ (center IPD) and, due to periodicity, at $\Delta\varphi = \pm\pi$ (lateral IPD).

In order to be able to evaluate the model predictions in a psychoacoustic experiment, it is necessary to connect $P(+|\Delta\varphi)$, the probability that the response to a given $\Delta\varphi$ is attributed to a positive IPD, to a hearing sensation property. In free-field conditions, positive IPDs mostly result from sound sources located to the right of the median plane (Blauert, 1999; Völk et al., 2017b), and positive IPDs in headphone presentation tend to evoke hearing sensations to the right of the perceived median plane (Cherry and Sayers, 1956; Völk et al., 2017b). On that basis, this chapter hypothesizes that $P(+|\Delta\varphi)$ describes the probability that a given IPD results in a hearing sensation on the right of the perceived median plane. The latter probability will be denoted $P(r|\Delta\varphi)$.

4.3 Experiment one

In order to address the validity of the model predictions, a psychoacoustic left-right task was conducted. This experimental paradigm is a one-interval, two-alternative forced choice (2-AFC) procedure, where the subjects were presented a stimulus via headphones and were subsequently asked to answer the question “left or right?” This procedure may be described in an attempt of addressing whether the hearing sensations occurred left or right of the perceived median plane. In the left-right task, the subjects had to base their decision on an ”internal” representation of the median plane (MP), as no “center marker” was employed, in contrast to similar earlier experiments of this type (e. g., Phillips and Hall, 2005). The presentation of an acoustic marker, aside from introducing a reference, also changes the experiment in that the subjects are then able to perform the relative task of comparing the hearing sensations corresponding to the test and reference sounds instead of the absolute task of judging a single hearing sensation. In order to resemble the modeled process, the left-right task used in this chapter presents an absolute task, as the decision has to be made directly after the presentation of a single stimulus. This method also reduces the need for internal mapping that may be present with a marker or in pointer tasks and magnitude estimation. None of the subjects indicated difficulties with the task and only little training was necessary to obtain consistent results that remained stable even when repeating the experiment after several weeks (Völk et al., 2017b).

f / Hz	125	250	500	1k	2k	4k
L/dB	72	64.5	59.5	57.8	57.9	55.5

Table 4.1: Presentation levels L for the tone impulses of different frequencies f for experiment 1, determined with the goal of equal loudness based on ISO 226:2006.

4.3.1 Subjects

A total of 13 subjects (5 females, 8 males) with the age ranging from 23 to 52 years (median 26 years) participated in this experiment. All subjects were without known hearing difficulties and showed less than 20 dB deviation from the monaural pure-tone thresholds according to ISO 389-7:2005 at all studied frequencies. None of the listeners was familiar with the design or aim of the experiment. The subjects participated voluntarily and were not reimbursed for their participation.

4.3.2 Stimuli

The stimuli were 700 ms pure-tone impulses with 160 ms Gaussian slopes (10% to 90% of the amplitude) at six one-octave spaced frequencies between 125 Hz and 4 kHz. The tone impulses were presented approximately equally loud at about 58 phon, with sound-pressure levels based on ISO 226:2006 (cf., Table 4.1). IPDs were introduced by changing the starting phase of the tone impulse at one ear relative to the other while leaving the envelope unchanged. This chapter investigated 49 equidistant IPDs from $-\pi$ to π.

4.3.3 Apparatus

Stimuli were generated digitally at double-precision word length and a sample rate of 96 kHz in Matlab R2016b (MathWorks, Natick, Massachusetts, USA). The signals were transmitted to the external audio interface (Babyface; RME, Haimhausen) using an ASIO/Matlab-interface software (WindAcoustics UG, Windach, Germany), enabling the use of the full D/A-converter word length of 24 Bit. The subjects sat in a sound-insulating booth (IAC 350; IAC Acoustics, Winchester, UK), with the window covered, in order to reduce uncontrolled visual stimulation. The same pair of circumaural headphones (HD 650; Sennheiser electronic GmbH, Wedemark) was used for all experiments (selected based on Völk, 2014). The headphones were electrically calibrated by means of the nominal headphone sensitivity at 1 kHz, and the left and right headphones were equalized at the stimulus frequencies in amplitude and phase (Völk et al., 2017a) using a coupler according to IEC 60318-1:2009. Two buttons (left/right) of a wireless presenter (R400; Logitech Europe S.A., Lausanne, Switzerland) were used by the subjects for indicating their responses.

4.3.4 Procedure

Every condition (combination of IPD and frequency) was presented ten times in individually-randomized order. The data were collected in two sessions, each comprising half of the ten presentations. The two sessions were separated by at least one, but typically several hours, and lasted about 45 min each. A single session was split into two parts, with a short break (typically 5 min) after half the stimuli, during which the subject left the booth. Within the experiment, the stimuli were presented one by one to the subject, which subsequently had to respond to the question "left or right" by pressing one of the two buttons on the presenter. To prevent temporal masking effects (Fastl and Zwicker, 2007, pp. 83f), a silent interval of at least 0.3 s, individually elongated by the response time, was inserted after each stimulus.

4.3.5 Results

As expected from previous studies on IPD perception (e. g., Sayers, 1964; Yost, 1974), none of the subjects showed systematic results at 2 kHz or 4 kHz, so that results for these frequencies will not be discussed. Most of the subjects experienced a reduced ability to give consistent judgments also at 1 kHz. Figure 4.3a) depicts the results of the two subjects S5 and S8 by means of the relative frequency of the response 'Right' (black dots). Both subjects showed distinct sensitivity to IPDs between 125 Hz and 500 Hz, while only S8 was able to respond comparably reliable at 1 kHz.

At all frequencies, the results of S5 where approximately symmetric to the origin (cf., left column of Fig. 4.2), as predicted by the model, while those of S8 showed an asymmetry. For S8, the transition between left and right (represented by the relative frequencies 0 and 1, respectively) occurred shifted from π towards IPDs of smaller magnitude, to the extent that the subject categorized an IPD of $\pm\pi$ consistently as located to the left. Several subjects showed such asymmetries.

In order to quantify the extent of the asymmetry discussed in the previous paragraph, the individual data was approximated by a combination of two logistic functions. The first function described the left-right transition around $\Delta\varphi = 0$ (resulting in a more or less central hearing sensation), while the second function described the transition for lateral hearing sensations. The IPDs at which those transitions occur are referred to as center IPD $\Delta\varphi_c$ and lateral IPD $\Delta\varphi_l$, respectively. The periodicity of $\Delta\varphi$ was implemented by splitting the second function in two parts, accounting for the ranges $\Delta\varphi \in [-\pi, 0]$ and $\Delta\varphi \in [0, +\pi]$, respectively. This results in the function

$$\begin{aligned} f(\Delta\varphi, \theta) = {} & \frac{1 - 2\delta_f}{1 + e^{-k_c(\Delta\varphi - \Delta\varphi_c)}} + \delta_f \\ & + \frac{1 - 2\delta_f}{1 + e^{k_l(\Delta\varphi - \Delta\varphi_l - 2\pi)}} - \frac{1 - 2\delta_f}{1 + e^{-k_l(\Delta\varphi - \Delta\varphi_l)}}, \end{aligned} \tag{4.10}$$

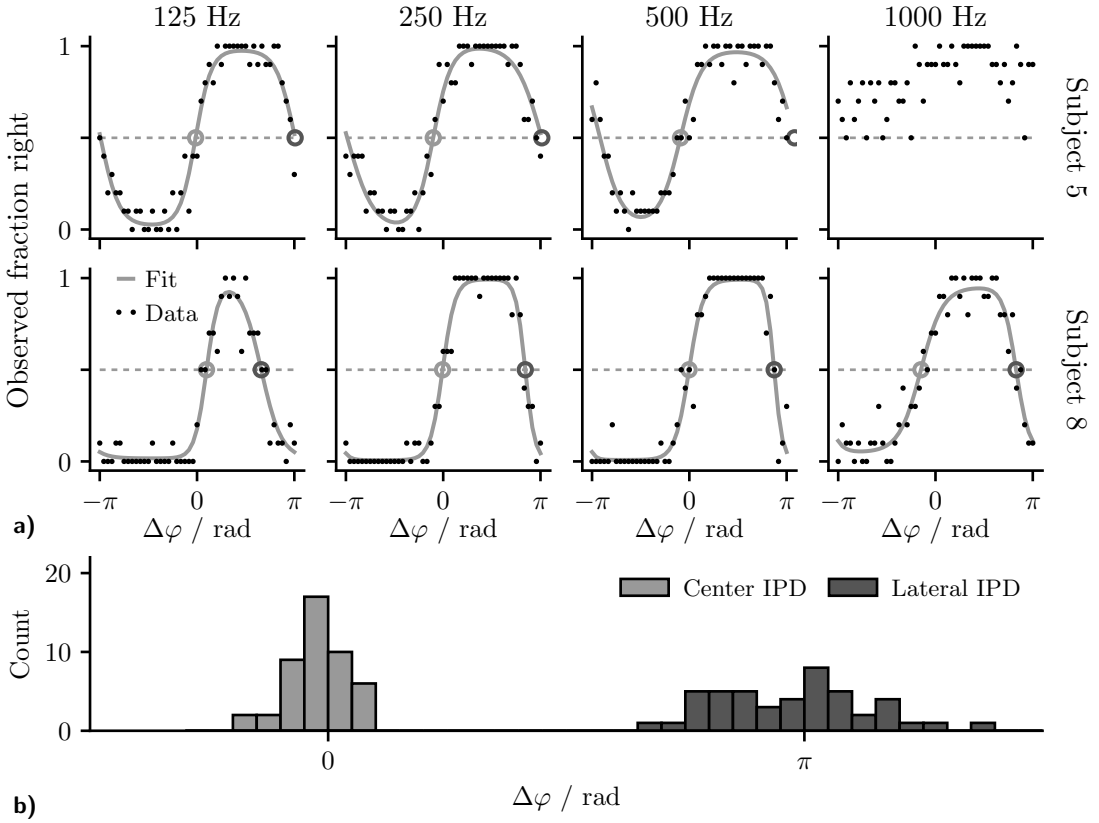

Figure 4.3: Results of experiment one (lateralization, left-right task). **a)** Individual data for the subjects S5 and S8 at different frequencies. Shown is the relative frequency of the response "right" for different IPDs (black dots), as well as fits to Eq. (4.10) (gray contours). The two circles mark the IPDs where the fitted functions result in chance level, the center IPD (light gray) and lateral IPD (dark gray). **b)** Interindividual histograms for the center IPDs and the lateral IPDs from the fits. The distributions are located at $-0.02\pi \pm 0.06\pi$ and $-0.97\pi \pm 0.16\pi$, respectively.

where θ is a vector composed of the parameters $\Delta\varphi_c$, $\Delta\varphi_l$, the slopes k_c and k_l of the two logistic functions, and δ_f, which adjusts the maximum and minimum of the logistic functions to $1 - \delta_f$ and $0 + \delta_f$, respectively.

Equation (4.10) was fitted to the individual data sets using a log-likelihood approach. To reduce the effect of outliers, δ_f was restricted to values between 0 and 0.05 (Wichmann and Hill, 2001). $\Delta\varphi_c$ and $\Delta\varphi_l$ were restricted to $[-\pi/2, \pi/2]$ and $[\pi/2, 3\pi/2]$, respectively. As mentioned above, some subjects showed unstructured results at 1 kHz, which resulted in an unreliable fit. Based on a deviance-based goodness of fit test, three fits at 1 kHz were rejected ($D = \chi^2(44) = 64.14$, 66.20, and 72.92; $p < 0.05$). The fits to the 1 kHz responses of three further subjects showed overall variations below 0.5, as for example S5 in Fig. 4.3a, indicating inconsistent reports. All six data

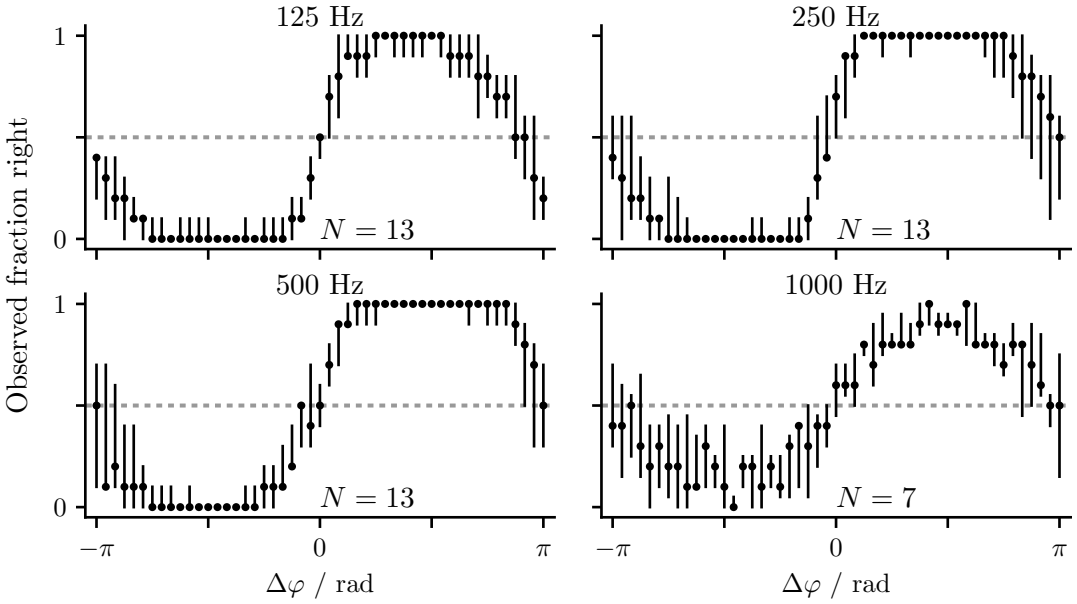

Figure 4.4: Results of experiment one (lateralization, left-right task). Medians and interquartile ranges of the individual relative frequencies of pushing the right button (10 individual repetitions per condition, 13 subjects). The results from subjects giving inconsistent reports at a specific frequency were not included (cf., inserts).

sets were excluded from further analysis.

Figure 4.3b shows the distributions of the center IPDs and lateral IPDs (N=46), as calculated from the fitted functions. The data were pooled over the frequencies, as similar distributions were found for the individual frequencies. While the center IPDs showed a rather narrow distribution around $\Delta\varphi = 0$ (M=-0.02π, SD=0.06π), the lateral IPDs, which, for clarity, are shown in the period $[0, 2\pi[$, were spread more widely (M=0.97π, SD=0.16π). Furthermore, center IPDs and lateral IPDs showed no significant correlation (Pearson's $r(46) = -0.063$, $p = 0.659$). Both findings indicate that the asymmetry is not caused by bias or a general shift of the perceived median plane, as these would influence center IPDs and lateral IPDs in a comparable manner.

Sayers (1964) and Yost (1981) both reported that large ITDs could result in multiple simultaneous hearing sensations, for example, two hearing sensations located at opposite sides of the perceived median plane. An occurrence of multiple hearing sensations at IPDs with larger magnitudes would explain the larger variance in the lateral IPD vs. the center IPD observed in this experiment: The left-right task forced the subjects to select one of the hearing sensations, which may not occur with a consistent criterion both, individually and between subjects. In order to investigate the occurrence of multiple hearing sensations, the follow-up experiment discussed in section 4.4 was designed.

The medians and interquartile ranges of the individual results of experiment one

are shown in Fig. 4.4. Due to the interindividual distribution of the asymmetries, the medians constitute approximately symmetric contours with regard to $(0, 0.5)$, with little to no asymmetry. As already shown by the histograms in Fig. 4.3b, the center IPDs of the median results shown in Fig. 4.4 are located close to $\Delta\varphi = 0$, and the lateral IPDs close to $\Delta\varphi = \pm\pi$. It is interesting to note that there are only minor structural differences between the results obtained at the different frequencies when plotted against IPD.

4.4 Experiment two

For investigating the appearance of multiple, spatially separated hearing sensations, especially in response to large IPDs, a second experiment was conducted. This experiment employed a free categorization task, where the subjects were presented a stimulus and were asked to report the number of spatially-separated hearing sensations. Compared to pointing tasks, this procedure enabled the subject to focus solely on the number of hearing sensations. In order to address whether the occurrence of multiple hearing sensations was connected to the extent of lateralization or the introduction of IPDs, the same experiment was also conducted with a reference set of pure-ILD stimuli.

4.4.1 Subjects

A total of 10 subjects (6 females, 4 males), with the age ranging from 22 to 32 years (median 26 years), participated in this experiment. All but one subject also took part in experiment one, and none of the listeners was familiar with the design or aim of the experiment. All subjects participated voluntarily and were not reimbursed for participation.

4.4.2 Stimuli and apparatus

The set of pure-tone stimuli from experiment one was also used in this experiment, but for reducing the duration of the experiment, the two higher frequencies 2 kHz and 4 kHz were omitted. The number of IPDs was reduced to 25, still in the range from $-\pi$ to π. In addition to the IPD stimuli, this experiment also included four ILD stimuli, with the ILDs $\Delta L = \pm 4\,\mathrm{dB}$ and $\Delta L = \pm 8\,\mathrm{dB}$. These values were chosen so that the maximum ILD resulted, based on empirical listening and Yost (1981) in a considerably lateralized hearing sensation. The ILDs were introduced by symmetrically amplifying and attenuating the right and left ear signals by half the desired ILD. No phase shift was introduced, and the ILD stimuli were otherwise identical to the IPD stimuli.

The identical equipment and settings as in experiment one were used, with the only exception that instead of a presenter, the subjects used a wireless numeric keypad (ID0120; LogiLink, Schalksmühle, Germany) for indicating their response.

4.4.3 Procedure

The basic experimental conditions were the same as in experiment one, but the data were collected in four sessions per subject. The sessions were separated by at least one, but typically several hours. Each session lasted about 8 min. In every session, all ILDs and IPDs of a single frequency were presented ten times each. The presentation sequence per session was individually randomized. Within the experiment, the stimuli were presented to the subject one at a time. The task was then to subsequently indicate the number of separate locations at which a hearing sensation had occurred. The subject was allowed to enter positive integers using the keypad. The next signal was presented after the input was confirmed by pressing the enter key. 0.3 s of silence, elongated by the individual response time, was inserted after each stimulus.

4.4.4 Results

Throughout the experiment, subjects consistently indicated to perceive hearing sensations at either one or two locations. Only one subject infrequently indicated three or four locations. In seven single trials, subjects entered numbers far above the typical range (mostly multiple keypress candidates as for example 111), these trials were excluded from further analysis. The results were analyzed by calculating the individual average over the denoted number per condition and subsequently calculating interindividual medians and interquartile ranges thereof. The results are shown in Fig. 4.5.

The first row of Fig. 4.5 shows the results over IPDs, the second row over ILDs, each column represents the results for one frequency. The dataset at $\Delta\varphi = 0$ is identical to that at $\Delta L = 0$. As in experiment one, a comparable trend was observed in the results at all frequencies. Hearing sensations corresponding to relatively small IPDs, approximately between $\pm 0.5\pi$, were reported mostly at one position. With increasing IPD, the subjects increasingly reported a second hearing sensation position. For ILDs on the contrary, the subjects on average reported consistently one hearing sensation, independently of ILD or frequency. These results suggest that the second reported hearing sensation at larger IPDs is related to the IPD, not the extent of lateralization.

4.5 Evaluating the Model

In order to address if and how the results of experiment one can be explained by the model introduced in section 4.2, the parameters of Eq. (4.9) were adjusted to fit the experimental data. For taking into account the asymmetries in the average results, presumably due to multiple hearing sensations, the model was fitted to the medians of the individual data. This procedure is assumed to result in a model of a typical subject (with symmetrical responses).

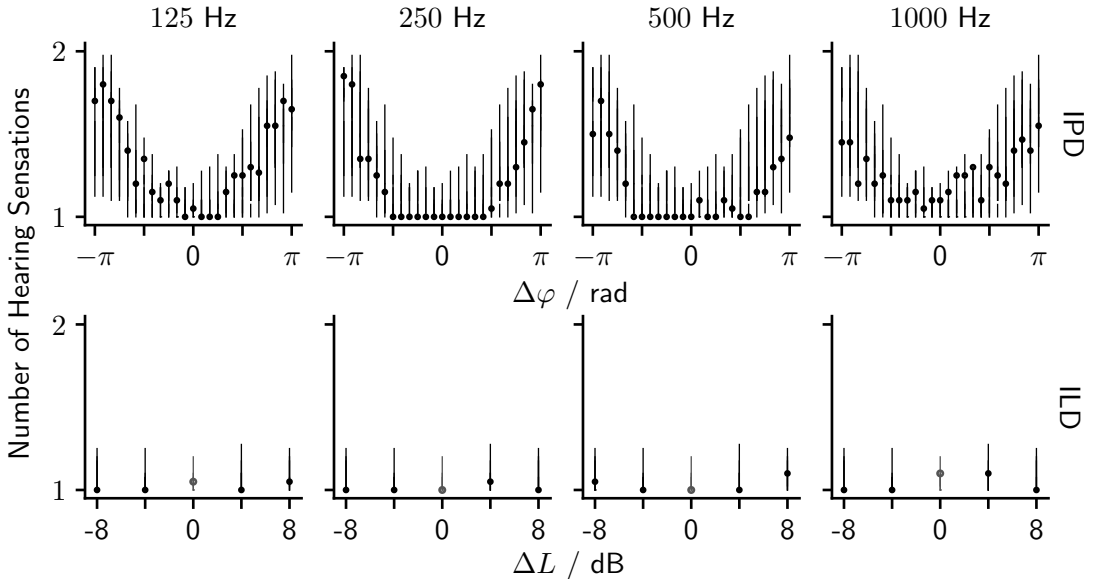

Figure 4.5: Results of experiment two (number of spatially separated hearing sensations for different ILDs, IPDs, and frequencies; N=10). Depicted are medians and interquartile ranges of individual arithmetic means over 10 repetitions per condition.

Equation (4.9) has two parameters, the best-IPD $\Delta\varphi_{\text{best}}$ and the variance of the neuronal response ${\sigma_\Delta}^2$. The variance resulting from the fit of the experimental data is not directly equivalent to the solely physiologically-motivated ${\sigma_\Delta}^2$ in Eq. (4.9). The latter only takes into account the noise of the neuronal responses, while the fits also include other noise sources, for example, introduced by the experimental method or the subjects (such as physiological and mental state, lapses, or response biases). The variance taking into account the combined noise in the experimental data is referred to as ${\sigma_{\text{fit}}}^2$, and ${\sigma_\Delta}^2$ must be considered a lower bound for ${\sigma_{\text{fit}}}^2$.

The fits of ${\sigma_{\text{fit}}}^2$ and the frequency-dependent $\Delta\varphi_{\text{best}}$ were achieved by minimizing the sum of the log-likelihoods for the median results at all frequencies. The fitting procedure revealed that the number of parameters could be further reduced by assuming ${\sigma_{\text{fit}}}^2$ to be independent of frequency, so that the only remaining frequency-dependent parameter was $\Delta\varphi_{\text{best}}$. The resulting parameters were $\sigma_{\text{fit}} = 0.28$ and the $\Delta\varphi_{\text{best}}$ values given in Tab. 4.2. As shown by Fig. 4.6a, this approach resulted in a plausible agreement between fits and data at all frequencies. It is apparent from Tab. 4.2 that the best-IPD increases with frequency. The effect of this increase on the model is visible in Fig. 4.6b, which shows the functions $\bar{R}_{\text{C,l}}$, $\bar{R}_{\text{C,r}}$, and ΔR as they result from the predictions shown in Fig. 4.6a. Summarizing, for describing the left-right task results well, the proposed model requires only a single frequency-dependent parameter, while the only further parameter can be kept independent of frequency.

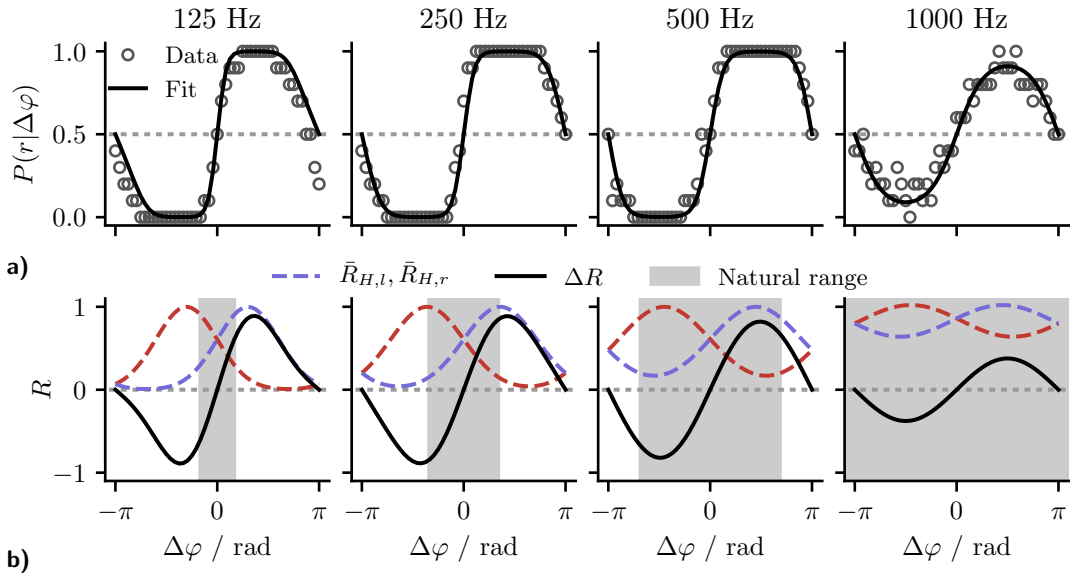

Figure 4.6: **a)** The physiologically-motivated model described by Eq. (4.9) was fitted (black contour) to the results from the psychoacoustic left-right task experiment (gray circles), resulting in $\Delta\varphi_{\text{best}}$ according to Tab. 4.2 and the frequency-independent parameter $\sigma_{\text{fit}} = 0.28$. **b)** Channel responses and their differences resulting from fitting the model parameters to the average data.

f / Hz	125	250	500	1k
$\Delta\varphi_{\text{best}}$ / rad	0.30π	0.36π	0.45π	0.66π

Table 4.2: Parameters resulting from fitting the model to the results of experiment one. The standard deviation was fitted frequency-independently as $\sigma_{\text{fit}} = 0.28$.

4.5.1 Robustness of IPD encoding

Up to this point, the neuronal left-right task was discussed based on the average neuronal responses in the two channels. A reliable neuronal code would have to convey unambiguous information over the whole naturally-occurring ITD range. In the presented model, presentation of a specific $\Delta\varphi$ results in an observed response $\mathbf{R}_{obs}$, drawn from a distribution defined by $\bar{\mathbf{R}}(\Delta\varphi)$ according to Eq. (4.6). If only $\mathbf{R}_{obs}$ is known, it is not possible to directly conclude the originating IPD, as the same $\mathbf{R}_{obs}$ could have resulted from different distributions defined by their respective $\bar{\mathbf{R}}$ $(\Delta\varphi)$. It is, however, possible, to calculate the probability for each IPD to result in the neuronal response $\mathbf{R}_{obs}$.

For discussing this process, it is helpful to derive how well the distribution defined by the mean response $\bar{\mathbf{R}}(\Delta\varphi_1)$, resulting from a first IPD $\Delta\varphi_1$, can be separated (distinguished) from the distribution defined by $\bar{\mathbf{R}}(\Delta\varphi_2)$, resulting from a second IPD

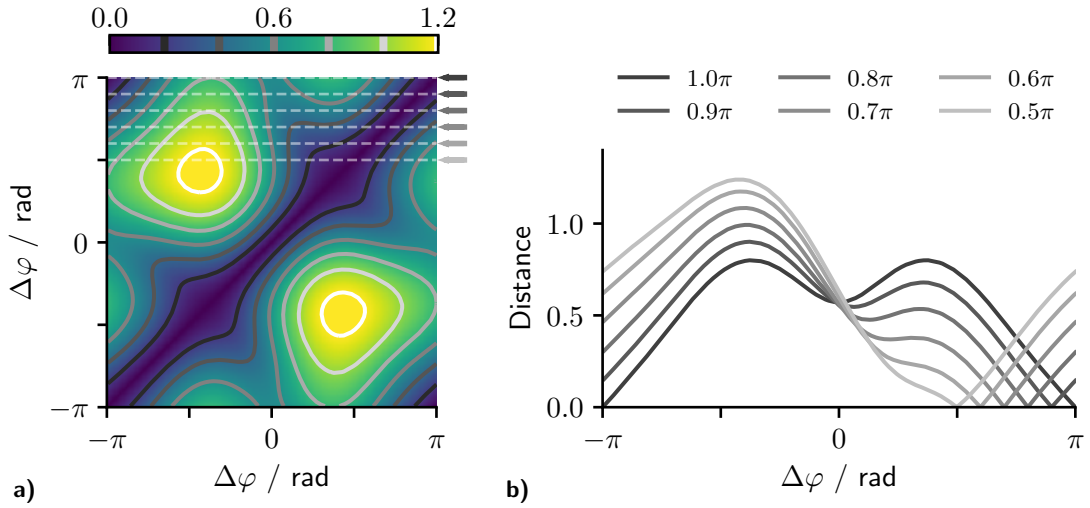

Figure 4.7: Separability of different IPD conditions. **a)** Distance between $\bar{\mathbf{R}}$ for any two IPDs as calculated from Eq. (4.11). A large distance indicates a high neuronal separability between two IPDs. **b)** Distances for a set of selected IPDs, marked by color-coded arrows in a). By definition, each contour shows zero distance at the corresponding selected IPD, while exhibiting an additional local minimum at $\Delta\varphi = -\pi$.

$\Delta\varphi_2$. The two mean responses $\bar{\mathbf{R}}(-\pi)$ and $\bar{\mathbf{R}}(+\pi)$, as an example, for being physically identical, define the same distribution and the separability of the two IPDs $+\pi$ and $-\pi$ is zero.

A measure for quantifying separability is d' (d-prime, Green, 1960). The d-prime value for two random variables A and B that share the same standard deviation is calculated as the distance between their mean values, divided by their common standard deviation. d' for the two IPDs in the proposed model is directly proportional to the distance

$$d(\Delta\varphi_1, \Delta\varphi_2) = \|\bar{\mathbf{R}}(\Delta\varphi_1) - \bar{\mathbf{R}}(\Delta\varphi_2)\| \tag{4.11}$$

between the corresponding $\bar{\mathbf{R}}$, where $\|\bullet\|$ indicates the Euclidean norm. This distance d instead of d-prime will be used to discuss the separability, as the calculation of d-prime would require speculating on σ_C.

Figure 4.7a shows the distance d between the $\bar{\mathbf{R}}$ of any two IPDs. By definition, d is symmetric around the diagonal, where $d = 0$. Figure 4.7b shows d for the selected horizontal cross-sections indicated by arrows and dashed white lines at the top of Fig. 4.7a. Consequently, Fig. 4.7b indicates the distance between $\bar{\mathbf{R}}$ for selected fixed IPDs $\Delta\varphi_1$ and all other IPDs $\Delta\varphi$, as functions of $\Delta\varphi$. As previously discussed, the distance for $\Delta\varphi_1 = \pi$ (black contour) is zero at $\Delta\varphi = \pm\pi$, while all other $\Delta\varphi_1$ show only one zero. At the same time, all depicted curves also exhibit a local minimum at $\Delta\varphi = -\pi$, indicating that any response to $\Delta\varphi_1$ for the respective contour also has a high probability to originate from $\Delta\varphi \approx -\pi$. There are two obvious strategies how

this ambiguity could be handled by the auditory system: I) The IPD $\Delta\varphi$ leading to the highest probability (lowest distance) is selected, which eventually would lead to a single hearing sensation at the most likely position. II) All IPDs $\Delta\varphi$ leading to a probability that fulfills a certain criterion are considered relevant and may eventually lead to hearing sensations, possibly with different loudness, dependent on the distance. Both, strategy I) and strategy II) would predict the results of experiment one. The results of experiment two, on the other hand, is not in agreement with strategy I), which would predict a single hearing sensation at all times, while strategy II) predicts additional hearing sensations, appearing especially at larger IPDs.

4.6 Discussion

4.6.1 Psychoacoustic experiments

Experiments similar to the left-right task in experiment one (Sec. 4.3) were previously published (e. g., Cherry and Sayers, 1956; Kunov and Abel, 1981; Abel and Kunov, 1983). While most of these studies were conducted using ITDs (whole waveform delays) instead of IPDs, they gave comparable results. In this chapter, some subjects showed a reduced or absent ability to solve the left-right task at 1 kHz, which is in agreement with Cherry and Sayers (1956), who also reported that some subjects had difficulties solving their task for frequencies from 800 Hz upwards. All cases where the subject was able to solve the left-right task resulted in periodic responses over IPD, especially with, as required, $P(r|+\pi) \approx P(r|-\pi)$. Some individual data sets showed a considerable deviation from the expected symmetry with respect to the origin, with $P(r|\pm\pi)$ resulting in probabilities far from chance level, which, as discussed above, could be attributed to the appearance of an additional hearing sensation, as previously reported (e. g., Yost, 1981; Sayers, 1964), combined with the individual response strategy. The occurrence of multiple hearing sensations was supported by the results of the second experiment (Sec. 4.4), in which subjects were asked to report the number of spatially separated hearing sensations. In this experiment, subjects increasingly started to report two hearing sensation for increasing IPDs, especially when $|\Delta\varphi| > \pi/2$. This is consistent with the data of Yost (1981) who, in an absolute localization task, found that $|\Delta\varphi| > \pi/2$ resulted in a tendency for the subjects to report two hearing sensations at opposite ends of the employed scale. While the left-right task does not indicate the hearing sensation position, informal listening and reports in the present chapter confirmed the discussion of Yost (1981).

4.6.2 Model of neuronal IPD processing

Section 4.2 introduced a simplified but physiologically-motivated description of the neuronal processing of IPDs. It is assumed in this description, that the auditory system makes use of the combined response of all neurons per hemisphere so that

the overall response can be described by the mean channel response formulated in Eqs. (4.3) and (4.4). This approach contradicts a labeled-line code (such as Jeffress, 1948), which would employ the information conveyed in single-neuron responses. There are several operations to combine the responses of the two channels, a popular one being the hemispheric difference, in which the two responses are subtracted (Goodman et al., 2013; Stecker et al., 2005; Brughera et al., 2013). This chapter does not directly assume a combining operation by treating both channels independently in the form of the response vector $\bar{\mathbf{R}}$. By only considering a left-right assignment task, it was possible to evade an assumption on the combining operation and predict the performance in the left-right task solely based on the statistics of the responses in the two channels. The derivation of this prediction also leads to a rate difference ΔR, the same measure previously used in hemispheric difference decoders. While ΔR in this chapter solves the left-right task, it is not a suitable measure to directly decode IPDs, for being ambiguous, with each ΔR encoding two IPDs (see Fig. 4.2a.

The model predictions of the probability to attribute a response $\mathbf{R}_{obs}(\Delta\varphi)$ to a positive or a negative IPD were evaluated by comparing the psychoacoustic left-right task results to the model experiment. This step assumes that a subject's ability to attribute a given stimulus to the left or right is determined by the statistics of the neuronal code: If the response to a positive IPD has a high probability to be confused with responses to a negative IPD, there should also be a high probability of perceptual confusion. The assumption that perception is determined by the underlying neuronal code is plausible at least as an upper boundary: Information not encoded in the neuronal code can also not be used for forming perceptions. Whether all information included in the neuronal code is also used for forming perceptions can only be assumed.

For all frequencies, a good agreement between model and experimental data could be reached by adjusting the frequency-dependent parameter $\Delta\varphi_{\text{best}}$ and the frequency-independent parameter ${\sigma_\Delta}^2$. $\Delta\varphi_{\text{best}}$ summarized in Tbl. 4.2 were determined solely based on the psychoacoustic data. As the model was strongly motivated by electrophysiological recordings, Fig. 4.8 compares the obtained values (black circles) to the results of single-neuron recordings in the inferior colliculus (IC) of cats (gray symbols: Joris et al., 2006; McAlpine et al., 2001). To obtain the points shown for Joris et al. (2006), data from 176 single neurons were digitized from their Fig. 2. Based on their characteristic frequency, the results were assigned to five log-spaced bins, and the mean value of each bin was calculated. A linear approximation following the equation

$$\Delta\varphi_{\text{best}}(f) = 2\pi\Delta t_0 f + \Delta\varphi_0 \tag{4.12}$$

was fitted to each data set where $\Delta\varphi_0$ can be interpreted as a constant phase and Δt_0 as a constant time delay. The fits resulted in $\Delta\varphi_0$ values of 0.25π for the model and 0.17π for both data sets. The Δt_0 values were 204 µs for the model and 110 µs, 59 µs for the data from Joris et al. (2006) and McAlpine et al. (2001) respectively. While some difference, especially in Δt_0 was found between the model and the measurement

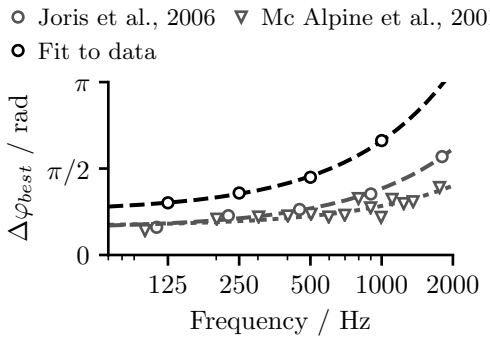

Figure 4.8: Comparison between the best-IPD parameter determined by fitting the model to the left-right task experiment (black circles) with single neuron recordings digitized from McAlpine et al. (2001) (gray triangle) and Joris et al. (2006) (gray circle). All three could be described well by a linear approximation (dashed lines)

data, the model followed the linear relationship of $\Delta\varphi_{\text{best}}$ that was also found in the physiological measurements. One reason for this difference could be the dissimilarity in the naturally-occurring free-field ITD-range due to the difference in head size between the two species.

In the experiments, IPDs with magnitudes above approximately $\pi/2$ increasingly resulted in multiple hearing sensations. Considering this phenomenon in the light of the presented model might give insight in the processing of the neuronal code. Similar to $\Delta\varphi = 0$, $\Delta\varphi = \pm\pi$ cannot be unambiguously attributed to left or right of the perceived median plane, but is at the same time encoded as a large IPD, as depicted by Fig. 4.1b. Intuitively, one would expect the hearing sensation to jump from one side to the other when transitioning from large positive IPDs past the lateral IPD of $\pm\pi$ towards negative IPDs with large magnitudes. Instead, the experimental data in combination with the result of absolute lateralization (Yost, 1981) suggest that, with increasing positive IPD, an additional but weaker hearing sensation, located towards the left, starts to appear. In the left-right task results, this additional hearing sensation increases in strength, until matching that of the other hearing sensation at $\Delta\varphi = \pm\pi$. The model explains this phenomenon when looking at the separability of the responses to different IPDs: Caused by the cyclic nature of $\bar{\mathbf{R}}(\Delta\varphi)$, as seen in Fig. 4.7b, a low separability occurs between IPDs with large magnitudes but different signs, in that an observed response $\mathbf{R}_{obs}$ due to a large positive IPD also has a high probability to be caused by a negative IPD with a large magnitude. This indicates that a hearing sensation position might be based on a probabilistic interpretation of the neuronal code: hearing sensations arise at all positions with high probabilities to result in the observed response $\mathbf{R}_{obs}$. This opposes a maximum likelihood approach, which would select the single location with the highest probability to result in the observed $\mathbf{R}_{obs}$ and would thus not result in two hearing sensations.

Unlike other recent models (e. g., Dietz et al., 2009; Takanen et al., 2014; Encke

and Hemmert, 2018), the presented model does not provide a framework that can be directly applied to sound signals but rather presents a formalized description of the neuronal representation of IPDs in response to dichotically-presented tone impulses. In its current form, all responses are considered time-invariant. As such, the model does not account for effects due to adaptation or stimulus duration, both of which are known to influence perception (Tobias, 1959; Phillips and Hall, 2005). The assumption of time-invariance appears justified for applying the model to the presented data, which was acquired for stimuli of constant length, in randomized order, and with a temporal sequence that minimizes the effects of adaptation.

Furthermore, the model only considers pure-tone stimuli and is, in its current form, not applicable to stimuli with multiple frequency components. This limitation was introduced deliberately, in order to focus on binaural interaction, with as little as possible additional effects, such as those of the auditory periphery or cross-frequency integration. It might be argued that only considering a left-right task to parameterize the model might be too simplistic but it has to be kept in mind that the model is completely described using only two parameters of which only one was allowed to vary with frequency. Using this minimal set of parameters, the model was still able to describe the outcome of the left-right task while at the same time resulting in the linear $\Delta\varphi_{\text{best}}$-frequency relationship found in single neuron recordings.

4.7 Summary and Conclusions

This chapter introduced a physiologically-motivated two-channel model of IPD processing. The model was evaluated using two psychoacoustic experiments: I) A left-right task that investigated which IPDs were perceived to the left or to the right of an internal center, referred to as perceived median plane and II) A categorization task which investigated the number of spatially separated hearing sensations that occurred for different IPDs. The main findings were:

- Using only two parameters, one frequency-independent, one adjusted frequency-dependently, the outcome of the left-right task was directly predicted by the statistics of the modeled neuronal responses.
- Large IPDs may result in multiple, spatially separated hearing sensations. Using the model, this is explained by the low separability between the response statistics for different IPDs. The ambiguous encoding of IPDs is resolved by several hearing sensation at positions corresponding to all IPDs that cause a probability for the observed response to exceed a certain criterion.

Concluding, the presented model is able to explain several aspects of pure-tone lateralization, while minimizing the number of parameters and, at the same time, retaining physiological plausibility.

CHAPTER 5

Predicting IPD Thresholds Using the Two-Channel Model

5.1 Introduction

Two different psychoacoustical measures are often used to characterize the performance of subjects in localizing sound sources in space. Absolute measures such as the deviation between the location indicated by a subject and the physical location of a sound source (Makous and Middlebrooks, 1990) or the left-right experiment used in the last chapter are measures for the accuracy of sound localization while, relative measures such as the minimal audible angle (MAA) or just noticeable differences (JNDs) (Mills, 1958) can be used to characterize the acuity of sound localization.

Several studies have previously reported interaural time difference (ITD) JNDs for tones (e.g. Klumpp, 1956; Zwislocki and Feldman, 1956; Brughera et al., 2013). Most of these studies determined the JNDs around a reference of zero ITD with only a few studies also determining difference thresholds around other references (e.g. Hershkowitz and Durlach, 1969; Domnitz and Colburn, 1977) and even these studies only provide results for 500 Hz and around a few other reference ITDs. Both studies showed a slow increase in JNDs with increasing reference ITD up to about a quarter stimulus period after which the JNDs increased steeply. Based on these results, several different modeling approaches have been used to investigate ITD or interaural phase difference (IPD) difference thresholds around zero ITD (e.g. Brughera et al., 2013; Encke and Hemmert, 2018) and also around other references (Colburn, 1973; Stern and Colburn, 1985; Hancock, 2004). In agreement with the psychoacoustic results, the model of Stern and Colburn (1985) predicts an increase in JNDs with reference ITD for a 500 Hz tone, but this model is based on an internal representation following the delay-line hypothesis of Jeffress which is difficult to reconcile with newer physiological evidence in mammals (Grothe et al., 2010). The model of Hancock (2004) does agree with mammalian physiology but only showed an increase in JND with reference ITD for noise and not for tonal stimuli. Therefore, to the best knowledge, none of the currently available models can account for both the dependence of JNDs on reference ITDs while still maintaining physiological plausibility.

In the previous chapter, it was shown that the two-channel model (TC-model) is able to predict the outcome of an absolute left-right experiment. This chapter investigates whether the neuronal representation as hypothesized for the TC-model can also be used to predict JNDs. To that end, a psychoacoustic experiment was conducted in which JNDs where determined at four frequencies and eleven references. This data is then compared against the predictions of the TC-model. The same model is further used to investigate the high-frequency limit of ITD sensitivity that was previously shown to exist at around 1500 Hz (Brughera et al., 2013; Zwislocki and Feldman, 1956).

5.2 Experiment

Only limited data on the dependence between ITD JNDs and reference ITDs is currently available for tones so that this experiment is aimed to provide the data needed to evaluate the TC-model. In agreement with the previous chapter, IPDs instead of ITDs will be used to describe the interaural time disparities. For tones, the two can be related using Eq. 4.1.

5.2.1 Subjects

A total of 11 subjects (5 female, 6 male) participated in the experiment. Their age ranged from 25 years to 38 years (median = 26 years). All subjects were without known hearing difficulties and showed less than 20 dB deviation from the monaural pure-tone thresholds according to ISO 389-7:2005 at all studied frequencies. The subjects participated voluntarily and were not reimbursed for their participation. All subjects also took part in the experiment described in Sec. 4.3

5.2.2 Stimuli

To enable a comparison between the two studies, the basic stimulus parameters were identical to the ones used in the previous chapter. The stimuli were 700 ms pure-tone impulses with 160 ms Gaussian slopes (10% to 90% of the amplitude) at four one-octave spaced frequencies between 125 Hz and 1 kHz. The tone impulses were presented approximately equally loud, at about 58 phon, with sound-pressure levels based on ISO 226:2006 (cf., Table 4.1). IPDs were introduced by changing the starting phase of the tone impulse at one ear relative to the other.

Each presentation consisted of two stimuli with the IPDs $\Delta\varphi_1$ and $\Delta\varphi_2$. The two stimuli were presented consecutively in random order with a 300 ms inter-stimulus gap. $\Delta\varphi_1$ and $\Delta\varphi_2$ were arranged symmetrically around one of 11 reference IPDs (RIPDs) $\Delta\varphi_{\text{ref}}$ which where evenly spaced between $\pm 0.3\pi$. The difference between the two IPDs $D_{\text{IPD}} = \Delta\varphi_1 - \Delta\varphi_1$ being one of 22 evenly spaced values between $\pm 0.4\pi$ which will be called delta IPDs (DIPDs). The actually presented IPDs can be calculated as

$$\Delta\varphi_1 = \Delta\varphi_{\text{ref}} + D_{\text{IPD}}/2$$
$$\Delta\varphi_2 = \Delta\varphi_{\text{ref}} - D_{\text{IPD}}/2. \tag{5.1}$$

Section 4.4 reported that IPDs larger than approximately π often resulted in multiple hearing sensation (HS) which were perceived at different locations, often at both sides of the medial plane. The method used in this chapter assumes a monotonic dependence of lateralization on IPD – an increase in IPD is assumed to result in an increase in perceived lateralization. Monotonicity might not be fulfilled if multiple HSs are perceived in different locations so that the experiment was designed to avoid IPDs exceeding π restricting the range of RIPDs to the values stated above.

5.2.3 Apparatus

The same apparatus as in the previous chapter was also used in this chapter, the description is repeated here: The stimuli were generated digitally at double-precision word length and a sample rate of 96 kHz in Matlab R2016b (MathWorks, Natick, Massachusetts, USA). The signals were transmitted to the external audio interface (Babyface; RME, Haimhausen) using an ASIO/Matlab-interface software (WindAcoustics UG, Windach, Germany), enabling the use of the full D/A-converter word length of 24 Bit. The subjects sat in a sound-insulating booth (IAC 350; IAC Acoustics, Winchester, UK), with the window covered, in order to reduce uncontrolled visual stimulation. The same pair of circumaural headphones (HD 650; Sennheiser electronic GmbH, Wedemark) was used for all experiments (selected based on Völk, 2014). The headphones were electrically calibrated by means of the nominal headphone sensitivity at 1 kHz, and the left and right headphones were equalized at the stimulus frequencies in amplitude and phase (Völk et al., 2017b) within a coupler according to IEC 60318-1:2009. Two buttons (left/right) of a wireless presenter (R400; Logitech Europe S.A., Lausanne, Switzerland) were used by the subjects for indicating their responses.

5.2.4 Procedure

Each of the 242 combinations of RIPD and D_{IPD} at each of the four frequencies was presented five times. The data was collected in two sessions per frequency each session lasting about 30 min. A single session was split into two parts with a short break (typically around 5 min) during which the subject left the booth. The order of presentations within each session was individually randomized.

Within the experiment a pair of stimuli with different IPDs where presented to the subjects which subsequently had to indicate, by pressing one of two buttons, whether they perceived the second stimulus to be located to the left or the right of the first stimulus.

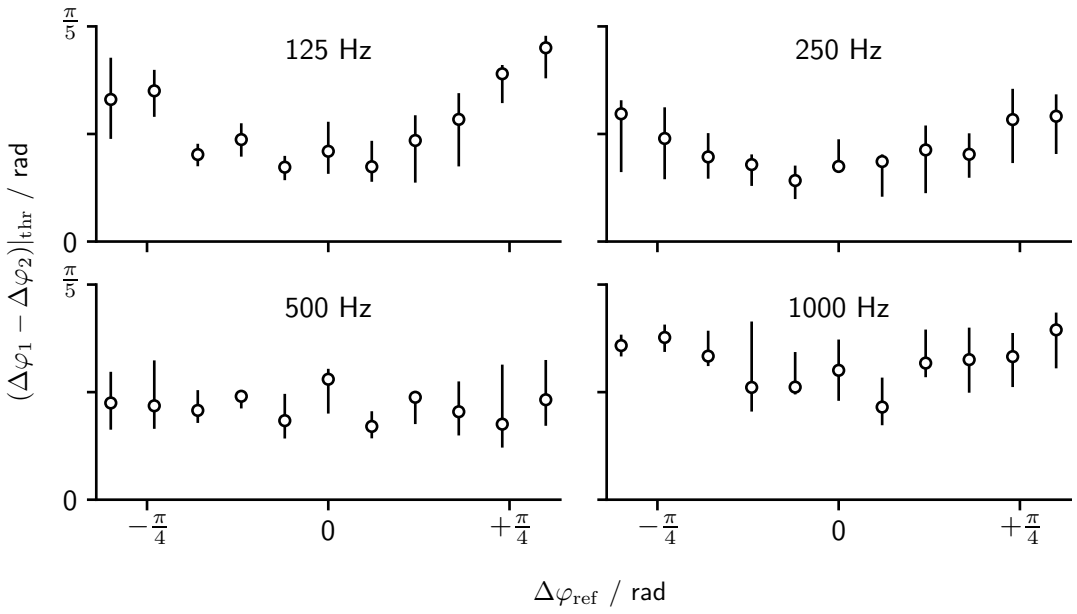

Figure 5.1: Interindividual median and interquartile ranges of the JNDs $(\Delta\varphi_1 - \Delta\varphi_2)\,|_{\mathrm{thr}}$ determined for the different RIPD at the four frequencies.

5.2.5 Results

For each RIPD and subject, psychometric functions where generated by calculating the mean response over the ten combined presentations of each $\pm\,(\Delta\varphi_1 - \Delta\varphi_2)$. A Weibull-function was subsequently fitted to each psychometric functions by using a log-likelihood approach. To reduce the effect of lapses, both the chance and the saturation level of the Weibull function where allowed to vary within the ranges of 0.45 to 0.55 and 0.95 to 1 respectively. (Wichmann and Hill, 2001). JNDs where defined as the DIPD at which the Weibull function crossed the 75 % correct line.

Figure 5.1 shows the interindividual results for the JNDs determined at the different RIPD and frequencies. Positive and negative RIPD of the same absolute value resulted in approximately the same JNDs. At lower frequencies, the JND increased with RIPD while this increase is less pronounced at higher frequencies with 500 Hz only showing little to no change in JND with RIPD.

5.3 Modeling IPD detection thresholds

To determine JND thresholds in the context of the TC-model which was introduced in the previous chapter, it is of interest how well the distribution of responses due to one IPD $\Delta\varphi_a$ can be separated from the responses due to a second IPD $\Delta\varphi_b$. Two IPDs that result in a very similar response distribution and thus a low separability are also difficult to distinguish based on their observed responses $\mathbf{R}_{obs}$. In the TC-model, the

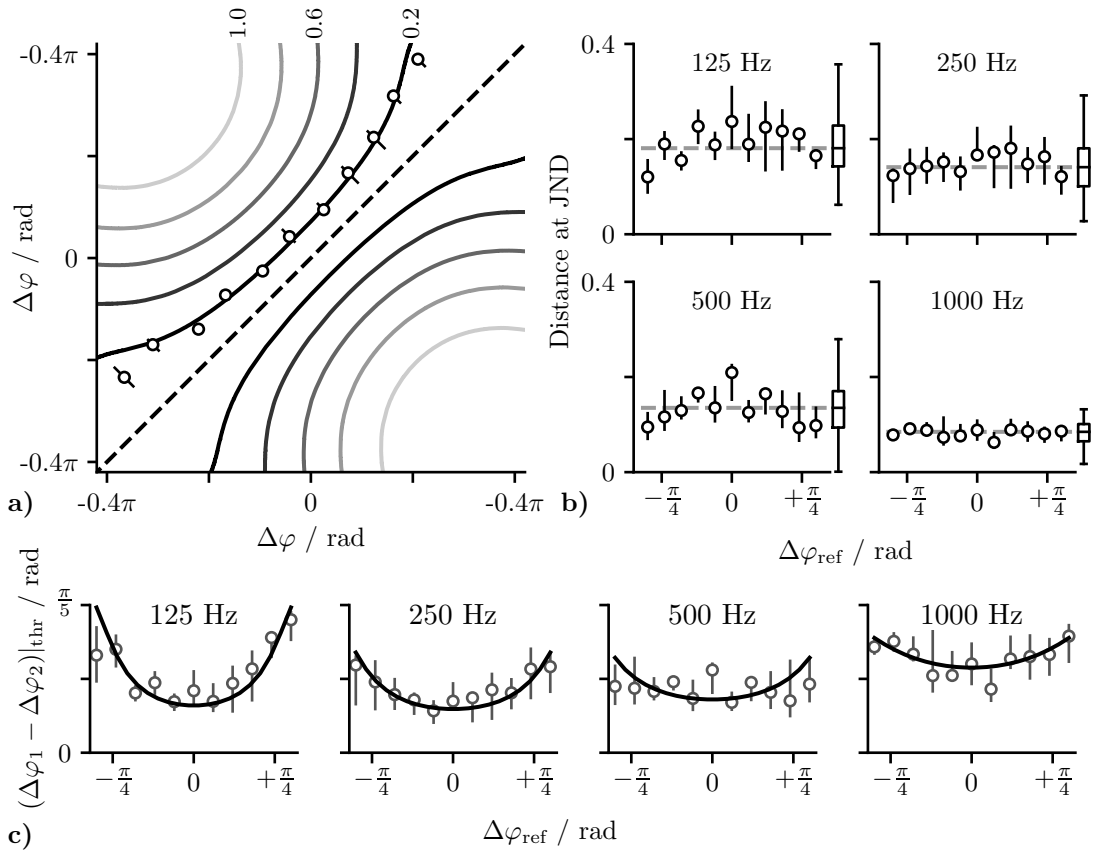

Figure 5.2: **a)** Contour lines for the distances between any two IPDs in the model parameterized for the 125 Hz condition. The contour lines are shown color coded with a spacing of 0.2. Distance values are denoted at the top border of the plot. The interindividual JNDs determined at different RIPD for 125 Hz were inserted as circles with lines indicating their inter-quartile ranges. **b)** Distances calculated for each JND that was determined in the experiment. The boxplot visualizes the pooled distribution of distances at a given frequency with the dashed lines indicating the median of these distances. **c)** The same data as in Fig. 5.1 with the dashed gray line indicating the JNDs predicted by the TC-model for d_{thr} set to the median distance of the respected frequency as indicated in b)

	125 Hz	250 Hz	500 Hz	1 kHz
d_{thr}	0.18	0.14	0.14	0.08

Table 5.1: Median distance values d_{thr}.

separability between any two IPDs is determined by the distance as given by Eq. (4.11). Only one parameter, $\Delta\varphi_{\mathrm{best}}$, is necessary to determine this distance. This frequency dependent parameter was previously determined in Sec. 4.5 by fitting the model to the results of a left-right task experiment. The parameters are listed in Table 4.2.

In this chapter, the IPD-JND is defined as the DIPD at which the subject is able to determine the perceived relative direction of the two stimuli with 75% confidence. If the perception of stimuli locations is constrained by the statistics of the underlying neuronal code as assumed in the TC-model, there should be a single distance d_{thr} which corresponds to the 75% performance at JND. Using Eq. (4.11) with the parameters summarized in Tbl. 4.2, it is possible to calculate the distances d between any two IPDs. Figure 5.2a shows the iso-distance lines for the TC-model at 125 Hz. In the same plot, the JND values that were determined at 125 Hz are shown in the form of circles with the attached lines indicating the inter-quartile ranges. From this visualization, it is visible that the JNDs are described well by the $d = 0.2$ iso-line, which supports the assumption of a constant d_{thr} at JND. Using Eq. (4.11) it is also possible to calculate the distance associated with each JND value. Figure 5.2b shows the interindividual distances calculated for the JNDs at different RIPD and frequencies with the dashed gray line indicating the interindividual median over all distances at a given frequency. While there are some systematic deviations from this line, they are small so that the distances will be approximated by one constant d_{thr} at each frequency. These values of d_{thr} are summarized in Tbl. 5.1. Having determined d_{thr}, these values can be used to calculate JNDs at any RIPD. This is done by finding the DIPD that, at a given RIPD, results in the distance of d_{thr}:

$$d(\Delta\varphi_{\mathrm{ref}}, (\Delta\varphi_1 - \Delta\varphi_2)\,|_{\mathrm{thr}}) = d_{\mathrm{thr}} \tag{5.2}$$

Figure 5.2c) shows the JND values calculated from the TC-model compared to the JND values determined by the experiment.

5.3.1 The High Frequency Limit of IPD perception

Using the TC-model introduced in the previous chapter, only two parameters are necessary to predict JNDs at any frequency. The TC-model is defined by $\Delta\varphi_{\mathrm{best}}$ and JNDs are calculated using d_{thr}. Up until this point, all parameters were based on experimental results gained from the same set of subjects. In the following section, this chapter generalizes the TC-model to be applicable at other frequencies. As discussed

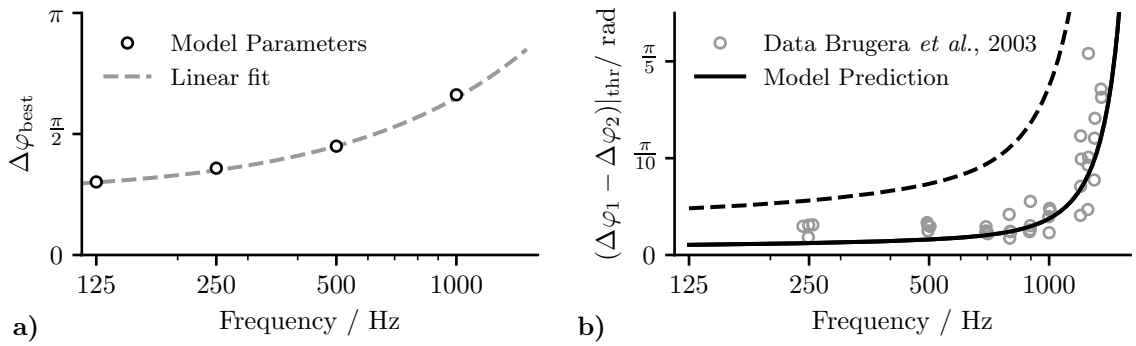

Figure 5.3: a) Linear approximation of $\Delta\varphi_{\text{best}}(f)$ following Eq. 5.3 with the parameters $t_{\text{best}} = 0.2\,\text{ms}$ and $\varphi_{\text{best}} = \pi/4$. **b)** JND predictions of the model for $d_{\text{thr}} = 0.05$ (solid black line) and for $d_{\text{thr}} = 0.14$ (dashed black line) compared to the JNDs of four subjects which were taken from Brughera et al. (2013)

in Sec. 4.6, the frequency dependency of $\Delta\varphi_{\text{best}}$ can be approximated by the linear function

$$\Delta\varphi_{\text{best}}(f) = 2\pi\Delta t_0 f + \Delta\varphi_0 \tag{5.3}$$

where $\Delta t_0 = 0.2\,\text{ms}$ and $\Delta\varphi_0 = \pi/4$. Using this approximation of $\Delta\varphi_{\text{best}}(f)$ enables the TC-model to predict JNDs at arbitrary frequencies.

Previous studies reported a steep increase in the JND at zero RIPD for frequencies around 1.5 kHz (Brughera et al., 2013; Zwislocki and Feldman, 1956). To test whether the high-frequency limit of IPD perception is predicted by the TC-model, JNDs at zero RIPD where calculated for frequencies from 125 Hz to 1500 Hz. To further simplify the calculation d_{thr} was assumed to be frequency independent. The gray circles in Fig. 5.3b) indicate JND values that where digitized from Brughera et al. (2013). As the absolute values of JNDs reported by Brughera et al. (2013) where considerably lower than the values determined in this chapter (discussed in Sec. 5.4.2), d_{thr} was reduced until the magnitude of predicted JNDs matched the reported values at $d_{\text{thr}} = 0.05$. The black curve in Fig. 5.3b) indicates the resulting prediction of the TC-model. From this plot, it is visible that the model prediction captures both, the overall trend of the JNDs at lower frequencies as well as the steep increase in JND around 1.5 kHz.

5.3.2 Improvement of Model Parameter Fittings

One problem with the linear approximation of the parameters as discussed above is that the interpolation $\Delta\varphi_{\text{best}}(f)$ crosses the value of 0.5π at a frequency of about $f = 700\,\text{Hz}$. As a result of the symmetric response function, $\Delta\varphi_{\text{best}} = 0.5\pi$ results in symmetric responses around $\Delta\varphi = \pi/2$ so that $\bar{\mathbf{R}}(\pi/2 + \Delta\varphi_a) = \bar{\mathbf{R}}(\pi/2 - \Delta\varphi_a)$ for any $\Delta\varphi_a$ which would also mean that $\bar{\mathbf{R}}(0) = \bar{\mathbf{R}}(\pi)$. As a result, an IPD of zero would not be distinguishable from an IPD of π which is not in agreement with perception. One

way for the model to remain applicable also in this frequency range is to move the case of $\Delta\varphi_{\text{best}} = 0.5\pi$ out of the physiological relevant frequency range. Figure 4.8 suggests that the slope Δt_0 of the $\Delta\varphi_{\text{best}}(f)$ approximation is larger than what is observed in single neuron recordings. If keeping the condition that connects the width of the channel response function to the best-IPD ($w_{\text{C}} = |\Delta\varphi_{\text{best}}|$) as introduced in Sec. 4.2, it is not possible to reduce the slope without compromising the JND predictions. If, on the other hand, this constrained is removed and only the slope of $\Delta\varphi_{\text{best}}(f)$ is changed while keeping the slope of $w_{\text{C}}(f)$, it is possible to find parameters with which the model reliably predicts both the results of the left-right task as well as JNDs while being applicable within the whole physiologically relevant frequency range. Figure 5.4 compares the predictions for two different parameter approximations: The original approximation as determined by fitting the results in the previous chapter as well as a corrected approximation where the slope of $\Delta\varphi_{\text{best}}(f)$ is set to 60 µs which is similar to the slope found for the data of McAlpine et al. (2001). Using this corrected approximation defined by the equations

$$\begin{aligned} \Delta\varphi_{\text{best}}(f) &= 2\pi \cdot 60\,\mu\text{s} \cdot f + 0.25\pi \\ w_{\text{C}}(f) &= 2\pi \cdot 200\,\mu\text{s} \cdot f + 0.25\pi, \end{aligned} \tag{5.4}$$

shifts $\Delta\varphi_{\text{best}}(f) = 0.5\pi$ to $f = 2084\,\text{Hz}$ which is located outside of the physiologically relevant frequency range. As visible from Fig 5.4, this approximation did not compromise the ability of the model to predict the outcome of the JND or left-right experiment.

5.4 Discussion

5.4.1 Psychoacoustic experiments

As previously discussed, surprisingly little results on JNDs at different RIPDs can be found in the literature. Domnitz (1973) and Hershkowitz and Durlach (1969) measured ITD JNDs for a 500 Hz tone at four or three reference-ITDs respectively. Both studies only found small to no increases in JNDs for reference-ITDs smaller than values that would be equivalent to RIPDs of 0.5π. Larger values resulted in a strong increase in JNDs with Hershkowitz and Durlach (1969) reporting that the subjects reported confusion and were unable to reliably solve the task at larger reference-ITDs. This inability to solve the task at large reference-ITDs is in agreement with the appearance of multiple hearing sensations as reported in Sec. 4.4 as the psychoacoustic task might become ambiguous in the presence of multiple HSs. The small change in JNDs with RIPD is agreement with the results presented in this chapter which also showed little RIPD dependent change in JNDs at 500 Hz. More data is available for ITD JNDs determined at zero RIPD (e.g. Klumpp, 1956; Zwislocki and Feldman, 1956; Brughera et al., 2013). All studies agree with a decrease in ITD-JNDs with increasing

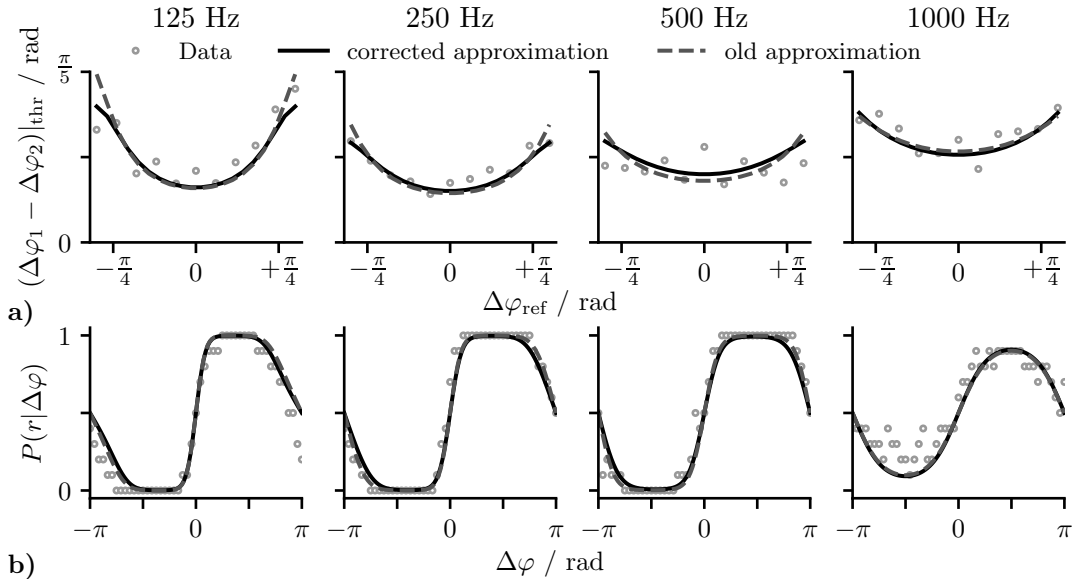

Figure 5.4: Comparison of the prediction for two different parameter approximations. The corrected approximation follows Eq. 5.4 while the old approximations follows The linear approximation shown in Fig. 5.3 with $w_C = |\Delta\varphi_{\mathrm{best}}|$.**a)** Comparison for the results of the JND experiment. **b)** Comparison for the results of the left-right experiment presented in Sec. 4.3

frequency until about 1000 Hz, which is followed by a steep increase in JNDs with a high-frequency limit located around 1500 Hz. While the psychoacoustic data presented in this chapter follows the same trend, the absolute JNDs were larger than those reported in previous studies. It is well documented that the performance in these JND tasks depends strongly on training and most JND experiments are conducted with highly trained subjects while our study was deliberately conducted with untrained subjects so that the results remained comparable to those obtained in the previous chapter. The impact of training on the modeling results will be discussed in the next section.

5.4.2 The model of neuronal IPD processing

In Sec. 5.3, the TC-model that was introduced in the previous chapter was used to predict the determined JNDs values. Requiring only one additional parameter, the threshold distance d_{thr}, the TC-model was able to predict JNDs at all tested frequencies while also capturing the frequency dependent increase of JNDs at different RIPDs (see Fig. 5.2c)). The determined values of d_{thr} as listed in Tbl. 5.1 changed only little with frequency with the only considerable deviation at 1 kHz. This is also visible from the boxplots shown in Fig. 5.2b). One possible explanation for this result in the context of the TC-model could be a small deviation in the TC-model parameter $\Delta\varphi_{\mathrm{best}}$

leading to a distorted representation of IPDs and as a consequence also to a deviation in d_{thr}. $\Delta\varphi_{\text{best}}$ values were determined by fitting the TC-model to the results of a left-right task (LRT) experiment. As noted in Sec. 4.3, some subjects had problems to solve the task of this experiment at 1 kHz which could have influenced the determined $\Delta\varphi_{\text{best}}$. A second interpretation of the reduced d_{thr} at 1 kHz would be a reduction in the internal noise at this frequency. The assumption that the separability between two IPDs can instead be represented by the value d as defined in Eq. (4.11) is based on the condition of equal internal noise. If this condition is not fulfilled, the separability can be determined by calculating d' (Green, 1960). Following signal detection theory, d' values can be calculated from the percent correct score of the experiment using the equation $d' = z(P_{\text{C}})\sqrt{2}$ where $z(P_{\text{C}})$ is the inverse of the cumulative distribution function of the standard normal distribution evaluated at the percent correct score P_{C}. As discussed in Sec. 4.5.1, d' can also be calculated from the distance d as defined in Eq. (4.11). Combining the two approaches at threshold ($P_{\text{C}} = 0.75$) results in the equation

$$d'_{thr} = z(0.75)\sqrt{2} = \frac{d_{\text{thr}}}{\sigma_{\text{jnd}}}, \tag{5.5}$$

where σ_{jnd} is the standard deviation due to internal noise. This internal noise can be calculated as

$$\sigma_{\text{jnd}} = \frac{d_{thr}}{z(0.75)\sqrt{2}} \approx d_{thr}. \tag{5.6}$$

Where $z(0.75)\sqrt{2} \approx 1$ so that d_{thr} can also be interpreted as an approximation of the internal noise σ_{jnd}.

The value of σ_{jnd} can not be assumed to equal σ_C, the noise of the modeled neuronal representation as defined in Eq. 4.6. Instead σ_{jnd} also includes contributions of all other sources of noise. In the context of the TC-model, σ_{jnd} can be decomposed into

$$\sigma_{\text{jnd}} = \sqrt{\sigma_C^2 + \sigma_{res}^2} \tag{5.7}$$

where σ_{res} combines contribution of all other noise sources such as those introduced due to the experimental procedure. As such, if the assumption of constant σ_C is retained, the observed decrease in d_{thr} could also be interpreted as a decrease in σ_{res} possibly due to a change in training or concentration of the subjects at this particular frequency. In the previous chapter the parameter σ_{fit}, an approximation of $\sigma_\Delta = \sqrt{2\sigma_C^2}$, was determined with a value of 0.28. Calculating an approximation of σ_C from this value results in $\sigma_C \approx 0.19$ which is close to the d_{thr} values determined in this experiment for frequencies up to 500 Hz.

As aforementioned, most previous studies reported considerably lower JNDs than the values found in the presented experiment which was justified with the fact that the

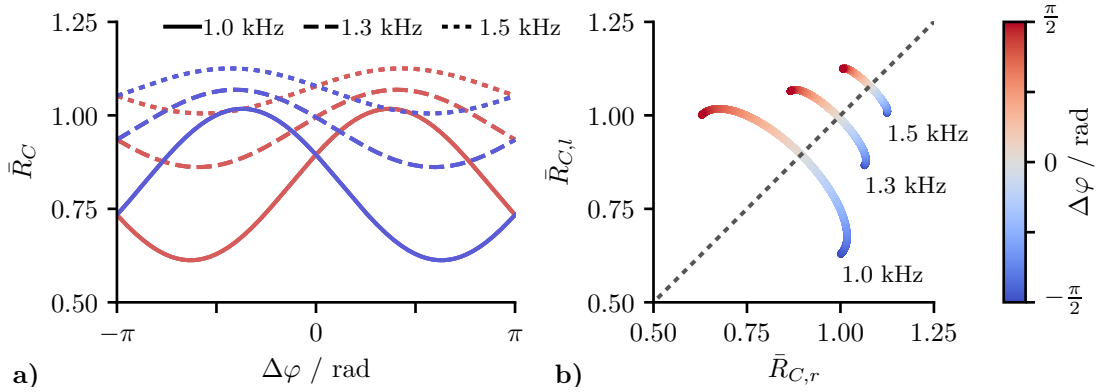

Figure 5.5: **a)** IPD-response functions of the TC-model model following the corrected parameter approximation show a decline of modulation. **b)** Response-plane trajectories of $\bar{\mathbf{R}}$ for IPDs between $-\pi/2$ and $\pi/2$ at different frequencies close to the high frequency limit.

subjects used in this study did not receive any training beyond being able to familiarize them self with the experimental procedure. Interestingly, the data of Brughera et al. (2013) could be explained by the TC-model solely by reducing d_{thr}. This could again be explained by assuming constant σ_c with different values of σ_{res}, where σ_{res} describes the influence of training and the experimental condition.

Multiple studies previously reported a steep increase in the JND at frequencies around 1500 Hz (Zwislocki and Feldman, 1956; Brughera et al., 2013). The TC-model also predicted the same increase if a linear dependency between $\Delta\varphi_{\text{best}}$ and frequency as well as a constant d_{thr} were assumed (see Sec. 5.3.1). The increase in JND can be explained when looking at the model responses at different frequencies. With increasing frequency, there is also an increase in the width of the channel-response $\bar{R}_{\text{C}}(\Delta\varphi_{\text{best}})$ functions defined in Eq. 4.2. At higher frequencies, this increase in width leads to a reduction in the modulation of $\bar{R}_{\text{C}}(\Delta\varphi_{\text{best}})$ (see Fig. 5.5a). As a result of this reduction, a larger IPD is necessary to reach d_{thr}. This relation can also be visualized in the response-plane of the TC-model which is created by plotting the response of the left channel $\bar{R}_{\text{C,r}}$ against the response of the right channel $\bar{R}_{\text{C,r}}$ at different IPDs. Figure 5.5b shows the response plane for the three frequencies (1 kHz, 1.3 kHz and 1.5 kHz). From this plot, it is visible that the trajectory in the response plane is compressed with increasing frequency so that an increase in JND is necessary to result in the same distance. In agreement with this discussion, a biophysical model used by Brughera et al. (2013) to explain the high-frequency limit did also show a strong loss in modulation with increasing frequency. This is also consistent with the biophysical model introduced in Chapter 3 which is clearly visible in Fig 5.6a), where the IPD-response functions of the model are shown for different frequencies. Fig 5.6b) shows the corresponding response planes. While both biophysical models, the one introduced in this thesis as well as the model of Brughera et al. (2013), also

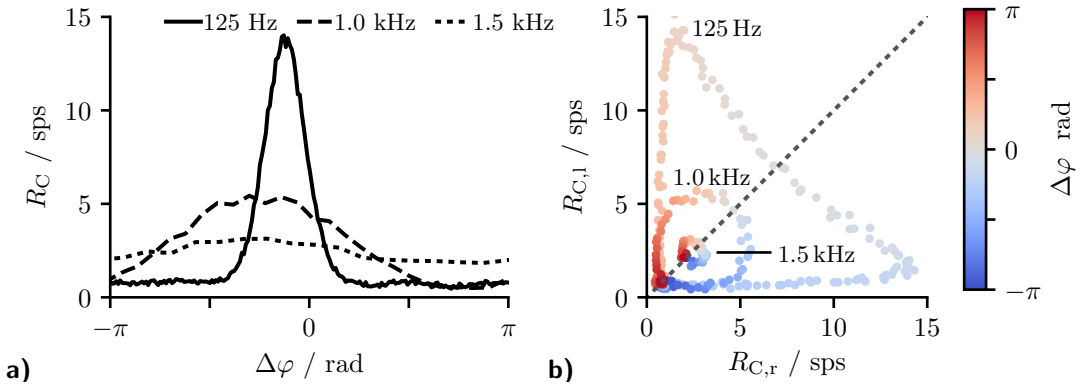

Figure 5.6: a) IPD-response functions of the biophysical model for three frequencies. Similar to the prediction of the TC-model-model, these IPD response functions also show reduced modulation at higher frequencies. **b)** The response plane trajectories for the responses shown in a).

showed an increasing base firing rate with frequency (minimum response in Fig. 5.5a), this increase was much smaller than in the TC-model. The stronger increase in the model is due to the normalization of the TC-model model. In the biophysical models both excitation as well as inhibition interact to result in different absolute responses depending on frequency (See Fig. 3.3) while the response of the TC-model model is normalized. As the TC-model is based on the relative location of different $\bar{\mathbf{R}}$ on the response plane, its predictions are not influenced by a response offset so that the increase in the base rate of the TC-model does not influence the results.

5.5 Summary and Conclusion

This chapter applied the TC-model that was introduced in the previous chapter to predict IPD difference thresholds for tonal stimuli. JNDs where determined at different RIPDs by conducting a psychoacoustic experiment on ten subjects. The main findings were:

- Within the tested region of RIPDs, JNDs increase with increasing RIPDs. This increase was found to be more pronounced at lower frequencies.
- Applying the TC-model with the same parameters determined in the previous chapter and using the additional parameter d_{thr}, predicted JNDs at all measured RIPDs and frequencies.
- The TC-model model is also able to predict JND results of previous studies when d_{thr} is adjustable to account for effects of training. Doing so, the TC-model also predicts the high-frequency limit of IPD sensitivity.

CHAPTER 6

Considering Wideband Signals

6.1 Introduction

In the two previous chapters, the two-channel model (TC-model) was only applied to tones and did not address how to handle signals with multiple frequency components. The following chapter introduces a basic concept of how the TC-model could be applied to wideband signals.

Most of the commonly used binaural models include model stages that account for the auditory periphery. An often used approach is to use a filterbank to split the signal into several frequency channels mimicking the mechanical bandpass filtering of the basilar membrane followed by additional stages that account for other processing steps such as the mechano-electrical transduction of the outer hair cells. The binaural stage of these model is then commonly applied separately to each frequency channel (e.g. Colburn, 1973; Dietz et al., 2009; Takanen et al., 2014). As an example, most models based on the delay-line approach represent the results in the form of a cross-correlogram. The cross-correlogram can be envisioned as a matrix where the rows represent the frequency channels and the columns the delays along the delay line (Colburn, 1973; Trahiotis and Stern, 1989; Stern et al., 2006). The values in this matrix represent the rate of the coincidence detectors at a specific frequency and delay. Interactions between the frequency channels can then be implemented by applying functions to this matrix.

In contrast to the biophysical model proposed in Sec. 3, the TC-model does not account for the processing within the auditory periphery so that it is only possible to discuss the basic concept of how to apply the model to non-tonal stimuli. To do so, this chapter will assume that the signal within each frequency channel can be described by an equivalent unmodulated tone with a frequency that equals the channel's center frequency and that the interaural differences can be described by one interaural phase difference (IPD) for each channel.

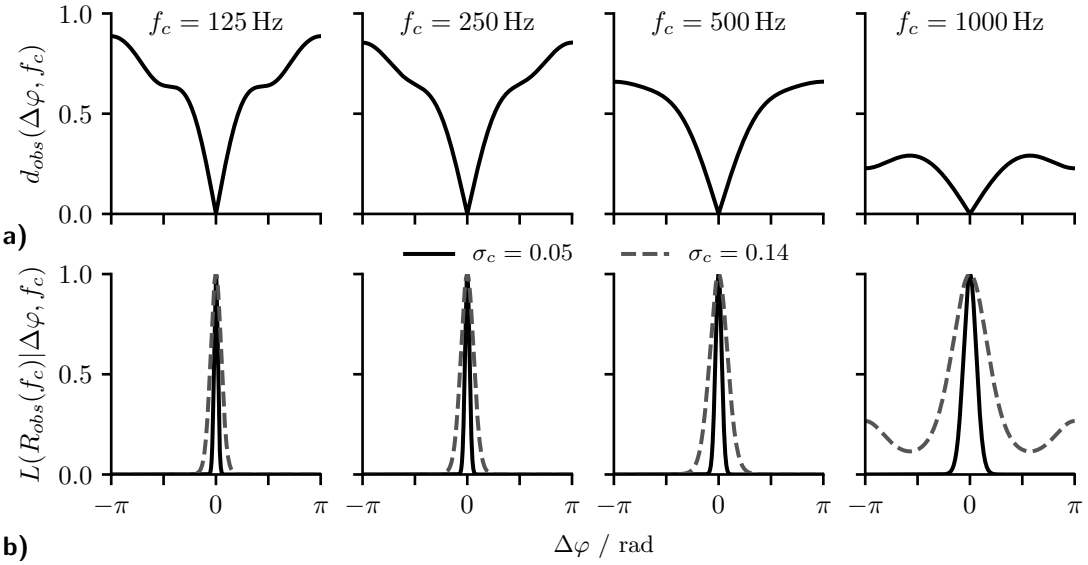

Figure 6.1: a) Distances between the observed responses $\mathbf{R}_{obs}(f_\mathrm{C})$ and the mean responses at different IPDs $\bar{\mathbf{R}}(\Delta\varphi, f_\mathrm{C})$ calculated for the case $\mathbf{R}_{obs}(f_\mathrm{C}) = \bar{\mathbf{R}}(0, f_\mathrm{C})$ **b)** Relative likelihood functions calculated for the distances shown in **a)**. Values were calculated for channel noises $\sigma_\mathrm{C} = 0.05$ and $\sigma_\mathrm{C} = 0.14$.

6.2 Calculating the Relative Likelihoods Functions

Assuming an arbitrary number of frequency channels, each defined by its center frequency f_C, the presentation of a signal would result in one observed neuronal response $\mathbf{R}_{obs}(f_\mathrm{C})$ per frequency channel. Using Eq. 4.11 it is possible to calculate the distances d_{obs} between this observed response and the expected mean response due to any $\Delta\varphi$ at this frequency:

$$d_{obs}(f_\mathrm{C}, \Delta\varphi) = ||\bar{\mathbf{R}}(\Delta\varphi, f_\mathrm{C}) - \mathbf{R}_{obs}(f_\mathrm{C})||. \tag{6.1}$$

Here, $\bar{\mathbf{R}}(\Delta\varphi, f_\mathrm{C})$ is the mean response vector at an IPD of $\Delta\varphi$ calculated for the frequency f_C. Figure 6.1a shows d_{obs} calculated at four frequencies assuming $\mathbf{R}_{obs}(f_\mathrm{C}) = \bar{\mathbf{R}}(0, f_\mathrm{C})$ so that $d_{obs} = 0$ at an IPD of zero. Due to the normal distribution of responses around any $\bar{\mathbf{R}}(\Delta\varphi, f_\mathrm{C})$ (See Eq. 4.6), the distribution of d_{obs} can be stated as

$$d_{obs} \sim \mathcal{N}(0, {\sigma_\mathrm{C}}^2), \tag{6.2}$$

where $\mathcal{N}(0, {\sigma_\mathrm{C}}^2)$ denotes the normal distribution around zero with the standard deviation of the channel responses ${\sigma_\mathrm{C}}^2$. Combining Eq. 6.1 and Eq. 6.2 enables to calculate the relative likelihood L to observe $\mathbf{R}_{obs}(f_\mathrm{C})$ for any given $\Delta\varphi$ as

$$L(\mathbf{R}_{obs}(f_\mathrm{C})|f_\mathrm{C}, \Delta\varphi) = \exp\left(\frac{-d_{obs}(f_\mathrm{C}, \Delta\varphi)^2}{2{\sigma_\mathrm{C}}^2}\right). \tag{6.3}$$

Eq. 6.3 was normalized to result in a relative likelihood of one if d_{obs} equals to zero so that all other likelihoods can be interpreted relative to this value. As the name implies, the relative likelihood functions (L-functions) can be interpreted as a measure for the likelihood that an observed response $\mathbf{R}_{obs}$ encodes a certain IPD $\Delta\varphi$. A maximum likelihood decoder, as an example, would use the location of the peak of the L-function along the IPD axis as its prediction.

Figure 6.1b shows the L-functions calculated for the distances shown in Fig.6.1a. The functions are shown for a channel noise with the standard deviation $\sigma_{\mathrm{C}} = 0.05$ as determined for the data of Brughera et al. (2013) and $\sigma_{\mathrm{C}} = 0.14$ as determined for the experimental data in Sec. 5.3.1. As expected, all L-functions are maximal at the IPD resulting in the smallest d_{obs}. It is also visible that an increase in frequency results in an increase in the width of the L-functions peak which is in agreement with the increase in IPD just noticeable differences (JNDs) as presented in Sec. 5.3. A narrow peak indicates that the observed response is encoded with high acuity. This is also visible when comparing the L-functions for the two different σ_{C}. An increase in channel noise leads to an increase in the width of the function's peak, and thus a decreases in acuity (Compare Fig. 5.3). It is, in fact, possible to directly relate the L-functions shown in Fig.6.1 to the difference threshold. Following Eq. 5.5 and Eq. 5.6, the difference threshold of 75% correct responses is reached if the distance between two responses approximately equals the standard deviation of the noise. From Eq. 6.3 follows that this is the case if $L = \exp(-0.5) = 0.6$.

In the case of multiple frequency channels, one L-function is calculated for each channel. Figure 6.2a shows these functions for the frequencies between 100 Hz and 1 kHz calculated for a signal with a frequency independent IPD. As in the previous visualization, $\mathbf{R}_{obs}(f_{\mathrm{C}}) = \bar{\mathbf{R}}(0, f_{\mathrm{C}})$ was assumed. The L-functions are shown over an IPD range from -2.5π to 2.5π to visualize the 2π periodicity of the likelihood functions which is due to the underlying IPD periodicity. In these plots, the underlying IPDs of zero and 0.3π are clearly visible as the location of the L-functions peak. The acuity with which the IPD is encoded can be estimated from the width of this peak.

In a natural setting, a signal with a frequency independent IPD as shown in Fig. 6.2a is of little relevance. As discussed in Sec. 2.1, a given sound source location results in an approximately frequency independent ITD which is equivalent to a frequency dependent IPD. As a consequence, for signals with multiple frequency components it is more useful to present the relative likelihoods as a function of ITD instead of IPD. Figure 6.2b shows some examples for signals where an ITD was introduced in all frequency components. The first column in this figure shows two examples of ITDs within the physiological range of humans. Due to the $\frac{1}{f}$ relationship between ITD and IPD, these functions are wider at lower frequencies than at higher frequencies. This is in agreement with the fact that ITD difference thresholds decrease with frequency while IPD difference thresholds increase with frequency (discussed in Sec. 2.3). Interestingly, the L-functions shown in Figure 6.2 show a very similar structure as the cross-correlogram that would

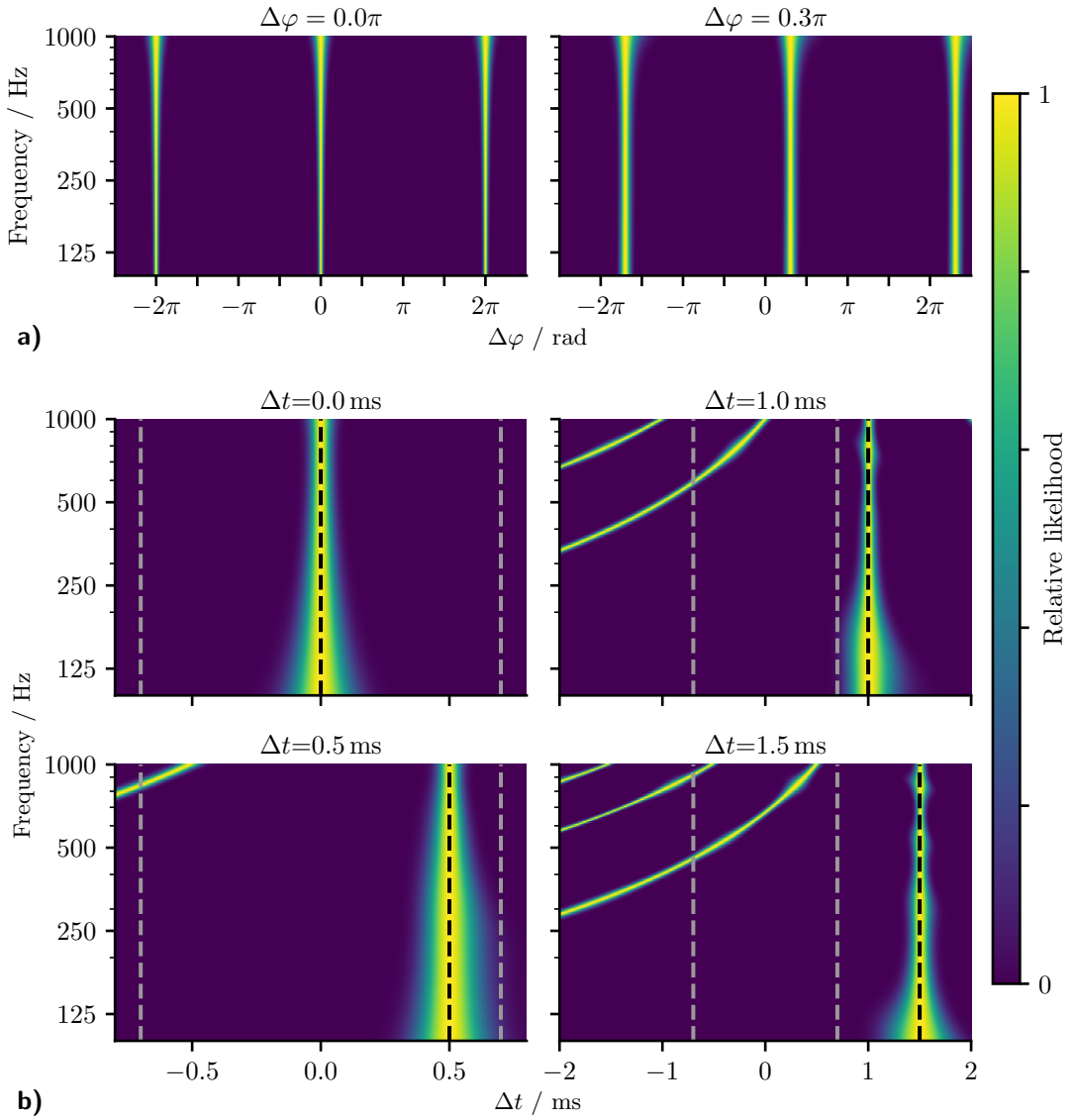

Figure 6.2: a) L-functions calculated for signals signal with the frequency independent IPDs $\Delta\varphi = 0$ and $\Delta\varphi = 0.3\pi$. In agreement with the increase of JND with increasing frequency, these functions also show an increase in the width of their peak. **b)** L-functions calculated for different ITDs. The functions are also shown over of ITD instead of IPD. The two dashed gray lines indicate approximations for the borders of the natural ITD range and the dashed black line indicates the stimulus ITD

result from applying a delay-line model to a delayed noise signal. This qualitative agreement between the two representations makes it possible to transfer some ideas that have been developed to explain cross-frequency processing in delay-line models to the TC-model. One prominent example is the concept of straightness as introduced by Trahiotis and Stern (1989). The right column of Fig. 6.2b shows L-functions calculated for ITDs outside of the physiological range. As a result of the large ITD, the main peak of the L-function is located outside of the physiological range (indicated by the dashed gray lines), but some of the secondary peaks, which exist due to the aforementioned periodicity, are located inside of the physiological range. The difference between the primary and the secondary peaks being that the arrangement of secondary peaks over frequency is curved while the location of the primary peaks is straight. The idea of the straightness concept following Trahiotis and Stern (1989) is to weight the peaks by their straightness over frequency. This approach would eliminate the curved secondary response leaving only the primary peak. Trahiotis and Stern (1989) also showed that the straightness concept is of behavioral relevance. In their study, they used bandpass filtered noise centered around 500 Hz at an ITD of 1.5 ms (see bottom right of Fig. 6.2b) and varied the signal bandwidth. At a narrow bandwidth, the subjects perceived the stimulus to be lateralized to the left while they perceived it to the right when the bandwidth was increased. These findings were explained by the fact that no or only little straightness weighting is possible in the case of the narrow bandwidth noise so that the hearing sensation was based on the secondary peak, which indicated a negative ITD. With increasing bandwidth, straightness weighting becomes effective, and the hearing sensation was lateralized based on the primary peak which indicates a positive ITD. A similar concept could easily be applied to the TC-model for example by averaging over the L-functions of multiple adjacent frequency channels.

Up to this point, all L-functions were calculated based on the constrained $\mathbf{R}_{obs}(f_\mathrm{C}) = \bar{\mathbf{R}}(0, f_\mathrm{C})$ which fixes the observed response to equal the mean response. While this is helpful for discussing the concept of the L-function, it also removes the influence of channel noise on the observed responses. Figure 6.3 shows L-functions calculated for 20 log-spaced frequency channels between 100 Hz and 1 kHz, where each $\mathbf{R}_{obs}(f_\mathrm{C})$ was randomly drawn from the probability distribution as defined in Eq. 4.6. L-functions are shown for a stimulus with an ITD of 0.3 ms ($\bar{\mathbf{R}}(\Delta\varphi = 2\pi f_\mathrm{C} \cdot 0.3\,\mathrm{ms}, f_\mathrm{C})$) and for two different channel noises with the standard deviations $\sigma_\mathrm{C} = 0.05$ and $\sigma_\mathrm{C} = 0.14$. As a result of introducing noise to the observed responses, there is also a random variation in the locations of the L-function peak along the ITD axis, which is visible in Fig. 6.3a. The location of these peaks along the ITD axis changes randomly following a normal distribution centered around the stimulus ITD. The standard deviation of the distribution within each frequency channel is defined by the channel noise. As a consequence of this distribution of the peaks, the variance in their position is in agreement with results of both, the left-right experiment of Chapter 4 and the JND experiment of Chapter 5. The influence of channel noise is visible in Fig. 6.3b, which

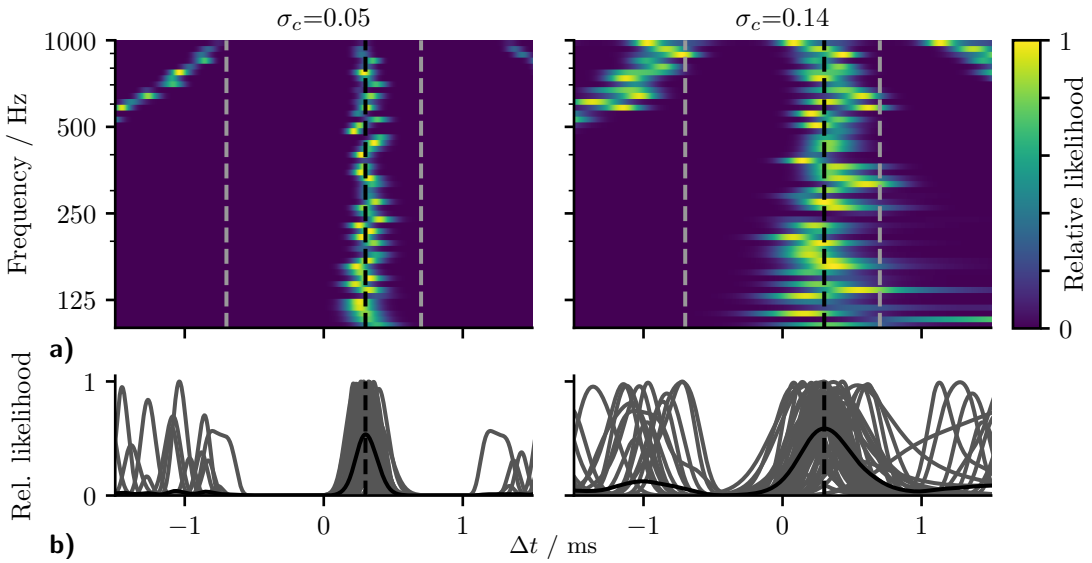

Figure 6.3: a) L-functions calculated for a signal with an ITD of 0.3 ms. The functions were calculated for $\mathbf{R}_{obs}(f_C)$ which were randomly drawn from a probability distribution as defined in Eq. 4.6.The two dashed gray lines indicate approximations for the borders of the natural ITD range and the dashed black line indicates the 0.3 ms ITD **b)** The superimposed L-functions of all frequency shown in b) (gray) as well as the mean over all functions (black). The dashed line indicates the 0.3 ms ITD.

shows the superimposed L-functions of all frequency channels (gray lines). An increase in channel noise leads to a larger variation of the positions of the L-function but also increases the width of the peaks, which accounts for the reduced acuity. The black curve in Fig. 6.3b shows the mean over the L-functions over all frequency channels, which is a simple way of applying the straightness concept. The result shows that the ITD is unambiguously encoded when combining the likelihood across frequency.

6.3 Summary and Conclusion

In the two previous chapters, the TC-model was discussed only in the context of tones without addressing how to handle stimuli with multiple frequency components. As the model in its current form lacks a description of the auditory periphery, this chapter is limited to a basic concept and abstains from comparing the model with psychoacoustic data. To visualize the acuity of the encoded IPD in a single frequency channel, this chapter derived the L-function, which can be interpreted as the likelihood that an observed response was caused by a certain IPD. When using the L-function, it is, however, important to note that any visualization where the function is shown over ITD instead of IPD might be misleading. The TC-model only represents ITDs in a

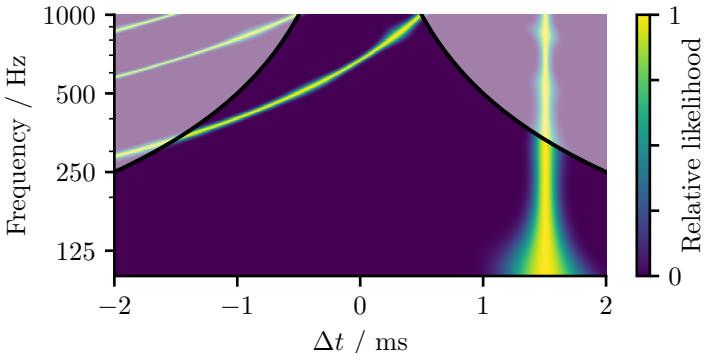

Figure 6.4: L-function calculated for an ITD of 1.5 ms (same as bottom right of Fig. 6.2b). The lighter areas visualize the region of ITDs that are not directly encoded in the TC-model but inferred by making use of the L-functions' periodicity.

range that is equivalent to $\pm\pi$ with the values at other ITDs being inferred from this range by making use of the L-functions' periodicity. This is visualized in Fig. 6.4, which shows the L-functions for the 1.5 ms ITD case but highlights the $\pm\pi$ region that is directly represented in the TC-model.

Interestingly, visualizing the L-functions over ITD, results in a representation very similar to the cross-correlogram which is often used to interpret the outcome of delay-line models. This agreement opens up the possibility to reuse some of the approaches that were introduced to explain across-frequency phenomena in delay-line models, such as the straightness concept. However, if the brain encodes IPDs, these would have to be remapped in a frequency specific manner.

Concluding, the L-function concept presented in this chapter seems to be a viable method of representing signals with multiple frequency components in the TC-model. For a more informative analysis, however, it is necessary to also account for the influence of phase and amplitude modulations within single frequency channels as they might arise when using complex signals and a realistic description of the auditory periphery.

CHAPTER 7

On the Influence of Low Frequency ILDs

The content of this chapter was previously published as a peer-reviewed article[i]:

Encke, J., Reimann, D., Hemmert, W., and Völk, F. On the role of interaural level differences in low-frequency pure-tone lateralization. *Acta Acustica united with Acustica*, 104, 2018

The Article is reproduced here without content-related changes. Figures have been reformatted to fit the format of this document but are otherwise unchanged.

7.1 Introduction

In his influential 1907-paper, Lord Rayleigh introduced the basis for what has come to be known as the "Duplex Theory" of sound localization (Rayleigh, 1907). Essentially, the duplex theory states that low-frequency sounds are localized predominantly based on interaural phase differences (IPDs), as the shadowing effect of the head only results in small interaural level differences (ILDs) in that frequency range. High-frequency sounds, on the other hand, are mainly localized using ILDs, as the impact of the head shadowing effect on ILDs increases with frequency. Additionally, IPDs become ambiguous at about 700 Hz, where the stimulus wavelength approximately equals the head circumference. Interestingly, the auditory system is able to detect IPDs past the ambiguity limit up to about 1300 Hz (Brughera et al., 2013), and the sensitivity to ILDs at low frequencies was found to be similar to the ILD sensitivity at high frequencies (Yost and Dye, 1988). These findings suggest an interaction between ILDs and IPDs in the low-frequency range. The nature of this interaction has been studied extensively (Sayers, 1964; Domnitz and Colburn, 1977) and, depending on their sign, ILDs were found to facilitate or counteract the lateralization evoked by IPDs. Due to the small magnitude of naturally occurring low-frequency ILDs in free-field conditions, their influence is often neglected. At the same time, additional cues are necessary in

order to resolve IPD ambiguities at sound frequencies above some 700 Hz in free-field conditions – the ILD being a potential and well-suited candidate. To our knowledge, no study has yet systematically investigated the resolving impact of ILDs on lateralization, especially of small ILDs as they naturally occur in free-field conditions or various natural listening environments. To that end, this study uses a pure-tone lateralization paradigm to investigate the influence of small ILDs on the lateralization evoked by IPDs. Aspects of the practical relevance of the results are discussed in the light of typical free-field ILDs.

7.2 Methods

In order to evaluate a potential interaction between ILDs and IPDs, we conduced a left-right task (LRT) experiment (Völk et al., 2017b). Tone impulses (700 ms, including 160 ms Gaussian onset and offset slopes) were presented dichotically at one-octave spaced frequencies between 125 Hz and 1 kHz, with IPDs introduced by shifting the starting phase of the tone at one ear relative to the other, without changing the stimulus envelope. The sound-pressure level (SPL) in the diotic condition at 125 Hz was set to 72 dB SPL, with the headphones calibrated based on their nominal sensitivity at 1 kHz. The levels at all other frequencies were determined based on the equal-loudness contours according to ISO 226:2006, with the aim of approximately balancing loudness across the conditions. ILDs were then introduced by amplifying the left and right signals by half the desired ILD, respectively. 21 evenly spaced IPDs between $-\pi$ and π were combined with the ILDs 0 dB, ± 0.5 dB, ± 1 dB, ± 1.5 dB, ± 2 dB, ± 4 dB and ± 8 dB. At frequencies below 1 kHz, each IPD was combined with only the ILDs with the sign of the IPD, while at 1 kHz, all IPDs were combined with all ILDs. Each stimulus condition was presented and evaluated ten times in random order. The results were collected in four sessions per subject (one session per frequency), each lasting on average 48 minutes, with a short break after half the stimuli.

Ten unpaid, naïve with respect to the experiment, and inexperienced volunteers (23 years to 41 years, median 27 years, all right handed) with no reported hearing deficits and absolute thresholds within ± 15 dB of ISO 226:2006 took part in the experiment. The subjects were seated in a sound-insulating booth and were presented with the stimuli using the same pair of circumaural headphones (HD 650; Sennheiser electronic GmbH, Wedemark). The specific headphones were selected due to their low inter-individual and intra-individual variability based on (Völk, 2014). They were interaurally equalized with respect to amplitude and phase (Völk et al., 2017a), so that the relevant sound-pressure time functions measured on a coupler according to IEC 60318-1:2009 were identical within the accuracy of the measurement procedure. The subjects were asked to sit upright, face a specific wall and indicate, by pushing one of two buttons, whether the hearing sensation (HS) occurred "left or right", with no further explanation (Völk et al., 2017b). Prior to the experiment and after reading

the written instructions, a basic understanding of the instructions was verified within a short example experiment. Answers to potential procedural questions were limited to "yes" or "no", and the example was repeated until the subject felt comfortable with the procedure. No further instructions were provided.

7.2.1 Analysis of IPDs and ILDs

For approximating the typical magnitudes of naturally occurring free-field IPDs and ILDs in human listeners, free-field binaural transfer function pairs of 139 subjects from the freely available database of the acoustics research institute (ARI) (Majdak et al., 2010) were analyzed. The transfer functions were estimated by individually calculating the discrete Fourier transforms (DFTs) of the finite impulse responses corresponding to the left and right ears that reflect the geometric arrangement of interest (variation of the sound-incidence azimuth in the horizontal plane at a fixed source distance of 1.2 m). The IPDs and ILDs for each position were estimated by subtracting the corresponding phase and level values of specific single DFT bins. The results are depicted in terms of the arithmetic mean values over the dataset in Figure 7.1a. Additionally, Figure 7.1b shows the frequency-dependent maximum values over all included source positions. The dashed lines indicate a possible descriptive approximations, where IPDs corresponding to 700 µs delay appear to fit the data frequency-independently well. Consequently, IPDs representing ±700 µs are taken as an approximation of the range of naturally-occurring free-field IPDs (abbrev.: natural range). Regarding ILDs, frequencies below 500 Hz result in maximum ILDs around 4.5 dB, while at frequencies of 500 Hz and above, the maximum ILD increases by about 6 dB/octave. As the results depend on measurement setup and transfer function definition, the data are considered an approximation of the actual values.

In order to evaluate whether the supportive effect of ILDs is relevant in natural settings, the average magnitudes of ILDs and IPDs, as they would occur in a typical free-field setup, were analyzed. The first two columns of Figure 7.1a show interindividually averaged IPDs and ILDs as functions of azimuth for different stimulus frequencies. For the three lower frequencies, both ILD and IPD show an approximately sinusoidal dependence on azimuth, with the maximum IPD below π. At 937.5 Hz, the maximum IPD is about $1.4\,\pi$, and IPD ambiguity is visible, as multiple azimuths correspond to the same IPD. The third column of Figure 7.1a shows ILDs plotted against the corresponding IPDs, for each azimuth. This representation visualizes that IPD ambiguities can be resolved by ILDs.

7.3 Results and Discussion

Figure 7.2 exemplarily shows the inter-individual medians of the LRT-experiment results for 500 Hz (left) and 1 kHz (right), represented as the fraction of stimuli rated

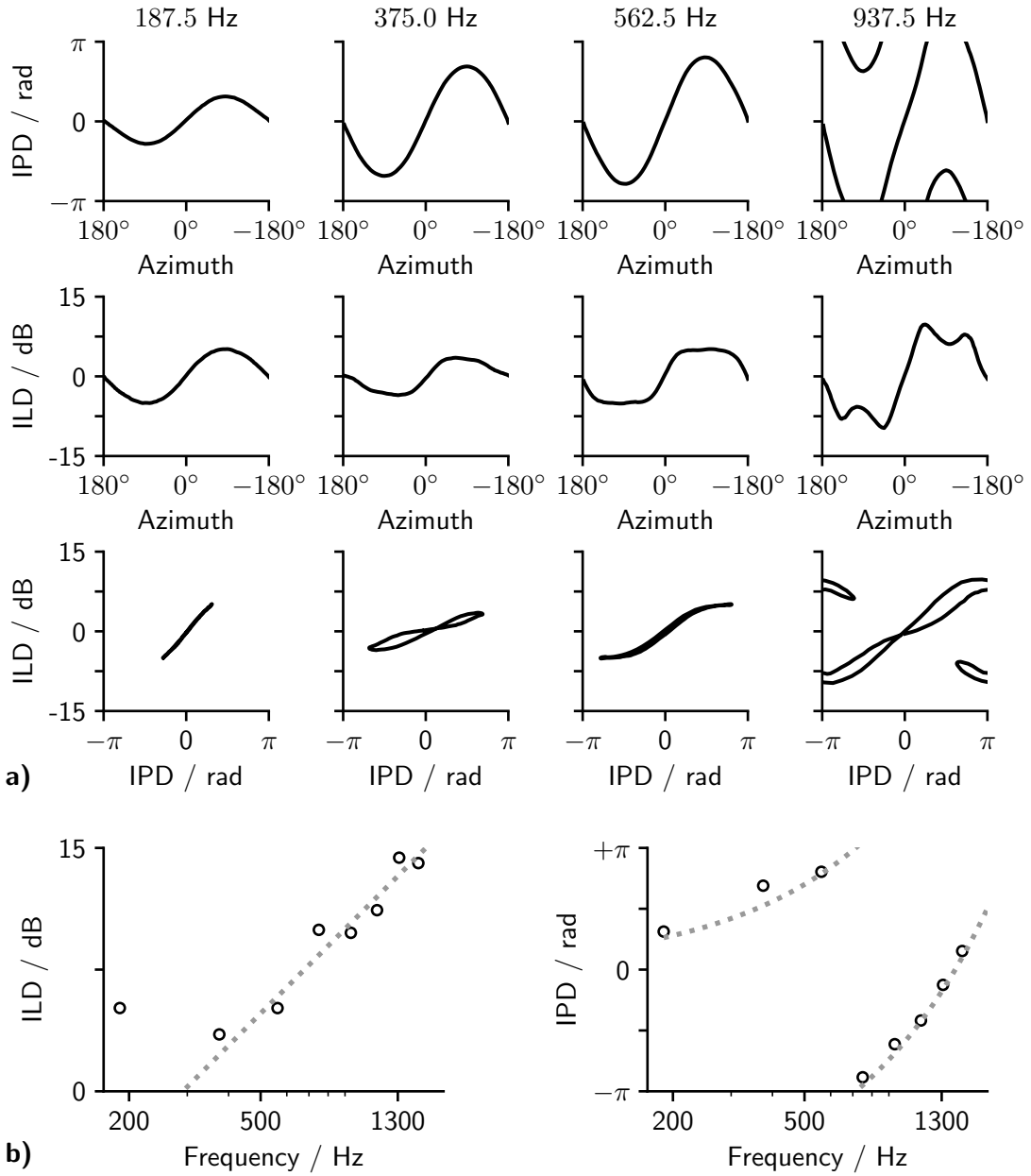

Figure 7.1: a) Average horizontal-plane IPDs (first column) and ILDs (second column), estimated from 139 binaural impulse response pairs of the ARI database (Majdak et al., 2010). The third column shows each ILD plotted against the corresponding IPD. **b)** Maximum estimated horizontal-plane ILDs (left) and IPDs (right) as functions of frequency. The dashed contours represent descriptive approximations: line with 6 dB/oct for maximum high-frequency ILDs (left) and IPDs corresponding to 700 µs delay for maximum IPDs (right).

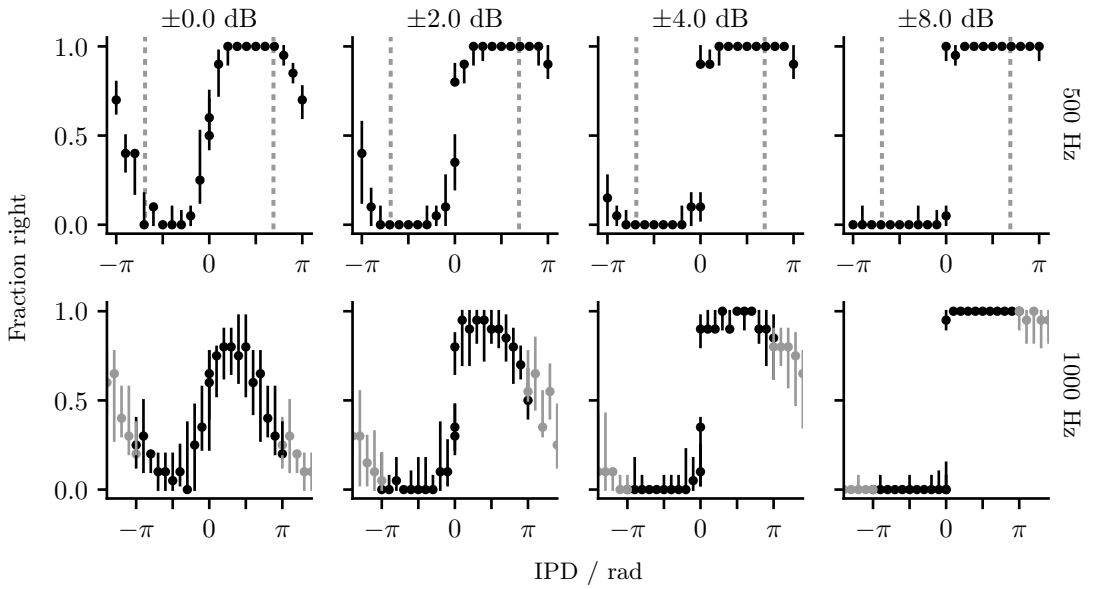

Figure 7.2: Results of the LRT experiment for 500 Hz and 1 kHz. Shown are the interindividual medians with 25% and 75% percentiles for all IPDs (abscissae), combined with selected ILDs (rows). The vertical dashed lines indicate the IPDs equivalent to ±700 µs. Duplicate IPDs (i. e. $|\mathrm{IPD}| \geq \pi$) are shown lighter.

on the right. As found in previous experiments (Völk et al., 2017b), the results for the two frequencies below 500 Hz only showed minor differences to those obtained at 500 Hz so that the results will be discussed based on the two frequencies shown in Figure 7.2. The columns in this figure represent the frequencies of the tone impulses, the rows indicate selected ILD conditions (label on the right). At 0 IPD and 0 ILD (horizontal centers, first row), the subject group responded at chance level. With increasing IPD (at 0 ILD; towards the right), the fraction of HS reported on the right increases, until saturating at 100 %. At large IPDs, the fraction of responses to the right starts to decrease again, likely due to an ambiguity introduced by IPD wrapping. In the 500 Hz condition, the ambiguity effect starts to manifest itself in the data approximately at the limits of the natural range (dashed vertical lines in Figure 7.2). At 1000 Hz, however, 0.7 ms corresponds to an IPD of approximately $1.4\,\pi$ (abscissa limits in Figure 7.2), resulting, according to Figure 7.2, more or less in a left/right reversal of the LRT results. This appears plausible, as an IPD of $1.4\,\pi$ is identical to $-0.6\,\pi$. Opposed to the 500 Hz conditions, ambiguity effects occur clearly within the natural range. As the ambiguity effects start to be visible at IPDs at the limits of the natural range at 500 Hz and lie inside this range at 1 kHz, it appears reasonable to assume that ambiguity effects within the natural range become perceptually relevant at frequencies of 500 Hz and above, with an amount that increases with frequency.

Each row in Figure 7.2 shows the results for a selected ILD condition. The sign of the added ILD is always identical to the sign of the IPD (e.g −2 dB at $-\pi$), so that

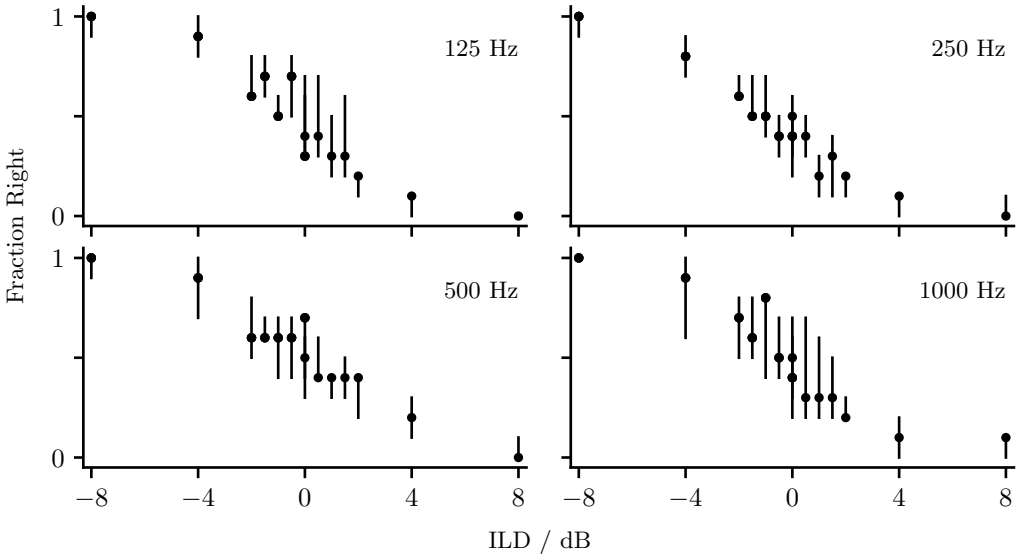

Figure 7.3: Results from the experiment of Figure 7.2, depicted as functions of ILD at 0 IPD.

mostly non-conflicting cues were presented. As expected from literature (Domnitz and Colburn, 1977), the addition of those ILDs mostly supports the lateralization due to the IPD (rows 2, 3, and 4 of Figure 7.2). At 2 dB, this effect is already clearly visible, and 8 dB is sufficient to more or less fully resolve the ambiguity effects with the intermediate ILD values (not shown) following this trend. Analyzing the magnitudes of the ILDs calculated from the free-field transfer functions at 937.5 Hz, the maximum ILD was found to be 9.8 dB, which, compared to the experimental data should be sufficient to fully resolve ambiguities.

The data in Figure 7.2 suggests that small ILDs may already influence the LRT results, at least at certain azimuths that are not affected by ILD-ambiguity effects (e.g. IPDs between approx. $\pm 0.5\pi$ in Figure 7.2), some combinations of which are naturally relevant IPD and ILD combinations (compare to data in Figure 7.1a. In the experimental paradigm used in this study, the influence of ILDs can only be examined where the LRT results are not saturated due to the IPD alone.

In order to address the effect of small ILDs in more detail, an analysis was conducted for the zero-IPD condition only. The results for this condition are shown in Figure 7.3. A two-way analysis of variance (repeated measurement ANOVA) was conducted over the LRT results with the two factors ILD and frequency. All thirteen ILDs and four frequencies were included. The effect of frequency yielded $[F(3, 27) = 1.07,\ p = 0.38]$, indicating no significant main effect of the factor frequency. This finding is in agreement with previous studies that report no influence of frequency on the sensitivity to ILDs (Yost and Dye, 1988). The factor ILD, on the other hand, shows a highly significant

main effect [$F(12, 108) = 83.20,\ p < 0.0001$], with no significant interaction with the factor frequency [$F(36, 324) = 1.11,\ p = 0.32$]. A post-hoc comparison according to Tukey reveals significant ($p < 0.05$) differences to occur first between the zero ILD condition and ILDs of ± 1.5 dB, with a non-significant comparable trend for smaller ILDs.

The main interest of this study was the impact of small ILDs on the lateralization of tone impulses. The experimental results showed that the exclusive use of IPDs results in ambiguity effects at frequencies of 500 Hz and 1 kHz. It was also shown that the addition of small ILDs of some 4 dB largely resolved these ambiguities. This finding does not directly support the conclusion that, in the low-frequency range, IPDs dominate ILDs (Wightman and Kistler, 1992). In the 1 kHz, $1.4\,\pi$ condition, the IPD clearly indicates a sound source in the left hemifield. Once an ILD of realistic magnitude is added, this "inversion" is resolved, resulting in a HS lateralized to the right. This suggests that IPD and ILD are not separate "cues", but that HS properties arise based on most plausible combinations of the information extracted from the stimuli by various neuronal processing stages (Völk, 2013), rather then one arbitrarily-defined (physical) "cue" dominating the other.

7.4 Summary and Conclusions

In agreement with previous localization studies (Hartmann et al., 2016), this study shows that ILDs can resolve IPD ambiguity in lateralization at all studied frequencies, which becomes relevant in free-field settings starting at frequencies as low as 500 Hz, the same frequency as stated in the original publication on the "duplex theory of localization" by Lord Rayleigh (Rayleigh, 1907). It is important to note that, for natural stimuli such as speech, ILDs are not the only mechanism that could resolve ambiguities. In these cases, both low-frequency amplitude modulations and the evaluation across frequencies may also convey unambiguous information.

Using the same experimental data, this study also addresses whether ILDs affect the LRT result in parts of the horizontal plane and at frequencies that are not subject to IPD ambiguity. At zero IPD, ILDs as low as 1.5 dB cause a significant effect in the data presented here. ILDs of 4 dB were found to be sufficient to reliably attribute the HS to left or right, even without an IPD. This finding suggests that two sound sources in opposite hemifields can be attributed correctly based on ILDs only, as long as the source positions correspond to ILDs larger than ± 1.5 dB. ILDs of this magnitude occur for sources in the horizontal plane at all audible frequencies, even below 500 Hz. The latter conclusion appears relevant to the interpretation of various IPD-related localization experiments conducted in the free sound field or with virtual-acoustics techniques, as the present study indicates that such experiments may include effects of both IPDs and ILDs, even at low frequencies.

CHAPTER 8

Summary and Discussion

In this thesis, the processing of interaural phase differences (IPDs) in the mammalian brain was investigated by developing computational models of the underlying neuronal representation based on the two-channel concept. For these models, the process that connects a sound event to a perceived hearing sensation was described in three parts: the encoder that translates a physical signal into the neuronal representation of the signal, the neuronal representation itself, and the decoder that interprets the neuronal representation to result in a property of a perceived hearing sensation (see Fig. 1). The underlying assumption of the proposed models was that a decoder can only leverage information encoded in the neuronal code. As a consequence, perception should also be restricted by this information, so that the acuity and accuracy of perception should mirror the acuity and accuracy with which a physical cue is encoded in the neuronal representation.

Chapter 3 presented a detailed biophysical model of the medial superior olive (MSO), which is believed to be the encoder for IPDs in mammals. It encodes temporally precise phase differences between the left and right ear into a more robust rate code, which can be evaluated at higher brain stages without loss of precision. The model's structure, as well as the biophysical parameters, were based on current physiological knowledge about the function of the MSO. The model consisted of three parts: the auditory periphery, a population of globular bushy cells and a population of MSO neurons with each part implemented twice – once per brain hemisphere. In practice, the model took digitized ear-signals as the input and resulted in spike trains for each of the two MSO. These neuronal responses (spike trains) were interpreted following the two-channel hypothesis, and two different decoders were implemented to interpret the representation. The hemispheric-difference or linear-decoder was shown to be able to decode IPDs from the neuronal responses using a simple subtraction (see Sec. 3.2.2), but the decoder was found to be limited to a narrow range of IPDs around zero. Additionally, the output of the decoder was also strongly influenced by variations in the firing rate of the auditory nerve fiber (ANF) as they might arise due to adaptation or changes in sound level. To circumvent these problems, an artificial neuronal network (ANN) predictor was introduced (see Sec. 3.2.3). This predictor received additional inputs directly from the ANF and, as a consequence, was shown

to be able to predict IPDs also for complex signals by compensating for variations in the ANN input rate. These results suggest that a neuronal representation following the two-channel hypothesis can be implemented by the neuronal circuit of the MSO, and that the two-channel representation is sufficient to convey information about the interaural time difference (ITD) with high acuity. Due to the highly detailed nature of biophysical models and the resulting number of parameters, it is, however, computationally expensive to use this implementation for a more detailed analysis of the neuronal representation itself.

To that end, Chapter 4 develops a simplified model of the neuronal representation following the two-channel hypothesis. This two-channel model (TC-model) was constrained based on biophysical data resulting in a condensed formulation requiring only two parameters: the response variance and the IPD of maximum response (best-IPD). Importantly, the model also considered the response variations due to noise inherent to any neuronal system. Analyzing the model responses, it was shown that it is possible to calculate the certainty with which a signal with a given IPD could be categorized as a positive or negative IPD (see Sec. 4.2). To parameterize this model, a psychoacoustic left-right experiment was conducted. This experiment measured the ability of the subjects to state whether they perceived the hearing sensation due to tones with a specific IPD as lateralized to the right or the left (see Sec. 4.3). By adjusting the two parameters of the TC-model, it was possible to get a good fit between the experimental data and the model prediction (see Sec. 4.5). Interestingly, the best-IPD parameters were found to increase linearly with the frequency which is consistent with the result of single neuron measurements reported for different mammalian species (McAlpine et al., 2001; Joris et al., 2006; Pecka et al., 2008).

While the left-right task used in Chapter 4 is a method to determine the IPD encoding accuracy, Chapter 5 investigated whether it is also possible to predict the IPD encoding acuity by analyzing the TC-model. It was shown that the IPD encoding acuity in the model could be determined by applying signal detection theory (see Sec. 5.3 and 5.4.2). Using the parameterization determined in Chapter 4, only one additional parameter d_{thr} had to be added to predict just noticeable differences (JNDs) in IPD. In a psychoacoustic experiment using tones, JNDs in IPD at different reference phases were determined. By adjusting d_{thr} it was possible to achieve a good agreement between the experimental results and the JNDs predicted by the model (see Sec. 5.3). It was also shown, that the TC-model was able to predict JND values determined in other studies after compensating for the effect of subject training by adjusting d_{thr} (see Sec. 5.3.1). The good agreement with previous studies was found for a large range of frequencies and the TC-model also predicted the sharp decrease in acuity reported for frequencies around 1300 Hz (Brughera et al., 2013).

The results of the Chapters 4 and 5 suggest that it is indeed possible to predict the acuity and accuracy of IPD perception by analyzing the encoding acuity and accuracy with the TC-model. The benefit of the proposed model is that it relies only on two

central assumptions:

I IPDs are encoded by the mean activity of the neurons in each hemisphere and the IPD dependent average response within one hemisphere can be described by a function similar to Eq. 4.2.

II The variance in the mean neuronal response can be approximated by Gaussian noise.

Reducing the model to this small set of assumptions was possible partly by limiting the signals in this thesis to tones. The benefit of using tones is that it is possible to ensure that only IPD cues are physically available to the subjects. This exclusive reliance on IPDs seems essential when relating perception to its neuronal representation. In previous discussions only a one-to-one correspondence between encoder and decoder was assumed (see also Fig. 1) but realistically one must assume the existence of several encoders and decoders as well as the possibility that one decoder interprets the outputs of multiple encoders. As a consequence, one would need the knowledge about the neuronal representation of all available physical cues to infer the acuity of perception from the neuronal representation. If the signal contains multiple frequency components, it is additionally necessary to consider how the information is combined across frequencies, and how the different frequency components interact on the level of the decoder; this would require a much more complex model.

Nevertheless, Chapter 6 introduced a concept of how signals with multiple frequency components could be represented using the TC-model. Due to the current limitation of the TC-model to tones, this discussion was based on the assumption that the signal within each frequency channel can be represented by an equivalent tone at the channel's center frequency and with an equivalent IPD neglecting any variations of amplitude or phase within the frequency channel. The discussion is based on the relative likelihood function (L-function), which can be interpreted as the likelihood that an observed neuronal response is due to a specific IPD. Calculating L-functions separately for each frequency channel resulted in a representation very similar to the cross-correlogram used in delay-line models. This makes it possible to transfer some concepts that have been developed to explain cross frequency processing in delay line models to the TC-model.

As discussed in Chapter 6, the L-function can also be used for maximum likelihood decoding. A maximum likelihood decoder would use the location of the L-function peak along the IPD axis as its prediction. Using this decoder in combination with the TC-model make it possible to simulate the experiments presented in this thesis directly. Figure 8.1a shows the results gained by simulating the left-right experiment of Chapter 4 using the TC-model in combination with the maximum likelihood predictor. The experiment was simulated by generating ten observed responses per IPD from the distribution following Eq. (4.6). The L-function was then calculated for each of the observed responses. If the peak of the L-function was located at a negative IPD, the

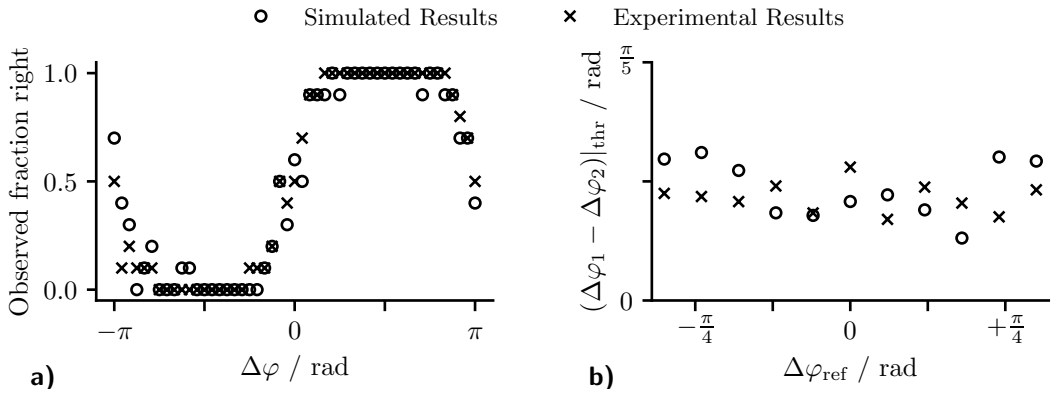

Figure 8.1: Comparison between the modeling and experimental results using a maximum likelihood decoder based on the L-function. **a)** Comparison between the simulated and experimental results of the left-right experiment at 500 Hz as presented in Chapter 4. The model was run with the corrected parameters following Eq. (5.4) and $\sigma_C = 0.19$ as determined for the experiment. **b)** Same as a) but for the JND experiment at 500 Hz (see Chapter 5.3). The model was run with the same parameters as in a) but with $\sigma_C = 0.14$ as determined for the experiment.

presentation was considered as "perceived to the left" while a peak at a positive IPD was considered as "perceived to the right". Figure. 8.1a depicts the mean over the ten presentations at each IPD.

Figure 8.1b shows the results of a simulated JND task as presented in Chapter 5. The experiment was simulated by using exactly the same parameters as discussed in Sec. 5.2. Each presentation of a stimulus was simulated by generating observed responses for each of the two IPDs following Eq. 5.1. The presentation was noted as 'correct' if the relative location of the two peaks matched the presented IPDs. This means that the peak of the L-function calculated for the observed response due to the smaller IPD was also located at a smaller IPD. JNDs were then determined as described in Sec. 5.2.5.

The agreement between experimental and simulated results as shown in Fig. 8.1 is not surprising. The model predictions, shown in the respective chapters, can be interpreted as the optimal, noise-free simulation results. The simulated results would converge towards the optimal predictions, if the simulation is run for a large number of presentations. Nevertheless, the simulations nicely visualize how the L-function can be used for maximum likelihood decoding.

In the left-right experiment of Chapter 4, IPDs with magnitudes above approximately $\pi/2$ increasingly resulted in multiple hearing sensations. This phenomenon was investigated in the psychoacoustic experiment described in Sec. 4.4, where tonal stimuli with different IPDs were presented to the subjects which subsequently had to indicate the perceived number of spatially separated hearing sensations. In this experiment, subjects reported only one hearing sensation for smaller IPDs between

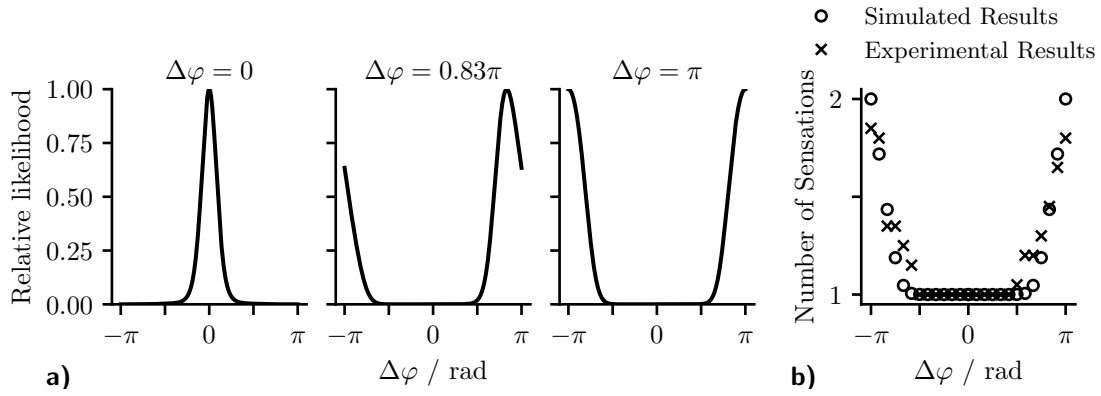

Figure 8.2: a) L-functions calculated for three selected IPDs at a frequency of 250 Hz assuming $\mathbf{R}_{obs}(\Delta\varphi) = \bar{\mathbf{R}}(\Delta\varphi)$. The model was run with the corrected parameters and $\sigma_C = 0.19$. At larger IPDs the peak of the L-function is split into two which might explain the appearance of a second hearing sensation. **b)** Comparison between the simulated and experimentally determined number of hearing sensations at 250 Hz.

approximately $\pm 0.5\pi$. With increasing IPD, the subjects also increasingly reported a second hearing sensation. In the discussion of Chapter 4, the occurrence of a second hearing sensation was used to argue against the use of a maximum likelihood decoder, which would not predict two hearing sensations but one randomly switching sides. Instead, it was proposed that a hearing sensation could be perceived in all locations where IPDs showed a high likelihood of resulting in the observed response.

This concept can also be implemented using the L-function based decoder. Figure 8.2a shows L-functions calculated for three selected IPDs assuming $\mathbf{R}_{obs}(\Delta\varphi) = \bar{\mathbf{R}}(\Delta\varphi)$. These figures visualize how the L-function for large IPDs is split into two peaks. Following the argument presented in Chapter 4, these peaks would result in two spatially separated hearing sensations. A simple way of evaluating the two peaks is by calculating the fraction of the L-function within each peak. At π, both peaks are of equal size resulting in a fraction of 0.5 for both peaks. With decreasing IPD, the fraction of the peak at positive IPDs increases, whereas the fraction of the peak at negative IPDs decreases. Assuming that the two hearing sensations are most salient at π,where most subjects also report both sensations, and that the salience decreases with a decreasing second peak, it is possible to gain a prediction for the experiment. Figure 8.2b shows a comparison between the experimental results and the predicted results. In this plot, the fraction of the primary peak is plotted, so that a fraction of 0.5 is denoted as two hearing sensations while a fraction of 1 is denoted as one hearing sensation and intermediate values calculated linearly. The good agreement between predicted and experimental data further supports the L-function approach.

Under normal listening conditions, the appearance of a second hearing sensation and the resulting ambiguity are only of relevance if the naturally occurring IPDs extend to values above approximately $\pi/2$. In these cases, additional cues would be necessary

to resolve the ambiguity. As noted in Sec. 2.1, IPDs and interaural level differences (ILDs) do not exist independently from each other under natural listening conditions. As a consequence, ILDs are a good candidate for resolving ambiguities in the IPD cue. In Chapter 7, the magnitude of naturally occurring IPDs and ILDs was investigated by analyzing free-field binaural transfer functions. The impact of ILDs on the perceived lateralization due to IPDs was evaluated in a left-right experiment similar to the one discussed in Chapter 4. Using these data, it was shown that IPD ambiguity becomes relevant in a free-field setting starting at approximately 500 Hz and that the ILDs available at these frequencies are sufficient to resolve the IPD ambiguity.

Recapitulating this discussion, the proposed TC-model was shown to explain several aspects of pure-tone lateralization, while minimizing the number of parameters and retaining physiological plausibility. These findings strongly support the two-channel hypothesis and its applicability to IPD processing in humans. While it is likely that this model needs to be extended to account for signals more complex than tones, the presented concepts should create a good basis for these extensions.

Bibliography

Abel, S. M. and Kunov, H. Lateralization based on interaural phase differences: Effects of frequency, amplitude, duration, and shape of rise/decay. *The Journal of the Acoustical Society of America*, 73(3):955960, Mar 1983. doi: 10.1121/1.389020.

Almeida, J. S. Predictive non-linear modeling of complex data by artificial neural networks. *Current Opinion in Biotechnology*, 13(1):7276, Feb 2002. doi: 10.1016/s0958-1669(02)00288-4.

Ashida, G., Kretzberg, J., and Tollin, D. J. Roles for coincidence detection in coding amplitude-modulated sounds. *PLOS Computational Biology*, 12(6):e1004997, 2016. doi: 10.1371/journal.pcbi.1004997.

Baumann, V. J., Lehnert, S., Leibold, C., and Koch, U. Tonotopic organization of the hyperpolarization-activated current (I_h) in the mammalian medial superior olive. *Frontiers in Neural Circuits*, 7(July):117, jan 2013. doi: 10.3389/fncir.2013.00117.

Beiderbeck, B., Myoga, M. H., Mller, N. I. C., Callan, A. R., Friauf, E., Grothe, B., and Pecka, M. Precisely timed inhibition facilitates action potential firing for spatial coding in the auditory brainstem. *Nature Communications*, 9(1), May 2018. doi: 10.1038/s41467-018-04210-y.

Belliveau, L. A. C., Lyamzin, D. R., and Lesica, N. A. The neural representation of interaural time differences in gerbils is transformed from midbrain to cortex. *Journal of Neuroscience*, 34(50):16796–16808, 2014. doi: 10.1523/jneurosci.2432-14.2014.

Bernstein, L. R. and Trahiotis, C. Lateralization of low-frequency, complex waveforms: The use of envelope-based temporal disparities. *The Journal of the Acoustical Society of America*, 77(5):18681880, May 1985. doi: 10.1121/1.391938.

Bernstein, L. R. and Trahiotis, C. An interaural-correlation-based approach that accounts for a wide variety of binaural detection data. *The Journal of the Acoustical Society of America*, 141(2):11501160, Feb 2017. doi: 10.1121/1.4976098.

Blauert, J. *Spatial hearing : the psychophysiscs of human sound localization*. MIT Press, 1999. ISBN 0-262-02413-6.

Boudreau, J. C. and Tsuchitani, C. Binaural interaction in the cat superior olive s segment. *Journal of Neurophysiology*, 31(3):442454, May 1968. doi: 10.1152/jn.1968.31.3.442.

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. Precise inhibition is essential for microsecond interaural time difference coding. *Nature*, 417 (6888):543–7, may 2002. doi: 10.1038/417543a.

Bremen, P. and Joris, P. X. Axonal recordings from medial superior olive neurons obtained from the lateral lemniscus of the chinchilla (chinchilla laniger). *Journal of Neuroscience*, 33(44):1750617518, Oct 2013. doi: 10.1523/jneurosci.1518-13.2013.

Briley, P. M., Kitterick, P. T., and Summerfield, A. Q. Evidence for opponent process analysis of sound source location in humans. *Journal of the Association for Research in Otolaryngology*, 14(1):83101, Oct 2012. doi: 10.1007/s10162-012-0356-x.

Bronkhorst, A. W. The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Attention, Perception, & Psychophysics*, 77(5):14651487, Apr 2015. doi: 10.3758/s13414-015-0882-9.

Brughera, A., Dunai, L., and Hartmann, W. M. Human interaural time difference thresholds for sine tones: The high-frequency limit. *Journal of the Acoustical Society of America*, 133(5):2839, 2013. doi: 10.1121/1.4795778.

Brughera, A. R., Stutman, E. R., Carney, L. H., and Colburn, H. S. A model with excitation and inhibition for cells in the medial superior olive. *Auditory Neuroscience*, 2:219–233, 1996.

Cant, N. B. and Casseday, J. H. Projections from the anteroventral cochlear nucleus to the lateral and medial superior olivary nuclei. *The Journal of Comparative Neurology*, 247(4):457476, May 1986. doi: 10.1002/cne.902470406.

Carr, C. E. and Konishi, M. Axonal delay lines for time measurement in the owl's brainstem. *Proceedings of the National Academy of Sciences*, 85(21):8311–8315, 1988.

Cherry, E. C. Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, 25(5):975979, Sep 1953. doi: 10.1121/1.1907229.

Cherry, E. C. and Sayers, B. M. A. human cross-correlator-a technique for measuring certain parameters of speech perception. *The Journal of the Acoustical Society of America*, 28(5):889895, Sep 1956. doi: 10.1121/1.1908506.

Colburn, H. S. Theory of binaural interaction based on auditorynerve data. I. general strategy and preliminary results on interaural discrimination. *The Journal of the Acoustical Society of America*, 54(6):14581470, Dec 1973. doi: 10.1121/1.1914445.

Colburn, H. S. Theory of binaural interaction based on auditory-nerve data. II. detection of tones in noise. *The Journal of the Acoustical Society of America*, 61(2): 525533, Feb 1977. doi: 10.1121/1.381294.

Colburn, S. H. and Durlach, N. I. Models of binaural interaction. In Carterette, E. and Friedman, M., editors, *Handbook of Perception*, volume 4, chapter 10, pages 467–518. Academic Press, New York, 1978.

Couchman, K., Grothe, B., and Felmy, F. Medial superior olivary neurons receive surprisingly few excitatory and inhibitory inputs with balanced strength and short-term dynamics. *The Journal of Neuroscience*, 30(50):17111–21, dec 2010. doi: 10.1523/JNEUROSCI.1760-10.2010.

Day, M. L. and Delgutte, B. Decoding sound source location and separation using neural population activity patterns. *Journal of Neuroscience*, 33(40):15837–15847, 2013. doi: 10.1523/jneurosci.2034-13.2013.

Dehmel, S., Kopp-Scheinpflug, C., Weick, M., Drrscheidt, G. J., and Rübsamen, R. Transmission of phase-coupling accuracy from the auditory nerve to spherical bushy cells in the mongolian gerbil. *Hearing Research*, 268(1-2):234249, Sep 2010. doi: 10.1016/j.heares.2010.06.005.

Dietz, M., Ewert, S. D., and Hohmann, V. Lateralization of stimuli with independent fine-structure and envelope-based temporal disparities. *The Journal of the Acoustical Society of America*, 125(3):16221635, Mar 2009. doi: 10.1121/1.3076045.

Dietz, M., Ewert, S. D., and Hohmann, V. Auditory model based direction estimation of concurrent speakers from binaural signals. *Speech Communication*, 53(5):592605, May 2011. doi: 10.1016/j.specom.2010.05.006.

Dietz, M., Lestang, J.-H., Majdak, P., Stern, R. M., Marquardt, T., Ewert, S. D., Hartmann, W. M., and Goodman, D. F. A framework for testing and comparing binaural models. *Hearing Research*, 360:92106, Mar 2018. doi: 10.1016/j.heares.2017.11.010.

Domnitz, R. The interaural time JND as a simultaneous function of interaural time and interaural amplitude. *J. Acoust. Soc. Am.*, 53(6):1549, 1973. doi: 10.1121/1.1913500.

Domnitz, R. H. and Colburn, H. S. Lateral position and interaural discrimination. *Journal of the Acoustical Society of America*, 61(6):1586–1598, Jun 1977. doi: 10.1121/1.381472.

Encke, J., Reimann, D., Hemmert, W., and Völk, F. On the role of interaural level differences in low-frequency pure-tone lateralization. *Acta Acustica united with Acustica*, 104, 2018.

Encke, J. and Hemmert, W. Extraction of inter-aural time differences using a spiking neuron network model of the medial superior olive. *Frontiers in Neuroscience*, 12, Mar 2018. doi: 10.3389/fnins.2018.00140.

Fastl, H. and Zwicker, E. *Psychoacoustics : facts and models*. Springer, 2007. ISBN 3-540-23159-5. doi: 10.1007/978-3-540-68888-4.

Franken, T. P., Roberts, M. T., Wei, L., Golding, N. L., and Joris, P. X. In vivo coincidence detection in mammalian sound localization generates phase delays. *Nature Neuroscience*, 18(3), 2015. doi: 10.1038/nn.3948.

Friauf, E. and Ostwald, J. Divergent projections of physiologically characterized rat ventral cochlear nucleus neurons as shown by intra-axonal injection of horseradish peroxidase. *Experimental Brain Research*, 73(2), Nov 1988. doi: 10.1007/bf00248219.

Glackin, B., Wall, J., McGinnity, T., Maguire, L., and McDaid, L. A spiking neural network model of the medial superior olive using spike timing dependent plasticity for sound localization. *Frontiers in Computational Neuroscience*, 4(18), 2010. doi: 10.3389/fncom.2010.00018.

Goldberg, J. M. and Brown, P. B. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *Journal of Neurophysiology*, 32(4):613636, Jul 1969. doi: 10.1152/jn.1969.32.4.613.

Goodman, D. F. M. The brian simulator. *Frontiers in Neuroscience*, 3(2):192197, Sep 2009. doi: 10.3389/neuro.01.026.2009.

Goodman, D. F., Benichoux, V., and Brette, R. Decoding neural responses to temporal cues for sound localization. *eLife*, 2(nil):nil, 2013. doi: 10.7554/elife.01312.

Green, D. M. Psychoacoustics and detection theory. *The Journal of the Acoustical Society of America*, 32(10):1189–1203, Oct 1960. doi: 10.1121/1.1907882.

Grothe, B., Pecka, M., and McAlpine, D. Mechanisms of sound localization in mammals. *Physiological Reviews*, 90(3):9831012, Jul 2010. doi: 10.1152/physrev.00026.2009.

Grothe, B. and Pecka, M. The natural history of sound localization in mammals, a story of neuronal inhibition. *Frontiers in Neural Circuits*, 8(October):1–19, oct 2014. doi: 10.3389/fncir.2014.00116.

Guinan, J. J., Norris, B. E., and Guinan, S. S. Single auditory units in the superior olivary complex: II: Locations of unit categories and tonotopic organization. *International Journal of Neuroscience*, 4(4):147166, Jan 1972. doi: 10.3109/00207457209164756.

Hafter, E. R. and De Maio, J. Difference thresholds for interaural delay. *Journal of the Acoustical Society of America*, 57(1):181, 1975. doi: 10.1121/1.380412.

Hancock, K. E. A physiologically based model of interaural time difference discrimination. *Journal of Neuroscience*, 24(32):71107117, Aug 2004. doi: 10.1523/jneurosci.0762-04.2004.

Harper, N. S., Scott, B. H., Semple, M. N., and McAlpine, D. The neural code for auditory space depends on sound frequency and head size in an optimal manner. *PLoS ONE*, 9(11):e108154, Nov 2014. doi: 10.1371/journal.pone.0108154.

Harrison, J. M. and Warr, W. B. A study of the cochlear nuclei and ascending auditory pathways of the medulla. *The Journal of Comparative Neurology*, 119(3):341379, Dec 1962. doi: 10.1002/cne.901190306.

Hartmann, W. M., Rakerd, B., Crawford, Z. D., and Zhang, P. X. Transaural experiments and a revised duplex theory for the localization of low-frequency tones. *The Journal of the Acoustical Society of America*, 139(2):968985, Feb 2016. doi: 10.1121/1.4941915.

Henning, G. B. Detectability of interaural delay in high-frequency complex waveforms. *The Journal of the Acoustical Society of America*, 55(1):8490, Jan 1974. doi: 10.1121/1.1928135.

Hershkowitz, R. M. and Durlach, N. I. Interaural time and amplitude JNDs for a 500Hz tone. *The Journal of the Acoustical Society of America*, 46(6B):14641467, Dec 1969. doi: 10.1121/1.1911887.

Husler, R., Colburn, S., and Marr, E. Sound localization in subjects with impaired hearing. *Acta Oto-Laryngologica*, 96(sup400):162, Jan 1983. doi: 10.3109/00016488309105590.

Jeffress, L. A. A place theory of sound localization. *Journal of Comparative and Physiological Psychology*, 41(1):35, 1948.

Jenkins, W. M. and Masterton, R. B. Sound localization: effects of unilateral lesions in central auditory system. *Journal of Neurophysiology*, 47(6):987–1016, 1982.

Jercog, P. E., Svirskis, G., Kotak, V. C., Sanes, D. H., and Rinzel, J. Asymmetric excitatory synaptic dynamics underlie interaural time difference processing in the

auditory system. *PLoS Biology*, 8(6):e1000406, Jun 2010. doi: 10.1371/journal.pbio.1000406.

Joris, P. X. and Yin, T. C. Envelope coding in the lateral superior olive. I. sensitivity to interaural time differences. *Journal of Neurophysiology*, 73(3):10431062, Mar 1995. doi: 10.1152/jn.1995.73.3.1043.

Joris, P. X., Carney, L. H., Smith, P. H., and Yin, T. C. Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. responses to tones at the characteristic frequency. *Journal of neurophysiology*, 71(3):1022–1036, 1994.

Joris, P. X., de Sande, B. V., Louage, D. H., and van der Heijden, M. Binaural and cochlear disparities. *Proceedings of the National Academy of Sciences*, 103(34): 12917–12922, 2006. doi: 10.1073/pnas.0601396103.

Keine, C. and Rübsamen, R. Inhibition shapes acoustic responsiveness in spherical bushy cells. *Journal of Neuroscience*, 35(22):85798592, Jun 2015. doi: 10.1523/jneurosci.0133-15.2015.

Keine, C., Rübsamen, R., and Englitz, B. Inhibition in the auditory brainstem enhances signal representation and regulates gain in complex acoustic environments. *eLife*, 5, Nov 2016. doi: 10.7554/elife.19295.

Keine, C., Rübsamen, R., and Englitz, B. Signal integration at spherical bushy cells enhances representation of temporal structure but limits its range. *eLife*, 6, Sep 2017. doi: 10.7554/elife.29639.

Kerber, S. and Seeber, B. U. Sound localization in noise by normal-hearing listeners and cochlear implant users. *Ear and Hearing*, 33(4):445457, 2012. doi: 10.1097/aud.0b013e318257607b.

Khurana, S., Remme, M. W. H., Rinzel, J., and Golding, N. L. Dynamic interaction of I_h and I_{K-LVA} during trains of synaptic potentials in principal neurons of the medial superior olive. *The Journal of Neuroscience*, 31(24):8936–47, 2011. doi: 10.1523/JNEUROSCI.1079-11.2011.

Klumpp, R. G. Some measurements of interaural time difference thresholds. *J. Acoust. Soc. Am.*, 28(5):859, 1956. doi: 10.1121/1.1908493.

Kunov, H. and Abel, S. M. Effects of rise/decay time on the lateralization of interaurally delayed 1khz tones. *The Journal of the Acoustical Society of America*, 69(3):769773, Mar 1981. doi: 10.1121/1.385577.

Künzel, T., Borst, J. G. G., and van der Heijden, M. Factors controlling the input-output relationship of spherical bushy cells in the gerbil cochlear nucleus. *Journal of Neuroscience*, 31(11):42604273, Mar 2011. doi: 10.1523/jneurosci.5433-10.2011.

Lehnert, S., Ford, M. C., Alexandrova, O., Hellmundt, F., Felmy, F., Grothe, B., and Leibold, C. Action potential generation in an anatomically constrained model of medial superior olive axons. *The Journal of Neuroscience*, 34(15):5370–84, apr 2014. doi: 10.1523/JNEUROSCI.4038-13.2014.

Lindemann, W. Extension of a binaural crosscorrelation model by contralateral inhibition. I. simulation of lateralization for stationary signals. *The Journal of the Acoustical Society of America*, 80(6):16081622, Dec 1986. doi: 10.1121/1.394325.

Magezi, D. A. and Krumbholz, K. Evidence for opponent-channel coding of interaural time differences in human auditory cortex. *Journal of Neurophysiology*, 104(4): 1997–2007, 2010. doi: 10.1152/jn.00424.2009.

Majdak, P., Goupell, M. J., and Laback, B. 3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training. *Atten. Percept. Psycho.*, 72(2):454–469, Feb 2010. doi: 10.3758/app.72.2.454.

Makous, J. C. and Middlebrooks, J. C. Twodimensional sound localization by human listeners. *The Journal of the Acoustical Society of America*, 87(5):21882200, May 1990. doi: 10.1121/1.399186.

Malhotra, S., Hall, A. J., and Lomber, S. G. Cortical control of sound localization in the cat: Unilateral cooling deactivation of 19 cerebral areas. *Journal of Neurophysiology*, 92(3):16251643, Sep 2004. doi: 10.1152/jn.01205.2003.

McAlpine, D. and Grothe, B. Sound localization and delay lines–do mammals fit the model? *Trends in neurosciences*, 26(7):347–50, jul 2003. doi: 10.1016/S0166-2236(03) 00140-1.

McAlpine, D., Jiang, D., and Palmer, A. R. A neural code for low-frequency sound localization in mammals. *Nature Neuroscience*, 4(4):396–401, 2001. doi: 10.1038/ 86049.

Mikiel-Hunter, J., Kotak, V., and Rinzel, J. High-frequency resonance in the gerbil medial superior olive. *PLOS Computational Biology*, 12(11):e1005166, Nov 2016. doi: 10.1371/journal.pcbi.1005166.

Mills, A. W. On the minimum audible angle. *The Journal of the Acoustical Society of America*, 30(4):237, 1958. doi: 10.1121/1.1909553.

Moushegian, G. and Jeffress, L. A. Role of interaural time and intensity differences in the lateralization of low-frequency tones. *Journal of the Acoustical Society of America*, 31(11):1441–1445, Nov 1959. doi: 10.1121/1.1907647.

Myoga, M. H., Lehnert, S., Leibold, C., Felmy, F., and Grothe, B. Glycinergic inhibition tunes coincidence detection in the auditory brainstem. *Nature communications*, 5 (May):3790, jan 2014. doi: 10.1038/ncomms4790.

Noble, W., Byrne, D., and Lepage, B. Effects on sound localization of configuration and type of hearing impairment. *The Journal of the Acoustical Society of America*, 95(2):9921005, Feb 1994. doi: 10.1121/1.408404.

Nuetzel, J. M. and Hafter, E. R. Lateralization of complex waveforms: Effects of fine structure, amplitude, and duration. *The Journal of the Acoustical Society of America*, 60(6):13391346, Dec 1976. doi: 10.1121/1.381227.

Pecka, M., Brand, A., Behrend, O., and Grothe, B. Interaural time difference processing in the mammalian medial superior olive: The role of glycinergic inhibition. *The Journal of Neuroscience*, 28(27):6914–25, jul 2008. doi: 10.1523/JNEUROSCI.1660-08.2008.

Phillips, D. P. and Hall, S. E. Psychophysical evidence for adaptation of central auditory processors for interaural differences in time and level. *Hearing Research*, 202(1-2):188199, Apr 2005. doi: 10.1016/j.heares.2004.11.001.

Phillips, D. P., Carmichael, M. E., and Hall, S. E. Interaction in the perceptual processing of interaural time and level differences. *Hearing Research*, 211(1-2):96102, Jan 2006. doi: 10.1016/j.heares.2005.10.005.

Plenge, G. On the differences between localization and lateralization. *The Journal of the Acoustical Society of America*, 56(3):944951, Sep 1974. doi: 10.1121/1.1903353.

Pulkki, V. and Hirvonen, T. Functional count-comparison model for binaural decoding. *Acta Acustica united with Acustica*, 95(5):883900, Sep 2009. doi: 10.3813/aaa.918220.

Rayleigh, L. Xii. on our perception of sound direction. *Philos. Mag. Series 6*, 13(74): 214–232, 1907. doi: 10.1080/14786440709463595.

Remme, M. W. H., Donato, R., Mikiel-Hunter, J., Ballestero, J. a., Foster, S., Rinzel, J., and McAlpine, D. Subthreshold resonance properties contribute to the efficient coding of auditory spatial cues. *Proceedings of the National Academy of Sciences of the United States of America*, pages 1–10, may 2014. doi: 10.1073/pnas.1316216111.

Roberts, M. T., Seeman, S. C., and Golding, N. L. A mechanistic understanding of the role of feedforward inhibition in the mammalian sound localization circuitry. *Neuron*, 78(5):923–35, jun 2013. doi: 10.1016/j.neuron.2013.04.022.

Rothman, J. S. and Manis, P. B. The roles potassium currents play in regulating the electrical activity of ventral cochlear nucleus neurons. *Journal of Neurophysiology*, 89(6):3097–3113, 2003. doi: 10.1152/jn.00127.2002.

Rudnicki, M., Schoppe, O., Isik, M., Völk, F., and Hemmert, W. Modeling auditory coding: from sound to spikes. *Cell and Tissue Research*, 361(1):159175, Jun 2015. doi: 10.1007/s00441-015-2202-z.

Rudniki, M. and Hemmert, W. High entrainment constrains synaptic depression levels of an in vivo globular bushy cell model. *Frontiers in Computational Neuroscience*, 11(16), Mar 2017. doi: 10.3389/fncom.2017.00016.

Salminen, N. H., Tiitinen, H., Yrttiaho, S., and May, P. J. C. The neural code for interaural time difference in human auditory cortex. *The Journal of the Acoustical Society of America*, 127(2):EL60EL65, Feb 2010. doi: 10.1121/1.3290744.

Salminen, N. H., Jones, S. J., Christianson, G. B., Marquardt, T., and McAlpine, D. A common periodic representation of interaural time differences in mammalian cortex. *NeuroImage*, 167:95103, Feb 2018. doi: 10.1016/j.neuroimage.2017.11.012.

Sayers, B. M. Acoustic-image lateralization judgments with binaural tones. *Journal of the Acoustical Society of America*, 36(5):923–926, May 1964. doi: 10.1121/1.1919121.

Sayers, B. M. and Cherry, E. C. Mechanism of binaural fusion in the hearing of speech. *The Journal of the Acoustical Society of America*, 29(9):973987, Sep 1957. doi: 10.1121/1.1914990.

Scott, L. L., Mathews, P. J., and Golding, N. L. Posthearing developmental refinement of temporal processing in principal neurons of the medial superior olive. *The Journal of Neuroscience*, 25(35):7887–7895, 2005. doi: 10.1523/JNEUROSCI.1016-05.2005.

Seeber, B. U., Baumann, U., and Fastl, H. Localization ability with bimodal hearing aids and bilateral cochlear implants. *The Journal of the Acoustical Society of America*, 116(3):16981709, Sep 2004. doi: 10.1121/1.1776192.

Sivian, L. J. and White, S. D. On minimum audible sound fields. *The Journal of the Acoustical Society of America*, 4(4):288321, Apr 1933. doi: 10.1121/1.1915608.

Smith, R. L. Short-term adaptation in single auditory nerve fibers: some poststimulatory effects. *Journal of Neurophysiology*, 1977.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, 416(6876):8790, Mar 2002. doi: 10.1038/416087a.

Spangler, K. M., Warr, W. B., and Henkel, C. K. The projections of principal cells of the medial nucleus of the trapezoid body in the cat. *The Journal of Comparative Neurology*, 238(3):249262, Aug 1985. doi: 10.1002/cne.902380302.

Stange, A., Myoga, M. H., Lingner, A., Ford, M. C., Alexandrova, O., Felmy, F., Pecka, M., Siveke, I., and Grothe, B. Adaptation in sound localization: from GABA(B) receptor-mediated synaptic modulation to perception. *Nature neuroscience*, 16(12): 1840–7, dec 2013. doi: 10.1038/nn.3548.

Stecker, G. C., Harrington, I. A., and Middlebrooks, J. C. Location coding by opponent neural populations in the auditory cortex. *PLoS Biology*, 3(3):e78, Feb 2005. doi: 10.1371/journal.pbio.0030078.

Stern, R. M. and Colburn, H. S. Lateral-position-based models of interaural discrimination. *The Journal of the Acoustical Society of America*, 77(2):753755, Feb 1985. doi: 10.1121/1.392345.

Stern, R. M. and Shear, G. D. Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay. *The Journal of the Acoustical Society of America*, 100(4):22782288, Oct 1996. doi: 10.1121/1.417937.

Stern, R. M., Brown, G. J., and Wang, D. Binaural sound localization. In Wang, D. and Brown, G. J., editors, *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*, chapter 5, pages 147–185. Wiley-IEEE Press, New York, 2006.

Takanen, M., Santala, O., and Pulkki, V. Visualization of functional count-comparison-based binaural auditory model output. *Hearing Research*, 309:147–163, 2014. doi: 10.1016/j.heares.2013.10.004.

Tobias, J. V. Lateralization threshold as a function of stimulus duration. *The Journal of the Acoustical Society of America*, 31(12):1591, 1959. doi: 10.1121/1.1907664.

Trahiotis, C. and Stern, R. M. Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase. *The Journal of the Acoustical Society of America*, 86(4):12851293, Oct 1989. doi: 10.1121/1.398743.

Van den Bogaert, T., Klasen, T. J., Moonen, M., Van Deun, L., and Wouters, J. Horizontal localization with bilateral hearing aids: Without is better than with. *The Journal of the Acoustical Society of America*, 119(1):515526, Jan 2006. doi: 10.1121/1.2139653.

van Hoesel, R. J. M. and Tyler, R. S. Speech perception, localization, and lateralization with bilateral cochlear implants. *The Journal of the Acoustical Society of America*, 113(3):16171630, Mar 2003. doi: 10.1121/1.1539520.

Völk, F. *Interrelations of Virtual Acoustics and Hearing Research by the Example of Binaural Synthesis*. PhD thesis, Technische Universität München, 2013.

Völk, F. Inter- and intra-individual variability in the blocked auditory canal transfer functions of three circum-aural headphones. *J. Audio. Eng. Soc.*, 62(5):315–323, Jun 2014. doi: 10.17743/jaes.2014.0021.

Völk, F., Encke, J., Kreh, J., and Hemmert, W. Relevance of headphone characteristics in binaural listening experiments: A case study. In *Audio Engineering Society Convention 143*, Oct 2017a.

Völk, F., Encke, J., Kreh, J., and Hemmert, W. Pure-tone lateralization revisited. In *Proceedings of Meetings on Acoustics*, pages 2611–2611. Acoustical Society of America, Oct 2017b. doi: 10.1121/2.0000784.

Von Békésy, G. and Wever, E. G. *Experiments in hearing*, volume 8. McGraw-Hill New York, 1960.

Wagener, K., Brand, T., and Kollmeier, B. Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests. *Zeitschrift für Audiologie*, 38:8695, 1999.

Wang, L., Devore, S., Delgutte, B., and Colburn, H. S. Dual sensitivity of inferior colliculus neurons to ITD in the envelopes of high-frequency sounds: experimental and modeling study. *Journal of Neurophysiology*, 111(1):164–181, 2013. doi: 10.1152/jn.00450.2013.

Wichmann, F. A. and Hill, N. J. The psychometric function: I. fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63(8):12931313, Nov 2001. doi: 10.3758/bf03194544.

Wightman, F. L. and Kistler, D. J. The dominant role of low-frequency interaural time differences in sound localization. *Journal of the Acoustical Society of America*, 91(3):1648–1661, Mar 1992. doi: 10.1121/1.402445.

Yin, T. C. and Chan, J. C. Interaural time sensitivity in medial superior olive of cat. *Journal of Neurophysiology*, 64(2):465–488, 1990. doi: 10.1152/jn.1990.64.2.465.

Yin, T. C. and Kuwada, S. Binaural interaction in low-frequency neurons in inferior colliculus of the cat. III. effects of changing frequency. *Journal of Neurophysiology*, 50(4):1020–1042, 1983.

Yost, W. A. Discriminations of interaural phase differences. *The Journal of the Acoustical Society of America*, 55(6):1299, 1974. doi: 10.1121/1.1914701.

Yost, W. A. Lateral position of sinusoids presented with interaural intensive and temporal differences. *The Journal of the Acoustical Society of America*, 70(2):397, 1981. doi: 10.1121/1.386775.

Yost, W. A. and Dye, R. H. Discrimination of interaural differences of level as a function of frequency. *Journal of the Acoustical Society of America*, 83(5):1846–1851, May 1988. doi: 10.1121/1.396520.

Zhou, Y., Carney, L. H., and Colburn, H. S. 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):3046–58, mar 2005. doi: 10.1523/JNEUROSCI.3064-04.2005.

Zilany, M. S. a., Bruce, I. C., and Carney, L. H. Updated parameters and expanded simulation options for a model of the auditory periphery. *The Journal of the Acoustical Society of America*, 135(1):283–6, jan 2014. doi: 10.1121/1.4837815.

Zwislocki, J. and Feldman, R. S. Just noticeable differences in dichotic phase. *Journal of the Acoustical Society of America*, 28(5):860–864, Sep 1956. doi: 10.1121/1.1908495.

Glossary and Symbols

Glossary

ANF	auditory nerve fiber
ANN	artificial neuronal network
ARI	acoustics research institute
BLT	binary lateralization task
CI	cochlear implant
CN	cochlear nucleus
DFT	discrete Fourier transform
DIPD	delta IPD
EPSC	excitatory post-synaptic current
EPSP	excitatory post-synaptic potential
FFT	fast Fourier-transform
GBC	globular bushy cell
HRTF	head-related transfer function
HS	hearing sensation
IC	inferior colliculus
ILD	interaural level difference
IPD	interaural phase difference
IPSC	inhibitory post-synaptic current
IPSP	inhibitory post-synaptic potential
ITD	interaural time difference
JND	just noticeable difference
L-function	relative likelihood function
LNTB	lateral nucleus of the trapezoid body
LRT	left-right task
LSO	lateral superior olive
MAA	minimal audible angle
MLP	multi-layer perceptron
MNTB	medial nucleus of the trapezoid body
MP	median plane
MSO	medial superior olive
PDF	probability density function
PMP	perceived median plane
RIPD	reference IPD
SBC	spherical bushy cell
SNN	spiking neural network
SPL	sound pressure level
TB	trapezoid body
TC-code	two-channel code
TC-model	two-channel model

Mathematical Symbols

$\Delta\varphi_{\text{best}}$	IPD of maximum response
ΔL	inter-aural level difference
$\Delta\varphi$	inter-aural phase difference
D_{IPD}	delta IPD
Δt	inter-aural time difference
d'	Sensitivity Index (d-prime)
ΔR	Difference between the responses or rates in the two hemisphers / channels
d_{thr}	distance at threshold
f_{C}	Center frequency
m, h, w, z, a	Gating variables
g_{leak}, g_{Na}, g_{KLT}, g_{h}	Ionic conductivities

I_{leak}, I_{Na}, I_{KLT}, I_{h}	Ionic currents	$\Delta\varphi_{\text{ref}}$	reference IPD
$I_{\text{syn,e}}$, $I_{\text{syn,i}}$	Synaptic Currents	$\bar{\mathbf{R}}$	Response vector
$(\Delta\varphi_1 - \Delta\varphi_2)\,\vert_{\text{thr}}$	Delta IPD	$\mathbf{R}_{obs}$	Observed response vector
P_{C}	Probability of a correct response	σ_{Δ}	Response STD of ΔR
$P(+\vert\Delta\varphi)$	Probability to observe a positive rate difference given $\Delta\varphi$	σ_{fit}	Response STD of fit
		σ_{C}	Response STD of channel
$P(r\vert\Delta\varphi)$	Probability to observe a hearing sensation (HS) to the right given $\Delta\varphi$	σ_{jnd}	Response STD determined in the jnd experiment
$\bar{R}_{\text{C}}$	Mean response in one channel	V_{m}	Transmembrane potential
$\bar{R}_{\text{C,l}}$	Mean response in the left channel	w_{C}	Determines the width of the channel response
$\bar{R}_{\text{C,r}}$	Mean response in the right channel		

List of Figures and Tables

List of Figures

List of Tables