Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Radioelectronics

FACULTY
OF ELECTRICAL
ENGINEERING
CTU IN PRAGUE

Models and experiments of binaural interactions

Doctoral Thesis

Ing. Jaroslav Bouše

Prague, 2020

Ph.D. programme: Electrical Engineering and Information Technology (P2612)
Branch of study: Acoustics (2609V001)

Supervisor: Prof. RNDr. MUDr. Petr Maršálek, Ph.D
Supervisor-Specialist: Ing. František Rund, Ph.D

Supervisor:

Prof. RNDr. MUDr. Petr Maršálek, Ph.D
Department of Radioelectronics
Faculty of Electrical Engineering
Czech Technical University in Prague
Technická 2
166 27 Prague 6
Czech Republic

Co-Supervisor:

Ing. František Rund, Ph.D
Department of Radioelectronics
Faculty of Electrical Engineering
Czech Technical University in Prague
Technická 2
166 27 Prague 6
Czech Republic

To whom I am missing

Declaration

I hereby declare that I worked out the presented thesis independently and I quoted all the sources used in this thesis in accord with Methodical instructions about ethical principles for writing academic thesis.

Prague, 31. August 2020 Jaroslav Bouse

Acknowledgments

I would like to thank at this place to all of the former and current Ph.D. students and colleagues at the department of radioelectronics. You have helped me a lot during hard times, and I appreciate this. Namely I can thank to just a few of you, to Dominik Storek, who was always optimistic about everything; Lukas Krasula, because being the same, but better than me, and for your tofu burger; Petr Janout, because of your calmness and kindness; the saint duo for listening to my never-ending rants, and finally Jakub Pospisil cause you were the funny guy even-tho you never knew it. Big thanks also go to my colleagues at Valeo, who supported me in a part of this journey. Namely, Zuzana Jakabova, whose help and mental support I appreciate a lot. And to all my friends as well. Like Michal Zajicek, who other than a good friend made for you agile development plan for the finalization of the thesis. And I shouldn't forget about my virtual colleagues which accompanied me in most of my acknowledgments, thank You T. Reks. and S. Uperjaarda.

I would also like to thank both of my supervisors, who supported me through my Ph.D. studies. And Vaclav Vencovsky, a colleague from the department and my former supervisor of both bachelor and master thesis, whose bright ideas finally led to this thesis.

Last but not least I would like to thank my family, which supported me financially and emotionally during the studies. Without you, I would never make it.

Work on this thesis was supported by: SGS20/180/OHK3/3T/13 - Audio Signal Processing Related to its Perception.

Quote of this thesis: '*It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it's wrong*' by Richard Feynman.

Jaroslav Bouse

Abstract

This dissertation thesis presents models and experiments of binaural interactions in human hearing. The rate-code models of medial and lateral superior olives (MSO and LSO) are presented. The models are inspired by recent neurophysiological findings and published in Bouse et al., J. Acoust. Soc. Am. 2019. A feature of these models is that they contain central stages of interaural time difference and interaural level difference (ITD and ILD) processing. These stages give subjective lateralization expressed in absolute numbers. The predictions made by both MSO and LSO models are compared with subjective data on the lateralization of pure tones and narrow band noises, discrimination of the ITD and ILD, and discrimination of the phase warp. The lateralization and discrimination experiments show good agreement with the subjective data.

The published models are further improved in this thesis to reduce computational demands. The improved model predictions are compared with both subjective experiments and former models data from the same test pool. Additionally, lateralization pure tone experiment on ITD versus IPD (interaural phase difference) was added to the test pool. Both versions of the models show good agreement with lateralization and discrimination subjective data. In some cases, new models show better performance than the old ones.

The experiments of binaural interactions shown in this thesis are lateralization of 1-ERB (equivalent rectangular bandwidth) wide narrow band noises with IPD or ILD, and audible quality assessment of DHRTF (differential head related transfer functions) artifact reduction methods, presented in Bouse et al., J. Acoust. Soc. Am 2019, and Storek et al., J. Audio Eng. Soc. 2016.

Keywords: binaural hearing, binaural models, lateralization, lateral superior olive, medial superior olive, opponent-coding, narrow band noise, psychoacoustics, psychoacoustic experiment, rate-code models

Abstrakt

Dizertační práce popisuje modely a experimenty binaurální interakce se zaměřením na lidské slyšení. Prezentovány jsou modely mediální a laterální superior olivy (MSO a LSO) fungujících na rate-code principu. Tyto modely jsou inspirovány nedávnými objevy v neurofyziologii a byly publikovány v Bouse et al., J. Acoust. Soc. Am. 2019. Modely navíc obsahují centrální části dekódující interaurální časové diference a interaurální úrovňové diference (ITD a ILD). Tyto části jsou pak schopné vyjádřit subjektivní lateralizaci v absolutních číslech. Predikce jak MSO, tak LSO modelu jsou porovnávány se subjektivními daty tj. lateralizací čistých tónů a úzkopásmových šumů, diskriminací ITD a ILD a diskriminací phase warpu. Jak lateralizační, tak diskriminační experimenty ukazují shodu mezi predikcemi modelů a subjektivními daty.

Publikované modely jsou v této práci dále vylepšeny s cílem snížit výpočetní nároky modelů. Predikce vylepšených modelů jsou porovnány se subjektivními daty a predikcemi původních modelů na stejných testovacích datech. Dodatečně je ještě přidán experiment s čistými tóny s ITD versus IPD (interaurální fázová diference). Obě verze modelů ukazují dobrou shodu s lateralizačními i diskriminačními subjektivními daty. V některých případech vykazují nové modely lepší výsledky než modely původní.

Experimenty binaurální interakce popisované v této dizertační práci jsou lateralizační experiment s 1-ERB (ekvivalentní pravoúhlá šířka pásma) širokými úzkopásmovými šumy s IPD nebo ILD a subjektivní hodnocení kvality metod odstraňujících rušení z DHRTF (differential head related transfer function), publikovány v Bouse et al., J. Acoust. Soc. Am 2019, a Storek et al., J. Audio Eng. Soc. 2016.

Klíčová slova: binauralní slyšení, binaurální modely, lateralizace, laterální superior oliva, mediální superior oliva, opponent-coding, narrow band noise, psychoakustika, psychoakustický experiment, rate-code modely

Contents

List of Abbreviations

Notation	Description	Page List
ANN	artificial neural network	25, 53
CF	central frequency	10, 30, 33, 34, 39, 40, 54, 64, 85
CN	cochlear nucleus	10, 11
DHRTF	differential-head-related transfer function	17, 18, 79, 87–95
DRNL	dual resonance nonlinear	29, 30, 59, 60
ERB	equivalent rectangular bandwidth	10, 29, 38, 55
FIR	finite impulse response	29
GBC	globular bushy cell	10, 11
HRIR	head-related impulse response	16, 17
HRTF	head-related transfer function	17, 18, 53, 54, 88–95
IHC	inner hair cell	59–61
IHC	inner hair cell	9
IIR	infinite impulse response	31, 33, 35
ILD	interaural level difference	4, 5, 10, 11, 13–16, 18, 23–26, 31, 35–41, 44–47, 51–55, 61, 63–65, 67, 69, 71–74, 81, 83–86
IPD	interaural phase difference	13, 17, 18, 23, 24, 26, 31, 33, 34, 37–40, 43, 44, 46, 51, 54, 59, 63, 64, 67, 68, 70, 72, 75, 79, 81–85
ITD	interaural time difference	4, 10–16, 18, 22–26, 31–34, 36–41, 47, 52–55, 60, 61, 63–65, 68, 70, 71, 73–75, 79, 85
ITF	interaural transfer function	24
KNN	K-nearest neighbour	53

Notation	Description	Page List
LSO	lateral superior olive	4, 10, 11, 24, 25, 29–31, 34–36, 38–41, 43–49, 51–55, 60, 61, 67, 69–75
MNTB	medial nucleus of the trapezoid body	10, 11
MSO	medial superior olive	4, 10–12, 22–25, 29–34, 36, 38–41, 43, 44, 46–49, 51–55, 59–61, 63, 67–75
NBN	narrow band noise	37–39, 46, 47, 64, 70, 72, 81–85
OHC	outer hair cell	9
SBC	spherical bushy cell	10, 11
SOC	superior olivary complex	10
VAS	virtual auditory space	17, 18

List of Figures

Part I

Introduction

INTRODUCTION

The German philosopher Immanuel Kant once wrote (Louden and Kuehn, 2006): ”Blindness separates us from things but deafness from people,” emphasizing hearing as a primary social sense of human beings. However, beforehand humans learn how to speak; they have to survive in nature full of predators. Of course, other skills helped humans develop into the state we now know, but without one sense that does not sleep (Campbell and Bartoli, 1986), it might be more difficult. Especially, when this sense, as in the case of all mammals, helps to locate predator in space while not seeing him (Grothe *et al.*, 2010). The sense of hearing, more specifically hearing with both ears, i.e., a binaural hearing had a significant role in every normal hearing person (Blauert, 1997). Sometimes it could even save the lives of many in situations, like hearing the horn from the lane on the highway to warn us that there is approaching car before we turn. We will most probably cause an accident if we do the turn. The knowledge of how we can hear spatially not only reveals us more about our self's as human beings. It also helps design better hearing aids for the ones with hearing problems (Laback and Majdak, 2008); to design better sound warning systems in cars (Lundqvist and Eriksson, 2019); to help game developers design more immersive games with excellent spatial sound (Larsen *et al.*, 2013), sound engineers to provide more natural sound image in cinema (Bilinski *et al.*, 2014) and many other applications.

Mathematical models showed great utility in explaining some attributes of the human spatial hearing and lead to more sophisticated physiological research (Blauert, 1997, 2013). Most of the up-to-date models can hardly be divided into two groups. The first is based on von Békésy (1930) assumption that we have two nuclei, both innervated from opposite ears. Based on the time and intensity differences of sound in both ears, they excite one nucleus more than the other. The central processor (brain) calculates the differences in excitation into the sound spatial map. This theory can be found in the literature under different names such as opponent coding, rate-coding, or count-comparison (Colburn, 1977; Pulkki and Hirvonen, 2009; Takanen *et al.*, 2014; Encke and Hemmert, 2018; Magezi and Krumbholz, 2010). The latter theory proposed by Jeffress (1948) assumes that both ears are connected to neuron populations with different time delays. This structure mathematically supplies the correlation operation, i.e., the spatial sound space is decoded by the position of the neuron with the time delay closest to the actual time delay between ears. This neuron fires up the most and gives the central processor a hint from where the sound origins. This theory is most of the time called in the literature as a delay line or correlation model (Stern and Colburn, 1978; Moore, 2003). The models from the Jeffress family are the most developed in the mean of modeling psychoacoustical data on humans. Although the Jeffress delay models are mathematically pleasant and physiologically found in bird brains, evidence in advantage to the rate-coding aroused during research on mammals and even on humans (Grothe, 2003; McAlpine and Grothe, 2003; Thompson *et al.*, 2006; Salminen *et al.*, 2010).

The caveat of the state of the art rate–code models, is that while they successively simulate the physiological truth, they lack the direct measure to compare their outputs with psychoacoustical data of humans as Jeffress family of models do (Dietz *et al.*, 2008; Pulkki and Hirvonen, 2009; Encke and Hemmert, 2018).

Another part of the modeling problem is a pool of the executed psychoacoustical experiments. While we have covered binaural interaction situations to a large extent (Blauert, 1997), we still have some which we are missing. For example, in virtual acoustic positioning systems, the quality of the positioning is evaluated most of the time, but not the quality of the overall audible impression. Another not covered topic is narrow band noises, which have the bandwidth equal to one auditory band. Naturally, they should have the same behavior as the pure tone with a frequency centered on the same band, but do they? Therefore, another topic of this thesis is to answer those questions or, at least, provide subjective data for other researchers.

Aims of the Doctoral thesis

The doctoral thesis's aim can be divided into two subjects, which are derived inevitably from the broad topic of the thesis.

The first subject (Part II and Part III) aims to further reduce the gap in the binaural modeling between neurophysiology and psychoacoustics, by introducing functional rate-code binaural models of medial superior olive (MSO) and lateral superior olive (LSO) which will meet three main criteria:

1. they will be based on current neurophysiological findings,

2. their output will give a quantitative representation of subjective lateralization based on interaural time difference (ITD) or interaural level difference (ILD) at a corresponding frequency,

3. they will have low structural complexity and low computational complexity for low power applications.

State-of-the-art binaural models withhold those conditions with limited success. Further discussion regarding the models' motivation and comparison against the criteria can be found in Chapter 3.3.

The second - the experimental subject aims to introduce new subjective data regarding both lateralization and subjective perception of virtually simulated acoustic space (Part IV).

State of the art

Due to the broad topic of the thesis, it is hard to concentrate the state of the art in a single section. In the Introduction part, we introduced two sections, which hold these topics. Regarding the experiments, it is a section named Psychoacoustics of binaural hearing. Regarding the binaural modeling, it is then the section Modeling of binaural hearing with an emphasis on the subsection Open topics of binaural modeling.

Structure of the Doctoral thesis

This doctoral thesis is divided into five parts as follows:

I. *Introduction*: The first part provides a reader a brief summary of the physiology and psychoacoustics of the binaural hearing. In the last chapter of this part, state-of-the-art binaural models are presented, and open topics of binaural modeling are stated.

II. *Designed Binaural models*: The second part is focused on description and verification of novel designed models of lateral and medial superior olives, as was published in Bouse *et al.* (2019).

III. *Designed Binaural models - modifications*: The third part focuses on the modification of the models from Part II, to further improve their ability to maintain Criterion 3 of this thesis aims. The modified models are thoroughly compared with the originals.

IV. *Psychoacoustical experiments*: The fourth part describes the psychoacoustic experiments designed and organized by the author of this thesis. In the first chapter of this part, lateralization experiment with narrow band noises with IPD or ILD is presented. This experiment was part of the authors' publication Bouse *et al.* (2019). In the second chapter of this part, the experiment about subjective evaluation of DHRTF artifacts is presented. This experiment was part of Storek *et al.* (2016) article, and this thesis author contribution was design, execution, and evaluation of the experiment.

V. *Conclusions*: The last part summarizes the results and relates them to the aims set in the introduction.

There are separate, detailed discussions within Part II, Part III, and Part IV.

Chapter 1

PHYSIOLOGY OF BINAURAL HEARING

In the present thesis, the models of the human auditory pathway from the periphery to medial and lateral superior olives are presented. The inspiration for their design comes from neurophysiological and psychoacoustical data from the literature. Henceforth, the nuclei and pathways relevant to the binaural processing are briefly reviewed in the following chapter (Figure 1.1).

1.1 Outer ear

The outer ear is the first frontier of an auditory system. It consists of two parts: the pinna and ear canal. The sound is transported through the pinna and ear canal as an acoustic wave towards the middle ear's tympanic membrane.

From a mechanistic point of view, pinna serves as a wind protector, but its more important

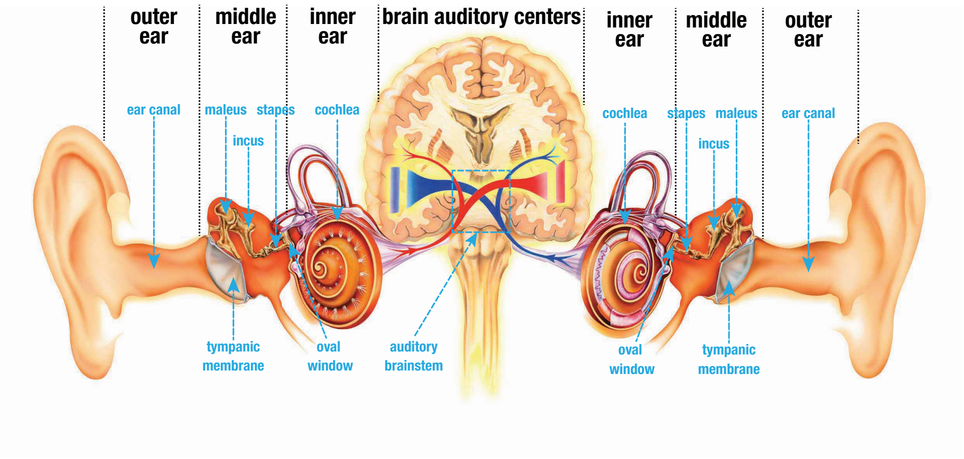

Figure 1.1: Illustration of human binaural hearing pathway. Edited from Kumar (2020).

property is the spectral filtering of the sound according to the direction of incidence (Blauert, 1997). In a horizontal plane, it sufficiently helps us to distinguish between front and back incidence. In a vertical plane, thanks to spectral unique attenuation patterns as a function of elevation, it provides meaningful cues to encode the elevation (Blauert, 1997).

The ear canal is approximately 20 mm long, an s-shaped tube opened at a lateral end, and medially terminated by a tympanic membrane at the other end (Silbernagl and Despopoulos, 2009). Its first third is made of cartilage, lined with skin containing glands, which produce cerumen (earwax). Tiny hairs and cerumen effectively add protection of the tympanic membrane from insects and small foreign particles (Silbernagl and Despopoulos, 2009). The other two-thirds are embedded in the temporal bone. Although protection function has its indisputable role, for a perception of sound, the ear canal, more importantly, acts as an acoustical resonator tuned approximately to 3–4 kHz (Wiener and Ross, 1946), i.e., it effectively amplifies salient frequencies of human speech. The sound level at the resonant frequency is increased by almost over 15 dB (Wiener and Ross, 1946).

1.2 Middle ear

The acoustic wave travels through the outer ear; in the end, it lands on a tympanic membrane, where it is transformed into mechanical vibration of the middle ear's bones.

The tympanic membrane is in a diameter of approximately 10 mm, and it is formed of three thin layers of tissue very sensitive to acoustic vibrations (Blauert, 1997). It medially terminates the middle ear cavity and assists in the protection of the delicate structures inside. The middle ear cavity is located in the temporal bone connected to the Eustachian tube. The Eustachian tube is usually closed but opens while chewing or swallowing. When opened, the air pressure between the outer and middle ear is equalized through the oral cavity (Silbernagl and Despopoulos, 2009).

The vibrations from the tympanic membrane are transferred to the oval window by the ossicular chain consisting of the middle ear's little bones, *malleus, incus*, and *stapes* (hammer, anvil, and stirrup). The *malleus* is attached directly to the tympanic membrane, the *stapes* inserts into the oval window. The *incus* is located in between the *malleus* and *stapes.* The whole middle ear amplifies the acoustic signal with a gain of about 1:18, 1:3 of gain is due to lever arm factor of the ossicular chain. However, its majority is due to the significant area difference between the tympanic membrane and the oval window (Silbernagl and Despopoulos, 2009). Such a substantial gain effectively works as the acoustical impedance matching device between the relatively small acoustical impedance of the air and the large impedance of the fluid inside the cochlea.

The middle ear protects the peripheral system against excessive stimulation by the auditory reflex. The protection works on a principle of decreasing efficiency of the ossicular chain by contracting small muscles attached to the *malleus* (*musculus tensor tympani*) and the *stapes* (*musculus stapedius*).

1.3 Inner ear

The inner ear consists of the vestibular system and cochlea. The cochlea is a tube similar in shape to a snail shell with two and a half turns in humans (Alberti, 2001). It is easier to think about the cochlea, imagining it as a straightened out tube, closed at the apex, and opened at the base with the round and oval windows and with a connection to the vestibular system (Alberti, 2001). The *stapes* transfer the mechanical vibration from the middle ear to the oval window. The vibration of the oval window is then transformed into the movement of the fluid inside the cochlea (Silbernagl and Despopoulos, 2009). This fluid is generally in-compressible; therefore, there is a need for a counterbalance in the labyrinth to allow the movement of the fluid (Alberti, 2001). The counterbalance is provided by an elastic membrane of a round window, which moves in an opposite phase than the base of *stapes* in the oval window (Alberti, 2001). The cochlear labyrinth is divided into three longitudinal sections: *scala vestibuli, scala media*, and *scala tympani* (Silbernagl and Despopoulos, 2009). Oval window vibrations are transferred through *scala vestibuli* to the cochlear apex, where they enter *scala tympani* and pass to the round window. The perilymph fills the sections. In-between them, *scala medial* is located separated from *scala vestibuli* by the Reisner's membrane and from *scala tympani* by the basilar membrane. Scala media is filled by endolymph.

The basilar membrane is composed of a significant number of taut, radially parallel fibers sealed between a gelatinous material of frail shear strength. These fibers are resonant at different frequencies as a function of distance from the oval window, i.e., at the apical end at low frequencies, at the basal end at high frequencies (Alberti, 2001). This tuning is called cochlear frequency selectivity. Four rows of hair cells lie on top of the basilar membrane, together with supporting cells. A single inner row, closest to the auditory nerve is composed of inner hair cells (IHCs), primary auditory sensors. The three outer rows, composed of outer hair cells (OHCs), receive mainly afferent nerve supplies and are considered active amplifiers of the traveling wave (Alberti, 2001). The whole structure of the hair cells is known as a tunnel of Corti (Silbernagl and Despopoulos, 2009). Any displacement of the cochlear fluid results in a motion of the tunnel of Corti and, consequently, a lateral displacement of the IHCs, bending of IHCs' cilia, followed by a neural discharge into the central auditory nervous system (Alberti, 2001).

From the functional point of view, this frequency to place distribution of sound across the cochlear body can be in a simplified manner translated into a bank of filters distributed across the hearing range. Moore and Glasberg (1983) used a notched noise method to determine the properties of these filters. The pure tone detection threshold was measured as a function of the width of notched noise, which was imposed around pure tone frequency. This measurement resulted in the equivalent rectangular bandwidth (ERB), which estimates the filter width relative to its central frequency:

$$\mathrm{ERB}(f_\mathrm{c}) = 24.7 + 0.108 f_\mathrm{c}, \tag{1.1}$$

where f_c corresponds to the central frequency (CF) of the peripheral filter in Hz.

1.4 Central auditory nervous system

The cochlear nucleus (CN) is the first synaptic station in the auditory pathway (Vetter, 2015). The CN is subdivided into the dorsal and ventral cochlear nucleus (DCN and VCN), each comprising several neural sub-types. The bushy cells in the VCN temporally enhance the input from the auditory nerve, allowing the phase-locking to the fine structure of sound up to 3 kHz, critical for binaural sound processing in ascending stages of superior olivary complex (SOC) (Grothe *et al.*, 2010).

SOC plays an indisputable role in decoding binaural cues of interaural level and time differences (ILD and ITD) (Joris and Yin, 1995; Joris, 1996; Grothe, 2003; McAlpine and Grothe, 2003; Tollin and Yin, 2005; Pecka *et al.*, 2008; Grothe *et al.*, 2010; Bures and Marsalek, 2013). It is composed of two nuclei medial and lateral superior olive (MSO and LSO).

1.4.1 Lateral superior olive

It is generally considered that the ILD processing lies in the lateral superior olive (LSO), which has excitatory inputs from the ipsilateral CN and inhibitory inputs from the contralateral cochlear nucleus (CN) (Tollin, 2003). The excitatory inputs of LSO are innervated from spherical bushy cells (SBCs) of the ipsilateral CN (Tollin, 2003). While the inhibitory inputs, LSO receives indirectly from globular bushy cells (GBCs) of the contralateral CN (Grothe *et al.*, 2010). Globular bushy cells possess superior temporal resolution than SBC due to the innervation from a more significant number of neurons from the auditory nerve and thickest diameter axons of any auditory nerve fibers (Grothe *et al.*, 2010). The GBCs project to the medial nucleus of the trapezoid body (MNTB) on the contralateral side and ascend to the LSO via calyx of Held (Grothe *et al.*, 2010).

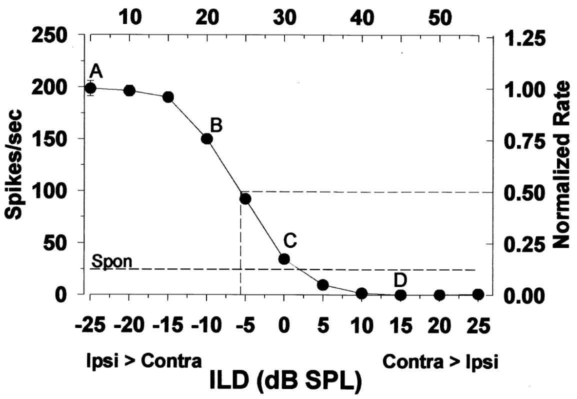

Figure 1.2: Mean discharge rate of a single cat's LSO neuron to pure tones with ILD. The discharge rate is minimal when the sound at the contralateral ear is more intense, and maximal when the sound at the ipsilateral ear is more intense. Picture taken from Tollin and Yin (2002).

The excitatory and inhibitory inputs are in LSO subtracted, forming sigmoid function (Figure 1.2). This function is minimal when the sound at the contralateral ear is more intense, and maximal when the sound at the ipsilateral ear is more intense (Tollin and Yin, 2002).

The thick diameter of the GBCs axons, and high speed and efficiency of synaptic transmission in the calyx of Held, allows the inhibitory signal to reach LSO by approximately 0.2 ms after initial contralateral cochlear excitation (Tollin and Yin, 2005). Due to this, the high temporal resolution of both excitation and inhibitory inputs allows LSO to be also ITD sensitive on low sound frequencies (Tollin and Yin, 2005).

1.4.2 Medial superior olive

While LSO is considered to decode both ILD and ITD (Tollin and Yin, 2005), MSO is known to be sensitive to the ITD with ability to resolve time differences up to two orders of magnitude shorter than the duration of action potentials bearing the information (Grothe *et al.*, 2010). The MSO receives excitatory inputs from both CNs from both hemispheres needed for coincidence detection, and inhibitory input from the contralateral ear (McAlpine and Grothe, 2003). The excitatory inputs are driven by SBCs, which accurately time-lock their discharges to the sound's temporal structure (Grothe, 2003). The dominant pathway for inhibition input originates from the MNTB, which is known for its exceptional temporal precision, and fastness, which allows the inhibitory input to arrive before the excitatory one (Grothe, 2003). Therefore MSO possesses both excitatory and inhibitory inputs with high temporal precision needed to decode the ITD from the incoming sound. In vivo recording of MSO cells responses showed that MSO is most sensitive, i.e., having the highest spike rate, around 0.12 of a cycle relative to their best frequency (frequency of sound the cell is most sensitive) (Grothe, 2003). It is assumed that this is the

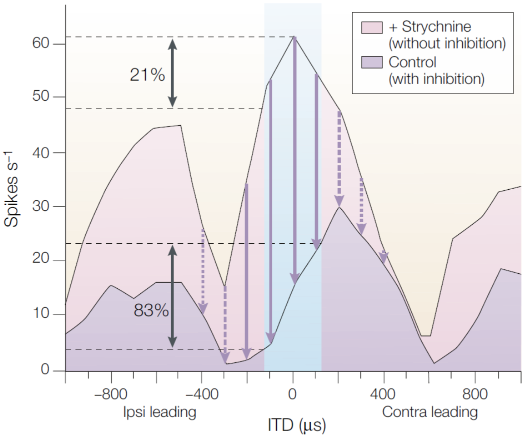

Figure 1.3: Mean discharge rate of a single gerbil's MSO neuron with and without strychnine inhibition blockage to a pure tones with ITD. The maximal firing rate of the control group occurred around ITD equal to 0.12 of the pure tone's cycle, outside of the ecological relevant ITD range (blue shading). While the cells with strychnine blocked inhibition exhibited a shift of the maximum towards zero ITD, and its increase. Picture taken from Grothe (2003).

most critical region in ITD decoding, which means that this peak is outside of the analyzed animals' ecological relevant ITDs. The same in vivo recordings (Grothe, 2003) show that when the inhibitory inputs are strychnine blocked, the peak shifts to zero ITD (Figure 1.3).

Chapter 2

PSYCHOACOUSTICS OF BINAURAL HEARING

In the previous chapter, we have described the physiology of human binaural hearing briefly. In this chapter, we will focus on how spatial hearing works from a psychology point of view. Branch of acoustics, psychoacoustics have studied this topic thoroughly since late of the 19-th century (Blauert, 1997). We will discuss results from some of the experiments briefly here, elaborate description of the binaural relevant experiments can be found in Blauert (1997).

2.1 Sound localization cues

The human auditory system receives the incoming sound through two ears located on opposite sides of the head. Temporal and spectral disparities between the signals in the two ears provide cues about the spatial location of the incoming sound (Figure 2.1). These cues are the differences in times and levels between the ears. The differences in time are known as interaural time differences (ITDs), in case of pure tones or harmonic complexes as interaural phase differences (IPDs). The differences in level are called interaural level differences (ILDs). ITD and ILD enable the sound to be localized on the horizontal plane. Moreover, reflection and diffraction from the upper body torso and pinna change the spectral composition of the incoming sound (Batteau, 1967; Blauert, 1997; Lopez-Poveda and Meddis, 1996; Steinhauser, 1879). These changes are unique to the angle of sound incidence and allow localization in the vertical plane. This cue is present even with one ear; therefore, it is called the monaural cue.

According to the duplex theory of sound localization, ITD and ILD are combined to localize a sound on the horizontal plane. The experiments first conducted by Lord Rayleigh lead to the tentative idea that for pure tones, ITD would dominate sound localization at low frequencies (originally only below 125 Hz) while ILD would dominate localization at higher frequencies (originally from 500 Hz onwards; Rayleigh (1907)). In free field conditions, pure tones can be localized on the basis of ITD at frequencies up to about 1.5-2 kHz and can be localized on the

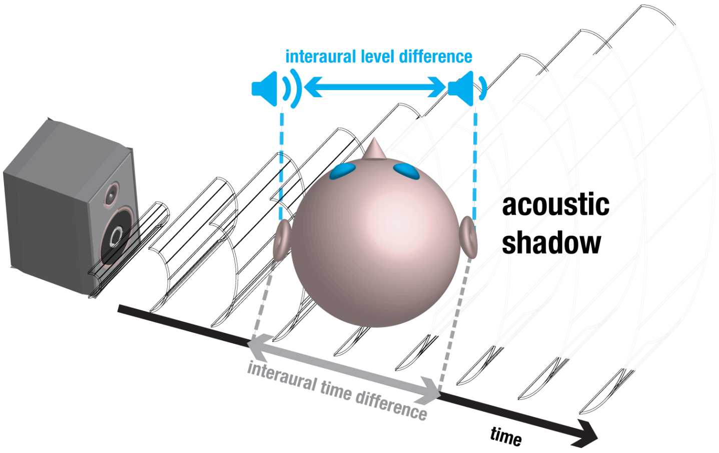

Figure 2.1: Illustration of ITD and ILD occurrence in free-field. Temporal and spectral disparities between the signals in the two ears provide cues about the spatial location of the incoming sound.

basis of ILD from about 1.5-4 kHz (Mills, 1960; Stevens and Newman, 1936). When noise bands were used, it was found that ITD dominates localization at low frequencies up to the boundary estimated at between 1.5 and 2.5 kHz (Blauert, 1997; Wightman and Kistler, 1992). A recent study by Hartmann *et al.* (2016) showed that ILD's natural salience in a free field might be high even at frequencies below 1 kHz.

2.1.1 Interaural time difference

For pure tones, the smallest detectable change in ITD depends mostly on the frequency (Figure 2.2). While the threshold can be as low as 10 μs at 0.8 kHz, it increases rapidly when the frequency is increased up to 1.2 Hz or is reduced to 0.2 kHz (Brughera *et al.*, 2013; Klumpp and Eady, 1956; Zwislocki and Feldman, 1956). Due to the direct mathematical bond between time and phase, in case of pure tones or harmonic complexes, IPD is often used instead of ITD to describe temporal differences. Physically, pure tones' IPD is limited to approximately 1.5 kHz, when the temporal difference exceeds the π phase limit, i.e., the wavelength of the incoming sound is smaller than the diameter of the head. If the listeners were presented with an artificially modified signal with a temporal difference equal to π-limit, they heard the sound with an ambiguous location between the left and right ear, or also oscillating location between those two (Yost, 1981).

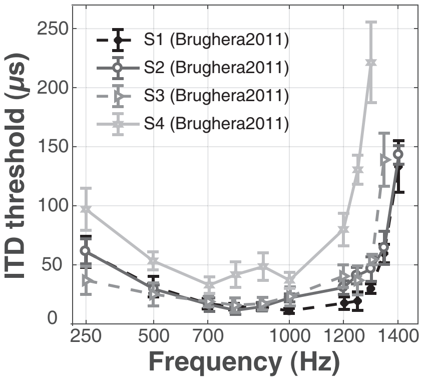

Figure 2.2: Human ITD threshold data for pure tones reproduced from Brughera *et al.* (2013)

2.1.2 Interaural level difference

In contrast to the ITD sensitivity, the ILD threshold (Figure 2.3) is almost constant (0.5 dB) across frequencies between 0.5 and 8 kHz with a peak at 1 kHz (Grantham, 1984; Mills, 1960; Yost and Dye, 1988). The ILD occurs mainly due to the acoustic shadow of the head and body torso, and sound's attenuation during propagation in the air. Therefore, for sound sources far from the listener, ILD occurs on higher frequencies (approx. 2 kHz) where the wavelength of the incoming sound wave is considerably smaller than the diameter of the head (Blauert, 1997). However, for sound sources near the listener's head, ILD occurs also at low frequencies due to the attenuation of the human head (Blauert, 1997).

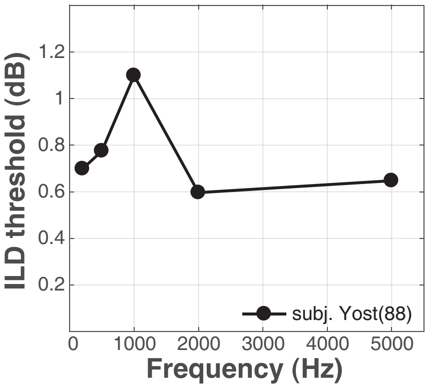

Figure 2.3: Human ILD threshold data for pure tones reproduced from Yost and Dye (1988)

2.1.3 Monaural cue

While ITD and ILD cues are according to duplex theory responsible for human localization in the horizontal plane, there are some caveats in horizontal localization they cannot account

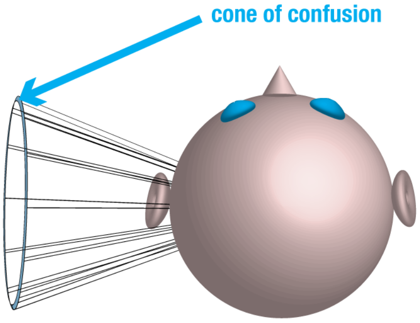

Figure 2.4: Illustration of ”cone of confusion”, the ambiguity of localizing of the sound source while using only ITD and ILD. The cone is centered on the interaural axis expanding from each ear entrance, representing on its surface locations of the same interaural differences (Wallach, 1939; Blauert, 1997; Mills, 1972).

for. For instance, experimental data shows that even listeners with hearing loss on one of the ear are capable, even-tho with limited accuracy, to localize sound sources (Slattery and Middlebrooks, 1994). Another example of the duplex theory limitation is the so-called cone of confusion (Figure 2.4). The cone is centered on the interaural axis expanding from each ear entrance, representing on its surface locations of the same interaural differences (Wallach, 1939; Blauert, 1997; Mills, 1972). In the case of a horizontal plane, it means that while using only ITD and ILD cues the listener would hardly differentiate between the sound source in front and behind him, which is also called front-back confusion (Blauert, 1997). Although front-back confusion is sorted partially by the asymmetry of ears positions and primary asymmetry of pinna towards the listeners' back (Blauert, 1997), another problem emerges in the vertical plane, where we need another cue to help us decode sound source position on the cone of confusion. This ”another” cue is the monaural cue.

The monaural cue, sometimes also referred to as a spectral cue (Langendijk and Bronkhorst, 2002), affects spectral components of incoming sounds above 6 kHz (Langendijk and Bronkhorst, 2002). Its spectral characteristics vary with the sound source's elevation, which inevitably helps to decode spatial information in a vertical plane (Blauert, 1997; Langendijk and Bronkhorst, 2002). These spectral changes are mainly caused by a sound reflection on the folds of pinna (Tianyi Yan and Jinglong Wu, 2007), and reflection from shoulders (Campbell *et al.*, 2008).

2.1.4 Head-related transfer function

All cues mentioned above: ITD, ILD, and monaural cue; can be obtained by measuring head-related impulse responses (HRIRs) in both ear canals for sound sources in various positions,

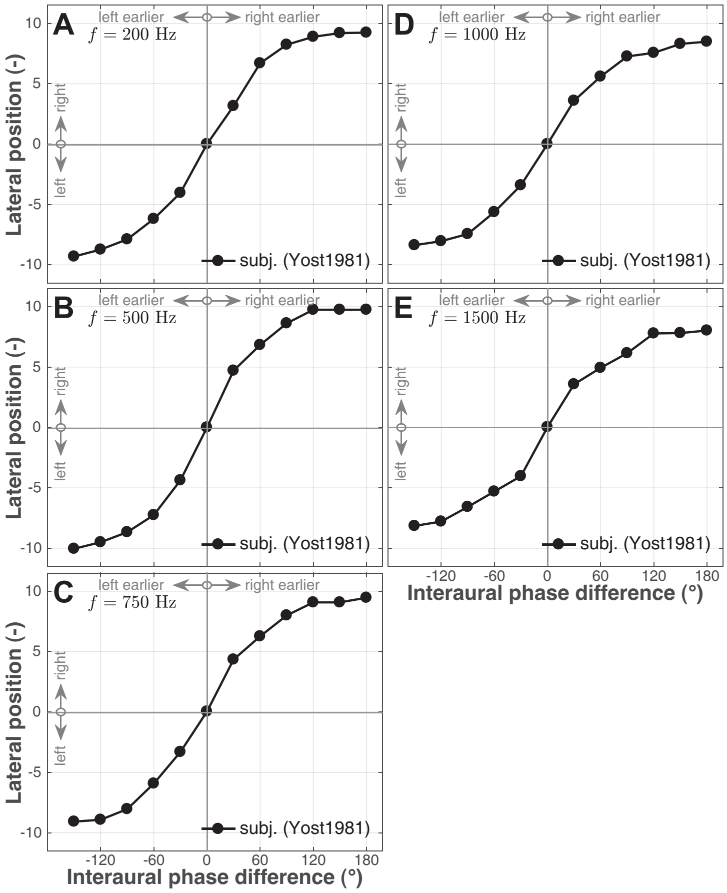

Figure 2.5: Lateralization data of pure tones with IPD reproduced from Yost (1981). The pure tones had frequencies of 0.2, 0.5, 0.75, 1, and 1.5 kHz, each depicted in separate panel.

which gives us unique pair of HRIRs for each sound source position. This pair is unique for the spatial position and to the listener, mainly due to alternating head sizes and pinna shapes (Watanabe *et al.*, 2007). We can obtain a pair of head-related transfer functions (HRTFs) by calculating the Fourier transform of the HRIRs. The pair of the HRTFs can then be used as filters for the left and right sound channel to move a monaural sound in virtual auditory space (VAS) to the position of the HRTFs had been recorded. Perfect results in the listener's spatial sensation can be only be achieved if the listener's personal HRTF had been utilized. Otherwise the spatial experience deteriorates, and audible artifacts might occur (Pec *et al.*, 2008).

Therefore, in VAS, if we want to move a sound, we have to process two sound channels, and if the HRTF is not personalized, we might introduce the audible degradation of the positioned sound. However, if we use a differential-head-related transfer function (DHRTF), we would need only one channel to process, and the audible artifacts will be reduced significantly for not personalized HRTF (Storek, 2014). The DHRTF is created by dividing the left ear HRTF with the right ear HRTF, producing the DHRTF used for the left ear, and vice versa for DHRTF for the right ear. The DHRTF is applied to the sound channel to which the virtual sound location is closer. By processing only one channel, the DHRTF reduces computational demands in VAS and provides an unfiltered signal to the other ear, improving the audible quality (Storek, 2014). However, due to notches presented in HRTF spectra and the mathematical processing used to create DHRTF, it can produce other audible artifacts. These artifacts can be reduced by conditioning each DHRTF by several methods, some of them are mentioned in Storek *et al.* (2016).

2.2 Headphones listening

In a free field listening, the ITD, ILD, and monaural cues naturally occur altogether. For an experimental procedure, however, it is often desirable to analyze each cue separately, which can be achieved by using a pair of headphones instead of loudspeakers. Listening through headphones often results in an auditory sensation localized within the head. The lateral position of this sensation is known as lateralization (Blauert, 1997; Moore, 2003). The lateralization can be moved towards one of the ears by manipulating ITD, ILD, or both (Sayers, 1964; Yost, 1981; Zhang and Hartmann, 2006).

The lateralization experiment of Yost (1981) with pure tones with IPD shows that the sensation of the lateral displacement as a function of IPD is invariant to the pure tone's frequency, which implied preference of IPD over ITD for pure tones (Figure 2.5). However, later experiment of Zhang and Hartmann (2006) showed that the results of Yost (1981) might be biased because of the test procedure, and proved that listeners prefer ITD over IPD in their experiment (Figure 2.6). An extra-aural experiment by Hartmann *et al.* (2016) also confirmed this fact.

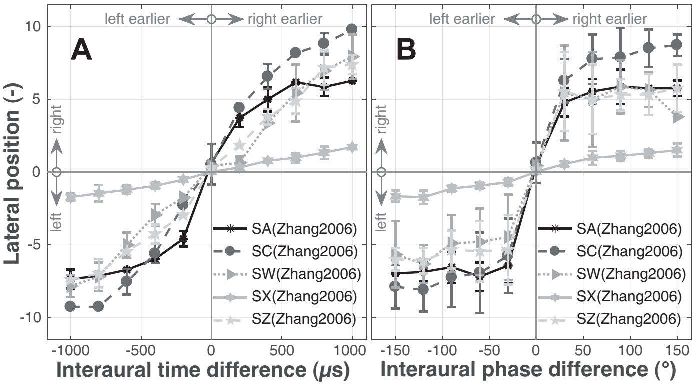

Figure 2.6: Lateralization data of pure tones as a function of ITD (Panel A) or IPD (Panel B) reproduced from Zhang and Hartmann (2006).

If the tones in the two ears have slightly different frequencies, it may cause binaural beats. The perceived position of the sound oscillates between the left and right ear for a frequency difference between the tones of about 2 Hz (Moore, 2003). Siveke *et al.* (2008) introduced a broadband binaural beat stimulus called phase warp that can be created by an up/down circular shift of the phase spectrum of the noise in one channel by a defined "beat" frequency. When such a sound is listened to through headphones, it produces the sensation of a sound source rotating around the listener's head. The beat frequency gives the frequency of rotation. The sensation of rotation disappears when the beat frequency exceeds 10 Hz (Siveke *et al.*, 2008). The sensation then changes to roughness.

Chapter 3

MODELING OF BINAURAL HEARING

In the previous chapters, we reviewed a principal of binaural hearing from physiological to psychoacoustic point of view. The main task of binaural modeling is to accommodate knowledge from one or both aforementioned and simulate their specific behaviour employing computer algorithms. Binaural models, according to Blauert (2013), can be divided into groups by several means.

Based on the system's behaviour, they are simulating, i.e., spatial direction/lateralization, distance, room properties, and many others. Based on sound scene complexity, simple models often account for single sound sources in an anechoic environment, while more elaborate models can account with several sound sources in rooms with reverberations. Furthermore, based on how the models simulate the human performance, whether they are inspired by a hearing system's physiological properties, or simulate the performance using signal processing independent of the physiological truth. The physiological nature can be simulated to some extent by the models. These models simulate each neuron and connection separately with known physiological parameters. However, they are often limited by the computational complexity and the fact that the neural response (activity) has to be fitted arbitrary to be able to reproduce psychoacoustical data (Encke and Hemmert, 2018). Therefore, standard practice in binaural modeling uses abstraction and simulates the binaural system function inspired by a physiological data rather than simulating complete underlying processes (Cherry and Sayers, 1956; Stern and Colburn, 1978; Gaik, 1993; Breebaart *et al.*, 2001; Moore, 2003; Pulkki and Hirvonen, 2009; Dietz *et al.*, 2008, 2009, 2011; Blauert, 2013; Takanen *et al.*, 2014). In that case, we are talking about the functional binaural models.

In this thesis, from all models mentioned above, we focus on the simple, functional binaural models of lateralization. Therefore, in the following chapter, a brief review of these models will be provided. These models can be divided into 'families' based on their approach (theory) of binaural part simulations. The first, two-channel binaural models were first introduced by von Békésy (1930). These account to two nuclei in the human brain whose population activity

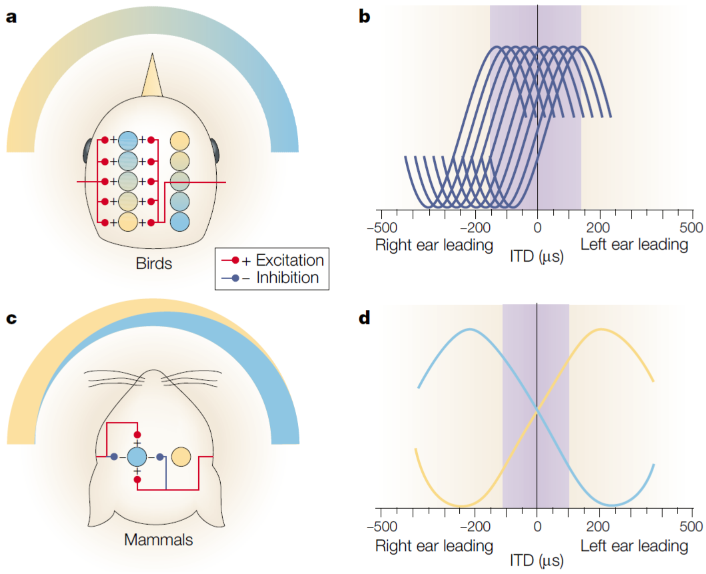

Figure 3.1: Illustration of different strategies of decoding ITD in birds and mammals. **Panel a** shows bird ITD-decoding structure, which resembles the function similar to Jeffress delay line (Jeffress, 1948). The coincidence detectors (circles) in both brain hemifields receive excitatory inputs from both ears connected through delay lines (induced by different axonal lengths). The coincidence detector neuron responds maximally if the ITD of the incoming sound is compensated by the internal neural delay provided by delay lines. Axonal length (delay line) arrangement, which covers ecologically relevant ITD range, creates a map of horizontal auditory space (Grothe, 2003). **Panel b** shows the maxima of the ITD functions from different coincidence neurons, which are evenly distributed across physiologically relevant range (shaded area). **Panel c** shows a mammalian ITD decoder, medial superior olive (MSO), as experimentally observed in gerbils and guinea-pigs. MSO neurons tuned to the same best frequency shows the response to the same ITD, and the ITD functions have their highest slope near zero ITD, as shown in **Panel c**. Sounds positioned to the left introduce the higher activity in the right MSO, and the stimuli positioned to the right activates more the left MSO (Grothe, 2003). The relative activity of both MSOs then decodes the auditory space. The picture was taken from Grothe (2003)

is compared to estimate the sound direction. The second family of models were introduced by Jeffress (1948), which is commonly named in literature Jeffress delay line models (Moore, 2003). These account with a structural array of connections with specific delay, which resembles into function similar to correlation.

3.1 Jeffress delay line models

A possible ITD detection mechanism was proposed by Jeffress (1948). He suggested that neural discharges coming from the left and right ears propagate through delay lines to coincidence detectors (Figure 3.1 Panel a). The detectors fire if the discharges from the left and right ears arrive within a short time window, i.e., when the input delay line connections effectively cancel out the ITD. The theory assumes that each neuron is tuned to a specific ITD, and its activity is highest for this ITD. This ITD characteristic to the neuron under examination is in literature called neuron's best or characteristic ITD (Grothe, 2003). For example, with the sound source next to the right ear, the sound first enters the right ear and then left ear with some ITD. In the proposed mechanism, the right ear signal propagates further through the delay line before it meets the signal from the left ear on the laterally displaced coincidence detector–specific delay of this detector from both ears than effectively decodes the spatial angle of the sound source.

Later, it was proven that the Jeffress mechanism could be simulated using interaural cross-correlation, if delay lines and coincidence cells consist of larger cell populations (Cherry and Sayers, 1956; Stern and Colburn, 1978). The Jeffress delay line mechanism was subsequently extended to account for the detection of ILD (Gaik, 1993; Breebaart *et al.*, 2001).

From the neurophysiological point of view, the Jeffress delay line theory expects for a single best frequency a population of neurons with best ITDs spread uniformly across ecologically relevant ITDs (see Figure 3.1 Panel a, lower right plot). A mechanism similar to the Jeffress delay line was found in the brain of the barn owl (Carr and Konishi, 1990), but not yet in any mammalian species.

3.2 Two-channels models

Recently, the delay line mechanism of Jeffress has been questioned in some neurophysiological studies, which have shown that neurons in the mammalian MSO respond maximally for ITD corresponding to a 45-degree IPD (0.125 of the cycle) (Grothe, 2003; McAlpine and Grothe, 2003). For a given best frequency, the best ITD was not only not distributed uniformly across ecologically relevant ITDs (see Figure 3.1 for comparison) but also the maximal responses laid outside ecologically relevant ITD range (Grothe, 2003). Those findings effectively contradict with the main premises for Jeffress delay line model theory in mammals. These neurophysiological studies propose that the difference in the relative spike rate between MSO neurons in the left and right sides of the brain encodes the sounds' spatial direction. In humans, the magneto-encephalography study by Salminen *et al.* (2010) gave evidence for such a hemifield rate-code of auditory space.

Models based on the hemifield rate-code are sometimes referred to as count-comparison (Pulkki and Hirvonen, 2009; Takanen *et al.*, 2014; Encke and Hemmert, 2018). This term was initially introduced by Colburn (1978) to describe the principle of the model designed by von Békésy (1930).

3.2.1 Binaural model of von Békésy

The von Békésy (1930) model assumes that each ear innervates the same neurons and that each neuron may be tuned left if the signal from the left ear arrives shortly before the signal from the right ear, or tuned right in the opposite case. The firing activity of neurons tuned left, and right is then compared to give the sound source direction. Since the number of excited neurons increases with a rising sound level, this model accounts for the time-intensity trading. Van Bergeijk (1962) later adjusted the model to account for the fact that MSO and LSO are composed of paired units placed symmetrically relative to the median plane.

3.2.2 Binaural model of Pulkki

One of the first rate-code models based on neurophysiological data is the functional model of the MSO and LSO designed by Pulkki and Hirvonen (2009). This model uses signal processing to reproduce data from neurophysiological studies.

The model accounts for the role of inhibition in the MSO, which was proposed by Grothe (2003), and incorporates it into the functional count–comparison model. As a front end, a simple peripheral ear model consisting of a gammatone (Johannesma, 1972) filter bank was employed.

Takanen *et al.* (2014) extended the model by integrating the MSO and LSO outputs into a visual map, which allowed a direct comparison between the listening test results and the model predictions. Besides, the model was extended with a hypothetical wideband MSO, which would account for the detection of ITD in the envelope of high-frequency sounds.

3.2.3 Binaural model of Dietz

Dietz *et al.* (2008) proposed in their model that the human auditory brainstem can effectively encode and decode the IPDs from the interaural transfer function (ITF) (Blauert, 1997). By filtering the peripheral model output using two parallel complex bandpass filters, Dietz obtained two ITFs that corresponded to the fine structure and the envelope of the IPD. The firing rates of the units simulating left and right MSOs are then calculated. When the fine structure and the ITF envelope are used, the model can account for binaural masking level differences (BMLDs) and lateralization data. With an extra-optimal detector, it has been proven that it could simulate experimental data on broadband binaural beats (phase warp) (Siveke *et al.*, 2008). The model was further improved by adding a module for calculating the lateralization based on ILD. The improved model was utilized as a front-end for a sound direction estimate of concurrent speakers in a binaural signal (Dietz *et al.*, 2011). Dietz *et al.* (2009) also extended the original IPD model with a quantitative estimate of the perceived lateralization, which successfully simulated higher auditory processing. The new model combines the temporal lateralization cues from the fine structure and the envelope of the binaural input signal, giving a single value for the overall perceived lateralization.

3.2.4 Binaural model of Encke

In addition to the phenomenological rate-code models mentioned above, Encke and Hemmert (2018) presented their physiologically-plausible, spiking neuron network model of the mammalian MSO. The authors used two methods to decode spatial information. The first method – the linear opponent decoder – was able to mimic the ITD threshold data, but its predictions were dependent on the stimulus's overall sound pressure level. The second method was based on a simple artificial neural network (ANN), with inputs from the spiking outputs of MSO and auditory nerve fiber models. With ANN, the model was able to predict static ITD imposed on the amplitude-modulated tone and speech stimuli. It was also able to track the transient change of ITD for a sine sweep stimulus.

3.3 Open topics of binaural modeling

The Jeffress family of models has been used successfully to predict human psychoacoustical data (Braasch *et al.*, 2013; Breebaart *et al.*, 2001; Colburn, 1977; Faller and Merimaa, 2004; Lindemann, 1986; Prokopiou *et al.*, 2017). However, from a physiological point of view, the existence of such a neural circuit in the mammalian brain is questionable (Grothe, 2003; Grothe *et al.*, 2010). Although, there is evidence for a rate-code in the mammalian brain. The earlier presented rate-code models have not quantitatively explained the psychoacoustical data (Dietz *et al.*, 2008, 2009, 2011; Encke and Hemmert, 2018; Pulkki and Hirvonen, 2009; Takanen *et al.*, 2014).

This thesis aims to reduce the gap between neurophysiology and psychophysics by introducing functional rate-code binaural models of MSO and LSO which will meet three main criteria:

1. they will be based on current neurophysiological findings,
2. their output will give a quantitative representation of subjective lateralization based on ITD or ILD at a corresponding frequency,
3. they will have low structural complexity and low computational complexity for low power applications.

The model designed by Pulkki and Hirvonen (2009) satisfies Criterion 1 and 3 but gives only the relative firing rate at the model output. Enhancing the model by a visual map Takanen *et al.* (2014) solved this problem, but significantly increased the overall structural complexity of the model.

The IPD model of Dietz *et al.* (2008, 2009, 2011) fulfills all three criteria, but calculating IPDs and ILDs from the ITF might be considered artificial.

The model of Encke and Hemmert (2018) satisfies Criterion 1 perfectly, as it is a physiological model, which also limits the model in computational complexity (Criterion 3). The model also accurately predicts human discrimination data with pure tones with ITD, and in other experiments, it predicts ITD. However, there is no mapping of this data to subjective lateralization.

For the reasons mentioned here, we propose new functional rate-code models of the human auditory brainstem, which will be presented in the following chapters.

Part II

Binaural Models Design and Verification

Chapter 4

DESIGN OF MEDIAL AND LATERAL SUPERIOR OLIVE MODELS

The proposed binaural interaction model is composed of two main parts: peripheral and binaural (see Fig. 4.1). The peripheral part of the model simulates the function of the auditory periphery. It utilizes algorithms adapted from the Auditory Modeling (AM) Toolbox (Søndergaard and Majdak, 2013). The binaural part consists of original models of MSO and LSO connected to the corresponding ITD and ILD central stages. A simpler version of the model was presented in Bouse and Vencovsky (2015). The version presented here was published in Bouse *et al.* (2019) and this thesis contains the text from there with slight changes. Matlab source codes of the model are available at `http://mmtg.fel.cvut.cz/rate-code-model/`.

4.1 The peripheral part of the model

The peripheral part models the incoming sound's transformation into the average response of a population of neurons tuned to the specific characteristic frequency in the auditory nerve fiber. This part consists of three functional blocks: the outer and middle ear, cochlear frequency selectivity, and the inner hair cells.

The frequency response of the outer ear (headphone to eardrum pressure) (Pralong and Carlile, 1996) is modeled by a 512th-order finite impulse response (FIR) filter. The transfer function of the middle ear (Goode *et al.*, 1994) is then modeled by an FIR filter of the same order. The model does not account for the middle ear reflex.

The cochlear selectivity is modeled by a dual resonance nonlinear (DRNL) filter bank (Lopez-Poveda and Meddis, 2001), which divides the input signal into 70 peripheral channels. The frequency spacing of the auditory filters in the bank was set to be equal to one-half of the equivalent rectangular bandwidth (ERB) (Moore and Glasberg, 1983), which is calculated us-

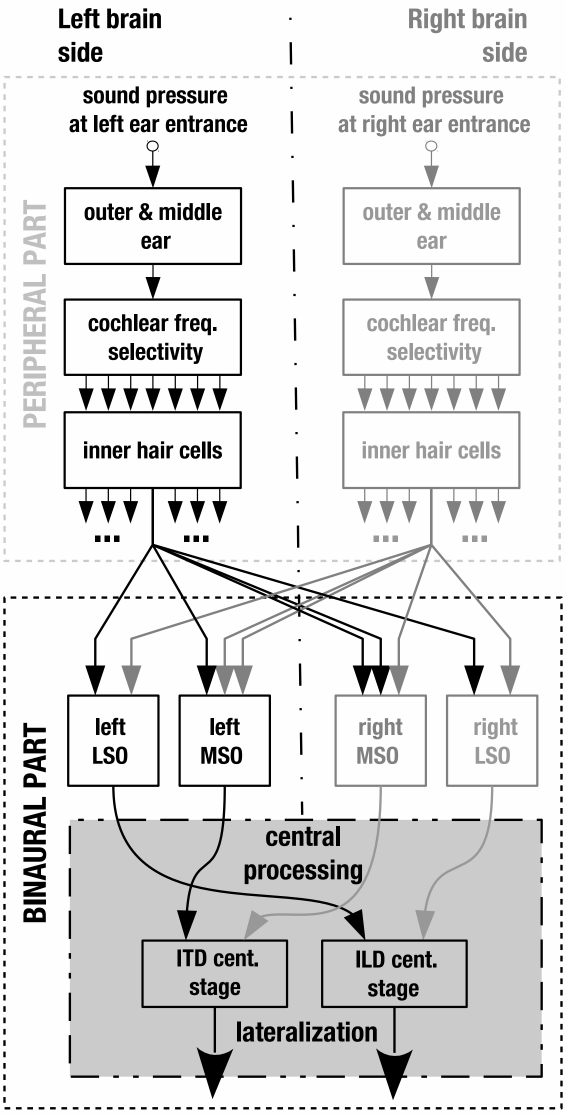

Figure 4.1: Schematic diagram of models of MSO and LSO. The diagram is divided vertically by a slashdotted line to account for the left and right sides of the brain, and horizontally by dotted boxes to split the models into the peripheral and binaural parts. The gray part highlights processing in the higher stages of the brain.

ing Equation (1.1). For all simulations, the central frequencies (CFs) of the DRNL filter bank were set to be in the range between 0.1 and 14 kHz. In each auditory filter of the bank, the signal propagates through two independent processing pathways: linear and nonlinear. The two paths are joined at the output, and the signals from the two paths are added together. The nonlinear processing path dominates the output at low signal levels and decays with the increasing signal level until the output becomes mostly dominated by the linear processing path (Lopez-Poveda and Meddis, 2001). We chose this type of model since it accounts for the compressive nonlinearity observed in the input/output functions of the basilar membrane response (Lopez-Poveda and Meddis, 2001).

The block simulating the mechano-electrical transduction by the inner hair cell and the auditory nerve fiber complex consists of a half-wave rectifier followed by a low-pass filter (LPF). While the half-wave rectification accounts for the actual mechano-electrical transduction, the low-pass filtering simulates the loss of phase locking of the neuronal signal to the fine structure of the incoming wave for frequencies above 1.5 kHz (Bernstein and Trahiotis, 1985; Weiss and Rose, 1988). In the present study, a low-pass filter with the same parameters as in studies by Breebaart *et al.* (2001) and Dietz *et al.* (2008) is utilized, i.e., a fifth-order Butterworth infinite impulse response (IIR) filter with a cut-off frequency of 760 Hz.

4.2 Binaural part of the model

The binaural part consists of two separate original computational models mimicking the medial and lateral superior olives (MSO and LSO). Both models follow the rate–code principle (von Békésy, 1930; Colburn, 1978; Dietz *et al.*, 2008; Encke and Hemmert, 2018; Pulkki and Hirvonen, 2009; Takanen *et al.*, 2014). There are separate MSO and LSO models for each side of the brain. Each model receives information from the left and right peripheries. The lateralization based on IPD/ITD (MSO) or ILD (LSO) is calculated in the corresponding central stages, where a comparison is made between the activity in the models for both sides of the brain. In the present study, the outputs of MSO and LSO are not combined like they are in the auditory pathway (Grothe *et al.*, 2010).

The internal processing of the models is symmetrical for both sides of the brain. Therefore, the signal coming from the periphery at the same side as analyzed in the MSO/LSO model will be referred to as ipsilateral, and the signal from the opposite side will be referred to as contralateral.

4.2.1 Model of the medial superior olive

A schematic diagram of the MSO models for both sides of the brain (black line = left, gray line = right) is depicted in Figure 4.2. The MSO model is an excitation-inhibition type, with two excitatory inputs from the ipsilateral and the contralateral sound peripheries, and one inhibitory input from the contralateral sound periphery. A low pass filter first processes each input; the third-order low-pass Butterworth filter with a cut-off frequency $f_{cut} = 1.1$ kHz and gain $G(f) = 1/\sqrt{(f/f_{cut} + f^6)}$ then reduces the model sensitivity to ITDs for frequencies higher than about 1.1 kHz. Although this filter has a higher cutoff than the preceding filter in the

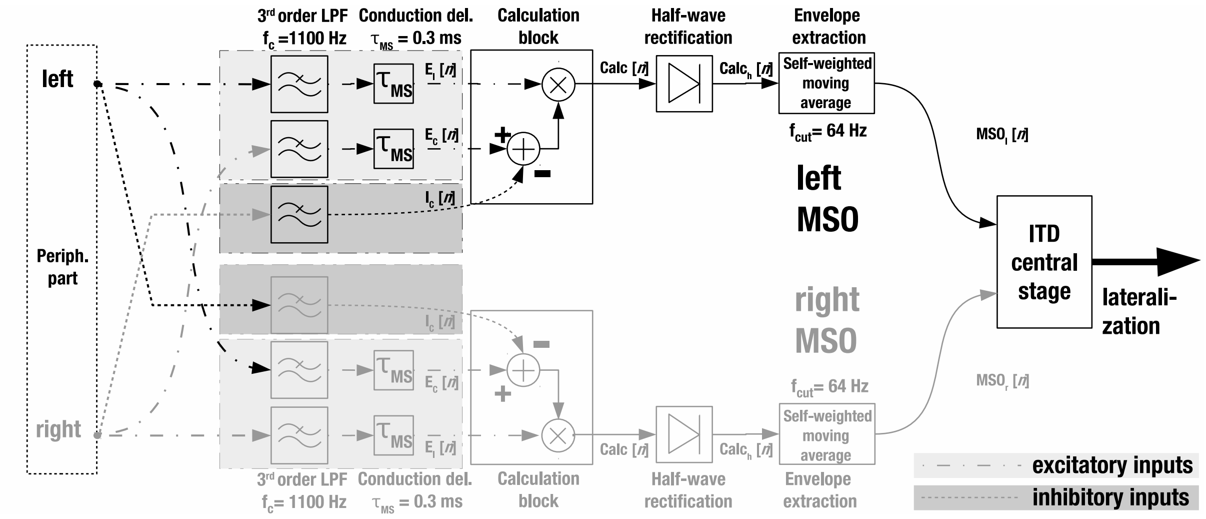

Figure 4.2: Schematic diagram of the MSO models for the left (black connection) and right (gray connection) peripheries with the ITD central stage. Each MSO model has three inputs, two excitatory inputs (light gray background) from both peripheries, and one inhibitory input (dark gray background) from the contralateral periphery.

peripheral part, its function is essential to account for the decreasing ITD discrimination at frequencies > 1 kHz.

The physiology data from bats' MSO (Grothe, 1994) and gerbils' MSO (Brand *et al.*, 2002; Roberts *et al.*, 2013) indicate that contralateral inhibition can precede ipsilateral excitation. A small constant delay τ_{MS} of 0.3 ms is inserted into each excitatory signal. In contrast, the excitatory inputs to the MSO from both peripheries have very similar overall conductance delays (Grothe, 2003). Therefore, there is no additional time lag between the ipsilateral and contralateral excitation signals in the MSO model.

The calculation block is the first binaural relay in the model. In the first step, it subtracts the preprocessed contralateral inhibition input I_{c} from the contralateral excitation input E_{c}. Then it calculates the mutual signal power of the product and ipsilateral excitation signal E_{i}. Thus, the output of the calculation block is given by

$$\mathrm{Calc}[n] = E_{\mathrm{i}}[n](E_{\mathrm{c}}[n] - I_{\mathrm{c}}[n]), \tag{4.1}$$

where n is the sample number. Since a neural signal cannot be negative, the output of the calculation block is half-wave rectified, and this signal is then denoted with subscript h.

The calculation block output contains several of sharp peaks followed by troughs, which can cause transient variation of the MSO model output even for stationary input signals. Therefore, an envelope is calculated from the half-wave rectified MSO signal in a self-weighted moving

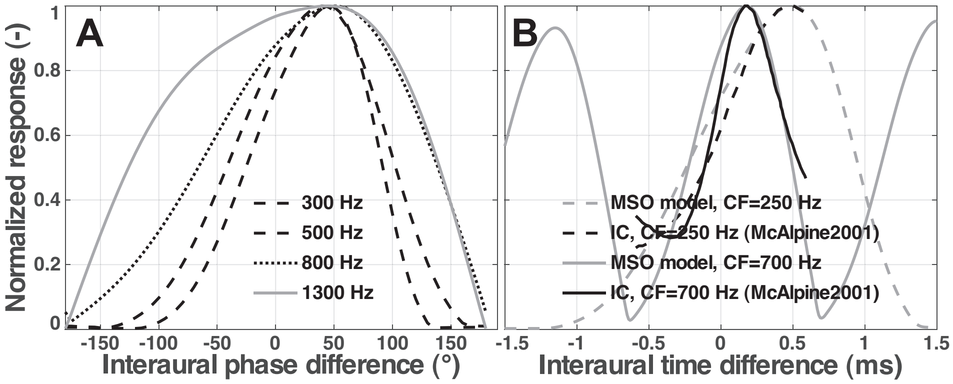

Figure 4.3: Panel A The normalized responses of the left MSO model to the pure tones of varying frequency and IPD. **Panel B** The normalized responses for two CFs (250 and 700 Hz) of the left MSO (gray line) with additional side-lobes and pooled responses from guinea pig's Inferior Colliculus (black line), reproduced from McAlpine *et al.* (2001), to interaurally delayed broad band noise.

average block (Pulkki and Hirvonen, 2009). The envelope is calculated using the formula

$$\mathrm{MSO}_{\mathrm{l\ or\ r}}[n] = \frac{\mathrm{Calc}_h^3[n] * H_1[n]}{\mathrm{Calc}_h^2[n] * H_1[n]}, \tag{4.2}$$

for

$$H_1[n] = (1 - \exp(-1/(f_s\tau_1)))\exp(n - 1/(f_s\tau_1)), \tag{4.3}$$

where $\mathrm{Calc_h}[n]$ is the half-wave rectified output of the calculation block, $H_1[n]$ is the impulse response of the first-order IIR filter, τ_1 is the time constant of the filter, f_s is the sampling frequency, and $*$ denotes convolution. The time constant of the filter τ_1 (2.5 ms) corresponds to the 64 Hz cut-off frequency, which simulates the relative decrease in temporal binaural resolution on higher frequencies reported by Siveke *et al.* (2008).

ITD central stage

In this thesis, we combine outputs from both sides of the brain in an ITD central stage. This stage is phenomenologically motivated to transform the MSO model outputs into the subjective lateralization scale. In the first place, we have to consider how the left and right MSO models' outputs vary in relation to the IPD. The right MSO model output is at a maximum when the signal at the left ear has an IPD of minus 50 degrees (0.15 of the cycle); at the same moment in time, the left MSO output is minimal, and vice versa for a 50-degree IPD. This behavior is consistent with the physiology of the mammalian MSO, where the maximal firing rate is observed around 45 degrees (0.125 cycles) when the contralateral ear is leading (Grothe, 2003). The maximum occurs around the 50-degree IPD, independently from the input signal frequency (see Figure 4.3 Panel A). The MSO model behaves in this way due to the algorithm that is used.

The algorithm multiplies the ipsilateral and contralateral inputs and then subtracts the result by the 0.3 ms shifted contralateral input multiplied with the ipsilateral input. This processing gives the maximum MSO response at about the same IPD for ecological ITD-relevant frequency range if the peripheral ear model contains half-wave rectification and low-pass filters limiting the phase locking. The LP filtering shapes the fine-structure of the half-wave rectified signal in a way that then produces the desired result. The MSO model also shows similar behavior to that measured on neurons in guinea pig's inferior colliculus neurons (McAlpine *et al.*, 2001), i.e., the broadest ITD function for low CFs with peaks at high ITDs and sharper tuning for high CFs with peaks at low ITDs (see Figure 4.3 Panel B). For both CFs, the predicted responses have a close to zero ITD shallower slope than the physiological data (McAlpine *et al.*, 2001).

The MSO outputs are half-wave rectified and, if one of the MSO models has a zero output, the other MSO model's output is also set to zero. The two ratios between the left and right MSO models are then computed as follows:

$$r_\mathrm{R}[n] = \frac{\mathrm{MSO_l}[n]}{\mathrm{MSO_r}[n]}\,, \qquad\qquad r_\mathrm{L}[n] = \frac{\mathrm{MSO_r}[n]}{\mathrm{MSO_l}[n]}\,, \tag{4.4}$$

where $\mathrm{MSO_l}$ is a signal from the left MSO model, and $\mathrm{MSO_r}$ is a signal from the right MSO model. Ratio r_R is inverse to r_L; this ensures that, for non-zero MSO model outputs, one of the ratios will always lie between zero and one. If the sound is perceived on the left, the ratio with subscript "L" will be larger than the ratio with subscript "R", symmetrically for the sound perceived on the right. Nevertheless, the larger ratio can be used as a bias between the left or right side, and the smaller ratio is better suited for evaluating lateralization, because it is bounded between zero and almost one. The lateralization map ranging from minus one (left ear) to plus one (right ear) is obtained by subtracting unity from r_L if the lateralization is "right-sided" and by subtracting r_R from unity if it is "left-sided":

$$\mathrm{L_{MSO}}[n] = \mathrm{sgn}\left(r_\mathrm{L}[n] - r_\mathrm{R}[n]\right)\left(\min\left(r_\mathrm{R}[n], r_\mathrm{L}[n]\right) - 1\right) + u[n], \tag{4.5}$$

where $\mathrm{L_{MSO}}$ is the lateralization predicted by the MSO model, $u[n]$ is the MSO internal noise, sgn is the signum operator, and min is the minimum operator. The model's overall sensitivity to the ITDs is limited only by the mathematical operations within the models. Thus, to match this sensitivity with human psychoacoustical data, the internal Gaussian noise $u[n]$ is added into all channels of the calculated lateralization. The variation of this noise was set experimentally to match model performance with the ITD discrimination data of Brughera *et al.* (2013) at 1 kHz, and it is constant for all the channels.

4.2.2 Model of the lateral superior olive

Figure 4.4 is a schematic diagram of the LSO models for both brain sides (black lines = left, gray lines = right). The LSO model is an excitation-inhibition type, with the excitatory input from

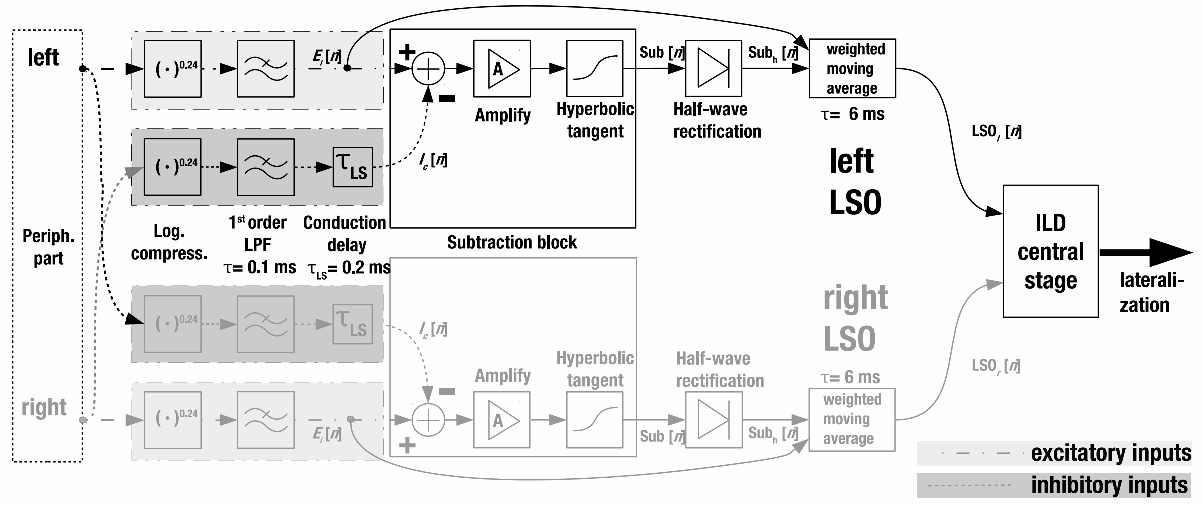

Figure 4.4: Schematic diagram of LSO models for the left (black connection) and right (gray connection) sides of the brain with the ILD central stage. The LSO model on one side has two inputs, an excitatory input (light gray background) from the ipsilateral periphery, and an inhibitory input (dark gray background) from the contralateral periphery.

the ipsilateral periphery and the inhibitory input from the contralateral periphery. Human sensitivity to the intensity disparities between the two ears has repeatedly been reported as logarithmically dependent (Moore, 2003). The inputs of the LSO models are, therefore, first compressed by a power of 0.24. The compression has a similar effect as the logarithm of the ipsilateral and contralateral signals. This compression is used even though our cochlear model also contains compressive nonlinearity. This is similar to the LSO model in Takanen *et al.* (2014). On the basis of data from cats' LSO (Joris, 1996), the contralateral inhibitory signal is delayed by τ_{LS} equal to 0.2 ms. The function of the mammalian LSO can be characterized as a fast subtraction unit between contralateral and ipsilateral neural signals (Bures, 2012; Bures and Marsalek, 2013; Joris and Yin, 1995). A first-order IIR filter with a time constant of 0.1 ms simulates the relative speed of the system. The process of subtraction is then simulated in a subtraction block, where the contralateral inhibitory signal is first subtracted from the ipsilateral inhibitory signal, sample by sample. The product is amplified by a gain $A = 100$ and is limited between -1 and 1 by a hyperbolic tangent function. The amplification gain, together with the limitation, successfully simulates the maximum firing rate of LSO cells, which occurs around 18 dB ILD (Tollin and Yin, 2005).

Thus, the processing of the subtraction block is given by

$$\mathrm{Sub}[n] = \tanh\left(A\left(E_{\mathrm{i}}[n] - I_{\mathrm{c}}[n]\right)\right), \tag{4.6}$$

where Sub is the output of the LSO subtraction unit, tanh is the hyperbolic tangent function, E_{i} is the excitation from the ipsilateral periphery, I_{c} is the inhibition from the contralateral periphery, and A is the linear gain.

All negative samples after the subtraction block are zeroed in the half-wave rectification block. The rectification, together with the subtraction, produces ripples at the output, which are smoothed by the weighted moving average unit. The design of this unit originates from Pulkki and Hirvonen (2009), and acts as a sample and hold circuit with small RC leakage with a time constant τ_2 of 6 ms. This processing is not inspired by neurophysiology, but it helps produce the desired decrease in the model's sensitivity around 1 kHz. The LSO output after the weighted moving average unit is given by

$$\mathrm{LSO}_{\mathrm{l\ or\ r}}[n] = \frac{(\mathrm{Sub}_h[n]\mathrm{E}_\mathrm{i}^2[n]) * H_2[n]}{\mathrm{E}_\mathrm{i}^2[n] * H_2[n]}\,, \tag{4.7}$$

for

$$H_2[n] = (1 - \exp(-1/(f_s\tau_2)))\exp(n - 1/(f_s\tau_2)), \tag{4.8}$$

where Sub_h is the half-wave rectified output of the LSO subtraction unit, E_i is the excitation from the ipsilateral periphery, and $*$ denotes convolution.

ILD central stage

The output of each LSO unit is proportional to the perceived lateral displacement of the sound source on the corresponding side of the brain. In the case of low-frequency signals, there is relatively high activity in both LSO units for near-zero ILDs. This activity is caused by the delay of the inhibitory signal from the contralateral brain side. Therefore, a simple central stage calculates the lateralization based on the ILD by

$$\mathrm{L}_{\mathrm{LSO}}[n] = \mathrm{LSO}_r[n] - \mathrm{LSO}_\mathrm{l}[n] + v[n], \tag{4.9}$$

where LSO_r represents the signal from the right LSO unit, and LSO_l represents the signal from the left LSO unit, and $v[n]$ the represents LSO internal noise. Here, sensitivity to the ILD of the LSO model, similarly to the MSO model, is limited only by the internal mathematical operations, which are more than one order lower than human psychoacoustical data. We, therefore, add Gaussian noise into every channel of the calculated lateralization. The variation of the noise was chosen to match human sensitivity to changes in ILD at 1kHz (Yost and Dye, 1988), and it is constant for all channels. The ILD central stage is phenomenologically motivated, though such simple processing is more plausibly presented physiologically than the ITD central stage. The predicted lateralization ranges in the interval between -1 and 1, where -1 stands for perception near the left ear, 0 for perception near the center of the head, and 1 for perception near the right ear.

Chapter 5

VERIFYING MODELS BY SIMULATIONS

The stimulus details and the simulation parameters are described below. In the Bouse *et al.* (2019) an experiment was conducted to obtain the lateralization of narrow band noise (NBN) with IPD or ILD (see Chapter 12). The data on lateralization of pure tones with IPD or ILD, discrimination of ITD, discrimination of ILD, and discrimination of phase warp were taken from the literature (Yost, 1981; Yost and Dye, 1988; Brughera *et al.*, 2013; Dietz *et al.*, 2008).

5.1 Stimuli - verification

The stimuli parameters used for verification of the models are described in this section. The parameters were held to be the same as in the psychoacoustic experiments (Yost, 1981; Yost and Dye, 1988; Brughera *et al.*, 2013; Dietz *et al.*, 2008), and Chapter 12.

All stimuli for verification were generated using 96 kHz sampling frequency.

5.1.1 Lateralization of pure tones with IPD or ILD

The data on the lateralization of pure tones with IPD or ILD were taken from Yost (1981). The pure tones had frequencies of 0.2, 0.5, 0.75, 1, and 1.5 kHz for the experiment with IPDs, and 0.2, 0.5, 1, 2, and 5 kHz for the experiment with ILDs. They were 100-ms long with an 8-ms-long on- and off-raised-cosine ramps. The sound level was 50 dB relative to the hearing level, according to ANSI S3.6-1996 (1996). In the IPD experiments, the IPDs were varied from -150 to 180 degrees with a 30-degree step, while the ILDs were varied from -18 dB to 18 dB with a 3-dB step in the ILD experiments.

5.1.2 Lateralization of narrowband noise with IPD or ILD

Narrowband noise (NBN) with a bandwidth equal to 1 ERB (see Equation (1.1)) was used as a stimulus in the experiment (see Chapter 12 for further details) and in the simulation. Two distinct central frequencies f_c (350 and 760 Hz) of NBN were used. NBNs were generated in each simulation run in the frequency domain with random amplitude and phase in the passband frequencies. The stimuli were 100-ms long with an 8-ms-long on- and off-raised-cosine ramps. The level was set to 50 dB relative to the hearing level at the central frequency f_c, according to ANSI S3.6-1996 (1996). In the IPD experiment, the IPDs were varied from -150 to 180 degrees with steps of 30-degrees. The IPD between the left and right ear channels was created in the frequency domain: the desired IPD was imposed onto the phase spectrum of the to-be-delayed signal, after which the signal was transformed back to the time domain. In the ILD experiment, the ILDs were varied within the range from the interval from -18 to 18 dB with steps of 3 dB with -20 and 20 dB ILD in addition. The ILD was imposed on the stimuli by amplifying one channel by ILD/2 and attenuating the other channel by the same amount.

5.1.3 Discrimination of ITD

The ITD discrimination threshold of the MSO model was obtained using pure tones of the same parameters as in the experiment conducted by Brughera *et al.* (2013). The pure tones were 500-ms long (100-ms-long on- and off-raised-cosine ramps), with frequencies of 0.25, 0.5, 0.7, 0.8, 0.9, 1, 1.2, 1.25, 1.3, and 1.35 kHz, and were presented at 70-dB SPL.

5.1.4 Discrimination of ILD

Pure tones with the same parameters as in Yost and Dye (1988) were used to obtain the ILD discrimination threshold of the LSO model. The 250-ms-long pure tones (10-ms-long on- and off-raised cosine ramps) were presented at a nominal SPL of 60 dB, with frequencies of 0.2, 0.5, 1, 2, and 5 kHz.

5.1.5 Discrimination of phase warp

In the simulation of the phase warp discrimination, the parameters of the stimuli were as in the experiment conducted by Dietz *et al.* (2008). The phase warp stimuli were 1-s (20-ms-long on- and off raised-cosine ramps) long and had a 65 dB SPL. The stimulus's left channel was created in the frequency domain from noise with a constant amplitude and random phase spectra (normally distributed) in the passband. The passband of the noise ranged from 10 Hz to either 550 Hz or 1100 Hz. The right channel was created by a cyclical-frequency shift of the phase spectrum of the left channel by beat frequency f_b of the phase warp stimulus, i.e., the whole phase spectrum was shifted in frequency by f_b, and the part of the spectra which was above the passband of the phase warp was moved to the start of the passband. In order to keep to the procedure of the original experiment, binaurally uncorrelated noise with a passband from 550 Hz or 1100 Hz to 48 kHz of the same spectral level was added to the phase-warp stimulus. The phase-warp beat frequency and its bandwidth were the variables in the first part of the experiment.

In the second part of the experiment, the phase warp $p[n]$ was mixed with binaurally uncorrelated Gaussian noise $w[n]$, which resulted in a mixed signal $s[n]$, given by

$$s[n] = p[n]r + w[n](1 - r), \tag{5.1}$$

where ratio r is calculated from modulation depth m by

$$r = \frac{1}{1 + \sqrt{1/m - 1}}\,. \tag{5.2}$$

The modulation depth m was the only variable in the second part of the experiment. The dB value of modulation depth is calculated as $10\log_{10}(m)$.

5.2 Simulation procedure - verification

The model results were obtained for the stationary part of the output signals only. The input transient was omitted. Therefore, the model is not intended to be used for studying the onset dominance of human sound localization or the precedence effect.

5.2.1 Lateralization of pure tones with IPD or with ILD

In the simulation of pure-tone lateralization, the mean values were calculated from the stationary part of the ITD or ILD central stage responses. Only a single band with CF nearest to the pure tone frequency was taken into account. In the IPD experiment, only the ITD central stage response was taken into account. By contrast, in the ILD experiment, only the ILD central stage was taken into account.

5.2.2 Lateralization of narrowband noise with IPD or with ILD

During the NBN noise simulation, only one stimulus with the interaural difference from the stimuli train (see Figure 12.2) was analyzed using the MSO or LSO models. Only a single band with CF nearest to the narrow band central frequency was taken into account. Afterward, the same simulation procedure was used as had been used in the lateralization of pure tones with IPD or with ILD.

5.2.3 Discrimination of ITD

In the ITD discrimination simulation, we assumed that the human auditory system could save the "pattern" of one stimulus and compare it with the pattern of another stimulus. In this case, discrimination index d' (Sakitt, 1973) is considered as the optimal observer performance:

$$d'(A, B) = \frac{|\mu_A - \mu_B|}{\sqrt{\sigma_A \sigma_B}}, \tag{5.3}$$

where μ_A, μ_B, σ_A and σ_B are the means and the standard deviations of the output of the MSO model ITD central stage for stimulus A and stimulus B. The discrimination index is calculated only for a single band, the CF of which is nearest to the central frequency of the pure tone. The observer judgments were based on the difference in ITD between stimulus A and stimulus B; henceforth, this difference will be denoted as ΔITD. Stimulus A had ITD equal to −ΔITD/2 (lateralized to the left ear), while stimulus B had ITD equal to ΔITD/2 (lateralized to the right ear). The discriminable ITD was detected when discrimination index d' exceeded the threshold limit and, at the same time, the model predicted the correct left or right lateralization shift between the tone in the first and second intervals. The threshold limit (value of 1.14) was chosen to estimate the 79.4 percent correct point of the psychometric function, according to Hacker and Ratcliff (1979). The same point of the psychometric function was targeted in the original listening experiment of Brughera *et al.* (2013), using a three-down, one-up adaptive staircase procedure (Levitt, 1971). The simulation followed the two-interval, two-alternative forced-choice (2AFC) paradigm, as in the listening experiment conducted by Brughera *et al.* (2013). The ΔITD started at 100 μs and was decreased with steps of 17 μs until four reversals were reached, after which the step size was reduced to 5 μs. After ΔITD decreased to below 11 μs, the step was further reduced to 2 μs. Overall, 14 reversals were simulated for each analyzed frequency, and the ΔITDs were estimated by computing the average of the last ten reversal points for each frequency.

5.2.4 Discrimination of ILD

The same optimal observer (Equation (5.3)), as in the previous section, was used to obtain the ILD discrimination threshold of the LSO model. The discrimination index is calculated only for a single band, the CF of which is nearest to the central frequency of the pure tone. The threshold detected by the observer will henceforth be denoted as ΔILD. Stimulus A had ILD equal to ΔILD/2 (lateralized to the right ear), while stimulus B had ILD equal to −ΔILD/2 (lateralized to the right ear). The simulation followed the 2AFC paradigm, as in the subjective experiment conducted by Yost and Dye (1988). The interval was considered as successfully predicted by the LSO model if d' was equal to or bigger than 0.95, and the model showed the correct direction of the lateral displacement. The selected d' criterion corresponds approximately to the 75 percent point on the psychometric function (Hacker and Ratcliff, 1979). The ΔILD value started at 1.5 dB in all cases and was decreased by steps of 0.25 dB until ΔILD of 0.4 dB was reached, after which the step size was reduced to 0.05 dB. Overall, 14 reversals were simulated for each analyzed frequency, and the ΔILDs were estimated by computing the average of the last ten reversal points for each frequency.

5.2.5 Discrimination of phase warp

A procedure similar to the original study (Dietz *et al.*, 2008) was implemented to obtain the discrimination threshold between the phase warp and the uncorrelated binaural noise. It was a two-down, one-up adaptive staircase procedure, that converges at the 70.7 percent correct point of the psychometric function (Levitt, 1971). The simulation followed the three-interval, three-alternatives forced-choice paradigm. One of the intervals contained the phase-warp stimulus, while the other two intervals contained the binaurally uncorrelated narrowband noises of the same bandwidths as the phase warp.

In the first part of the experiment, the maximum beat frequency detectable by the MSO and LSO models was analyzed for phase warp with a bandwidth of either 550 Hz or 1100 Hz. The beat frequency started at 50 Hz and was increased/decreased by 15-Hz steps until the second reversal, after which a step size of 10 Hz was used for the next two reversals before the final steps of 5 Hz was used for the remaining reversals. In the case of 1100-Hz bandwidth, the step size was doubled. Thirty reversals were simulated, and the mean was calculated from the last twenty values.

In the second part of the experiment, the smallest detectable modulation depth of phase-warp stimuli with fixed beat frequencies (10, 50, and 75 Hz) was analyzed. The modulation depth was first adjusted with a 4-dB step size until two reversals were observed, after which steps of 2 dB were used for the next two reversals before the final step size of 1 dB was used for the remaining reversals.

The same ideal observer, as in the article by Dietz *et al.* (2008), was used in this experiment. The outputs of the ITD and ILD central stages were transformed into the frequency domain. The magnitude spectra of the outputs, the f_c of which lie within the stimuli' bandwidth, were averaged. The ideal observer compared the average spectra of the three stimuli and chose the stimulus with the most energy within the frequency bin corresponding to beat frequency f_b.

Chapter 6

RESULTS OF SIMULATIONS

In this section, the results of simulations of MSO and LSO models are compared with subjective data. In the figures, the MSO data are depicted as blue-filled diamonds, and the LSO data are depicted as green-filled triangles, connected by a solid line of the corresponding color.

6.1 Lateralization of pure tones with IPD or ILD

The simulated data of the MSO model for pure tones with IPDs are depicted in Figure 6.1. These data are compared with subjective data taken from Yost (1981). We only present mode values calculated by Yost from the responses of four subjects. Therefore, the data shown here does not indicate that some of the subjects reported ambiguous lateralization percepts for IPD larger than ±90 deg. The MSO model data were multiplied by 10 to match the subjective scale. The figure is divided into five panels, each representing one pure-tone frequency (0.2, 0.5, 0.75, 1, and 1.5 kHz). The Pearson correlation coefficients between the simulated data and the subjective data were calculated for each pure-tone frequency, and are shown in the corresponding panels. The coefficients indicate a high correlation between the simulated MSO model and the subjective data. The root mean square error,

$$\mathrm{RMSE} = \sqrt{\left(\sum_{j=1}^{i} (x_j - y_j)^2\right) / i}, \tag{6.1}$$

between the simulated data (x) and the experimental data (y) was calculated for each pure tone frequency, where i represents the total number of tested IPDs. In all cases, the RMSE is quite high due to deviations at extreme IPDs (-180 deg). This will be discussed in the Discussion (Chapter 7). The best agreement between the prediction and the experimental data according to RMSE is at 0.2, 0.5, 0.75 and 1 kHz. In the case of 1 kHz, the model's performance

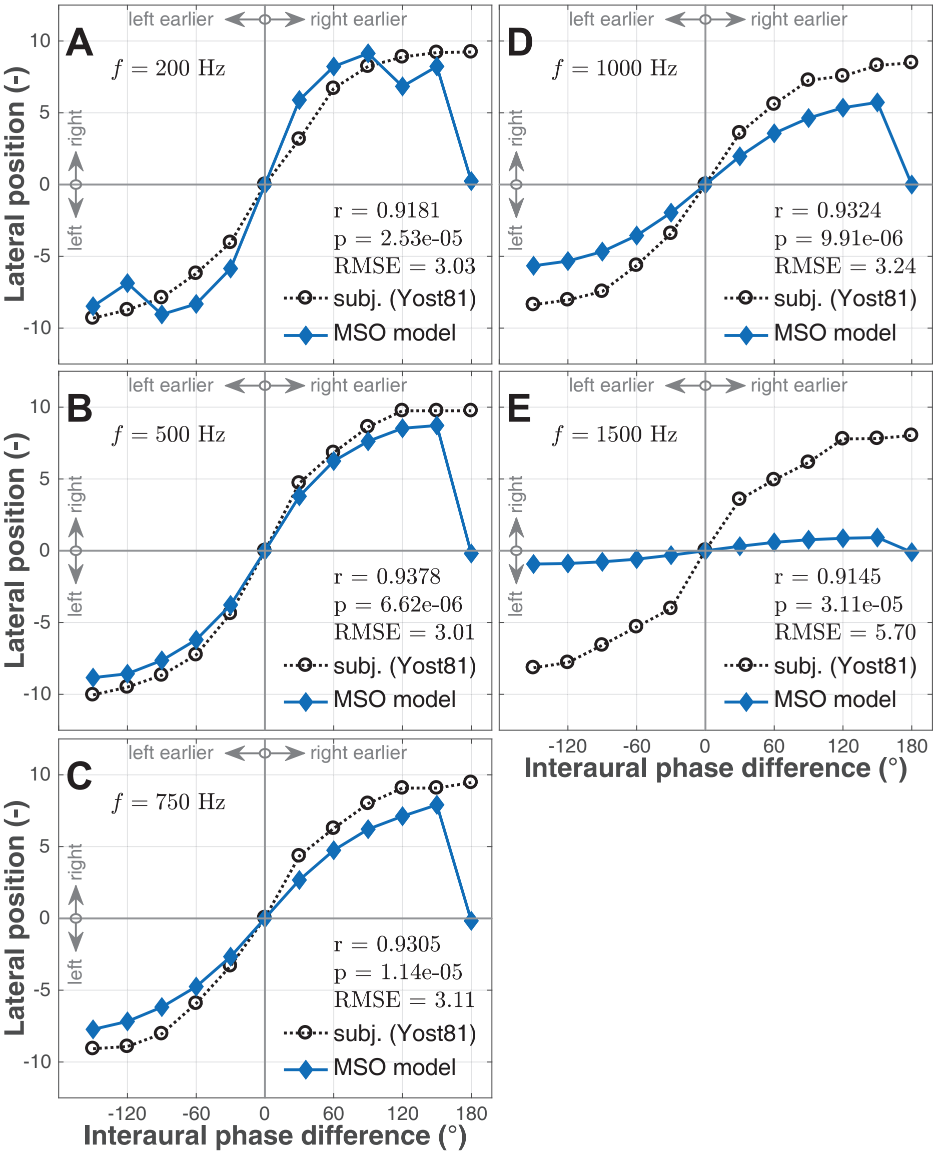

Figure 6.1: Results of the lateralization experiment with pure tones with interaural phase differences (IPD). The subjective data (Yost, 1981) are represented by circles (mode values), and are connected by a black dashed line. The responses of the MSO model are represented by diamonds connected by a solid blue line. Panels A–E show the data for pure-tone frequencies of 0.2, 0.5, 0.75, 1, and 1.5 kHz, respectively.

decays and shows a systematic decrease in sensitivity to IPDs towards higher frequencies. At 1.5 kHz, the model still follows the changes in IPD, but its output is almost damped.

The data for pure tones with ILDs are depicted in Figure 6.2. The LSO model data were compared with subjective data taken from Yost (1981). The LSO model data were multiplied by 10 to match the subjective scale. Data for pure tones of frequencies 0.2, 0.5, 1, 2, and 5 kHz are displayed in five separate panels. As in the previous case, the Pearson correlation

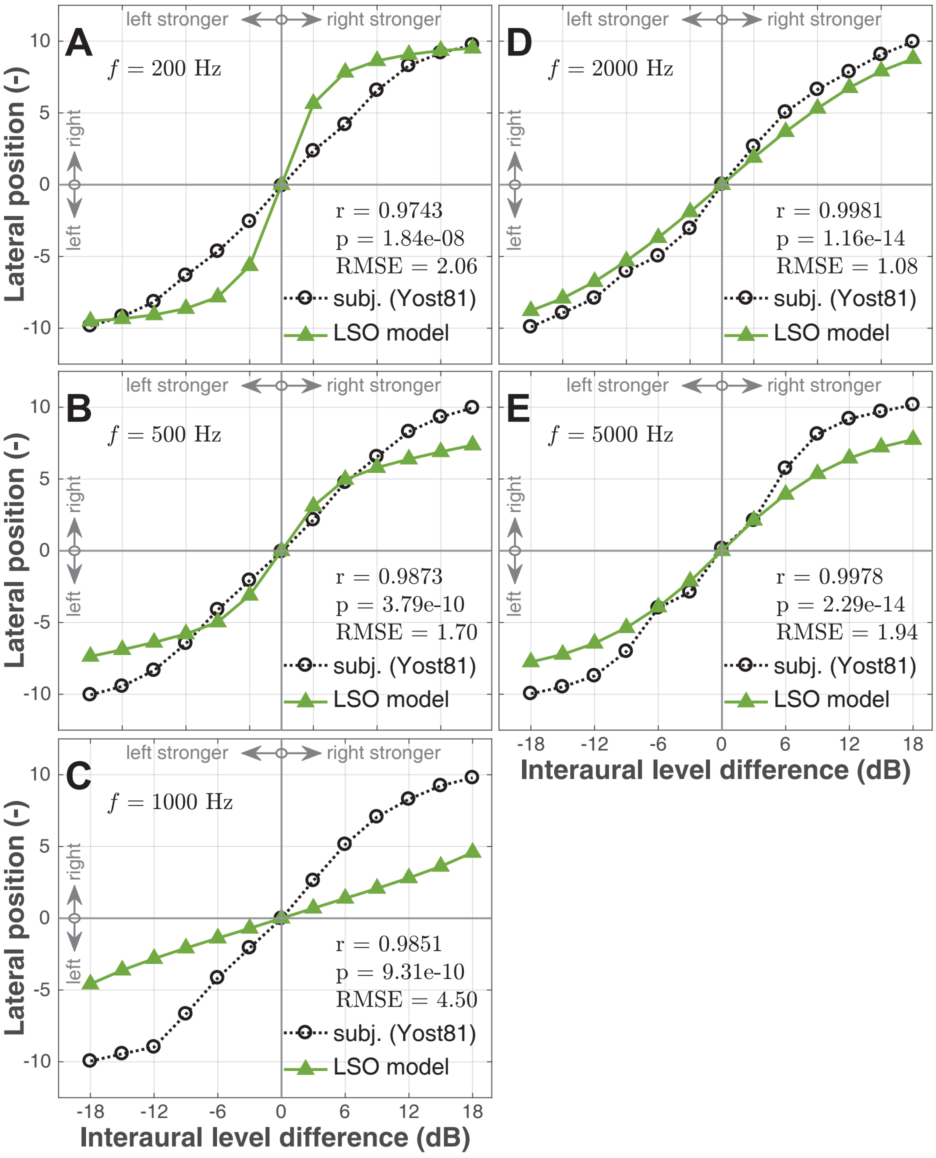

Figure 6.2: Results of the lateralization experiment with pure tones with interaural level differences (ILD). The subjective data (Yost, 1981) are represented by circles (mode values), and are connected by a black dashed line. The responses of the LSO model are represented by triangles connected by a solid green line. Panels A–E show data for pure-tone frequencies of 0.2, 0.5, 1, 2, and 5 kHz, respectively.

coefficients and RMSE between simulated and subjective data were calculated and are shown in each panel. The coefficients indicate a high correlation between the simulated LSO model and the subjective data. The performance of the model is better than human performance at 200 Hz. The best agreement between predictions and experimental data is at 0.5, 2, and 5 kHz. At 1 kHz, the model shows significantly lower sensitivity than the human subjects to ILD.

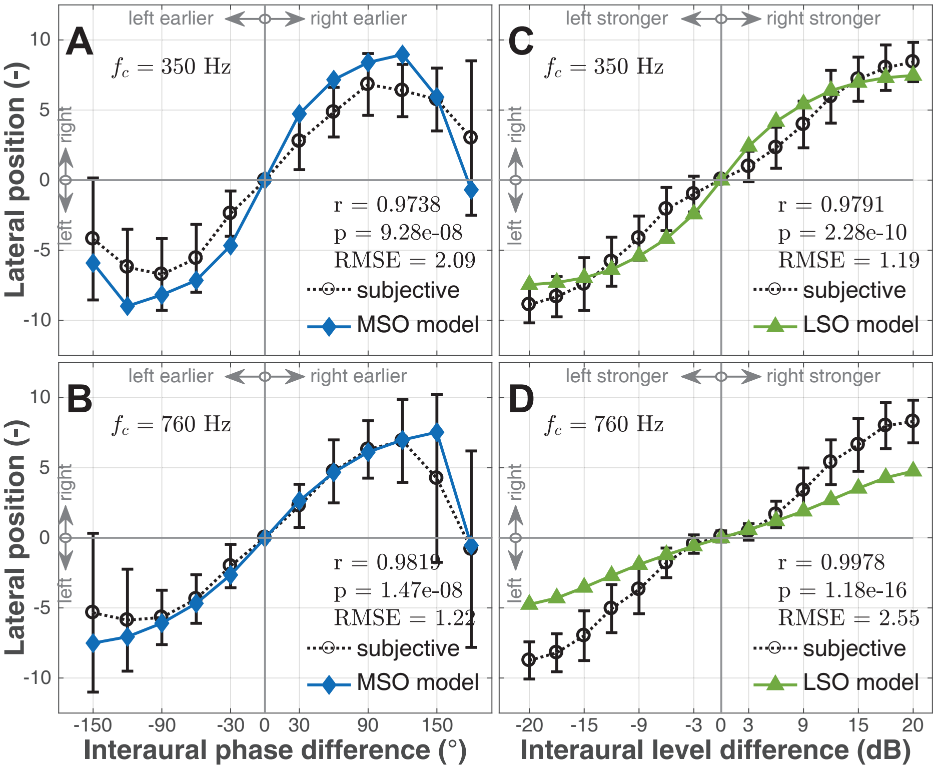

Figure 6.3: Results of the lateralization experiment with narrowband noise. The subjective data are represented by a circle (mean values) with a whisker (standard deviation) connected by a black dashed line. The responses of the MSO or LSO models are represented by a diamond or a triangle connected by blue or green line. Panels A and B show the results of the NBN experiment with IPD, in comparison with the response of the MSO model, while panels C and D show the NBN experiment with ILD in comparison with the response of the LSO model. The top row shows the results for NBNs with 350 Hz f_c and the bottom row shows the results for NBNs with 760 Hz f_c.

6.2 Lateralization of narrowband noise with IPD or ILD

The mean subjective data and their standard deviations from the experiment with narrowband noise with IPD or ILD are shown in Figure 6.3. The subjective data for the NBNs with IPD are compared to the data from the MSO model in panels A and B; and the subjective data for the NBNs with ILD are compared with the data from the LSO model in panels C and D. The MSO and LSO model data were multiplied by 10 to match the subjective scale. While the MSO data are qualitatively comparable with the subjective results, the LSO model deviates in the shape of the response, as is discussed in Chapter 7. The MSO model goes back to zero lateralization for IPD±180 deg, which results from the fact that we change the IPD of each spectral component in the noise. For IPD of ±180 deg., this manipulation creates two signals with an antiphase fine-structure but with the same time-domain envelope. Therefore, the listeners cannot use changes in the time domain envelope. In addition to pure tones with the opposite phase, these stimuli

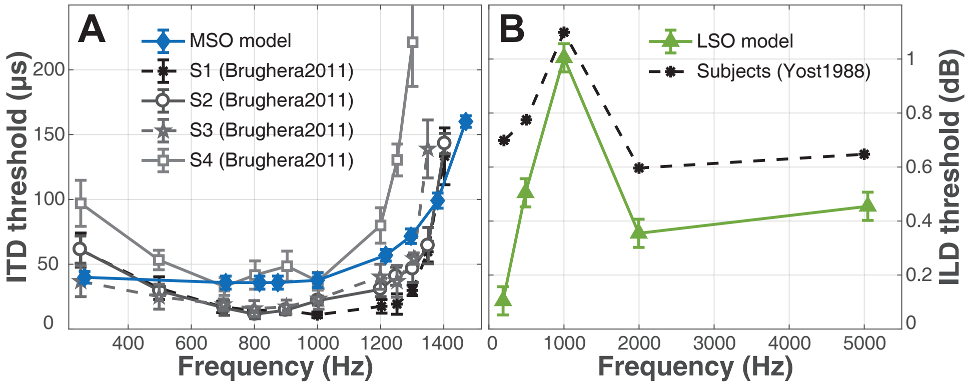

Figure 6.4: Panel A shows the results of the ITD discrimination experiment; the MSO model data are represented by a blue diamonds with a whisker, which represents the standard deviations, and are interconnected by a solid blue line. The mean subjective data from four subjects from Brughera *et al.* (2013) are depicted as circles, hexagons, stars, and squares. **Panel B** shows the results of the ILD discrimination experiment, where the LSO model data are represented by green triangles, and the subjective data reproduced from Yost and Dye (1988) are represented as black dots.

often create ambiguous lateralization on both sides. The RMSE and the Pearson correlation coefficient are calculated for each NBN test case.

6.3 Discrimination of ITD

The discrimination data from the MSO model are shown in Figure 6.4, panel A. The mean simulated data and their standard deviations are compared with the mean subjective data taken from Brughera *et al.* (2013). The simulated data shows good qualitative and quantitative agreement with the subjective data (S1 and S2) for higher frequencies. For lower frequencies, the ITD threshold of the model remain constant while the subjective values increase. The additional point at 1460 Hz shows the highest frequency at which the MSO model is able to discriminate ITD in the pure tone.

6.4 Discrimination of ILD

A comparison between the LSO model simulations and the subjective data is shown in Figure 6.4, panel B. The mean subjective data were taken from Yost and Dye (1988). The model accounts for the loss of sensitivity at 1 kHz, but the discrimination of the pure tones for both lower and higher frequencies exceeded the discrimination of the subjective data. The loss of

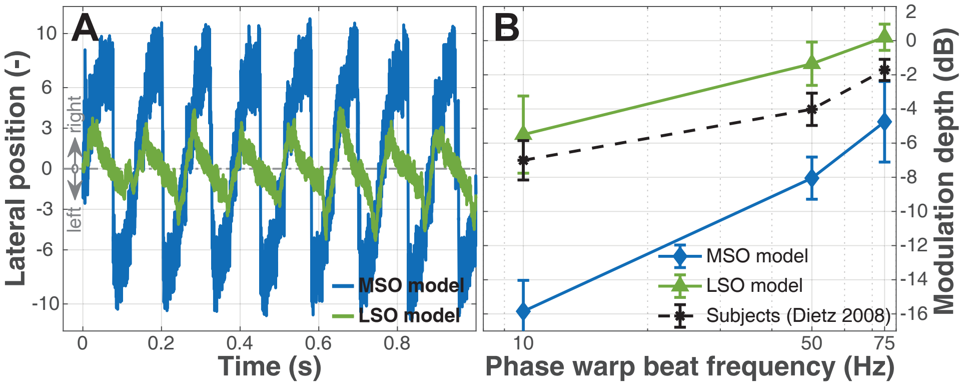

Figure 6.5: Panel A shows transient responses of the MSO and LSO models to a phase-warp stimulus with the beat frequency $f_u = 8$ Hz. **Panel B** shows the experiment results with a phase warp with variable modulation depth. The response of the MSO model is represented by diamonds connected by a solid blue line, and the response of the LSO model is represented by triangles connected by a solid green line. The subjective data reproduced from Dietz *et al.* (2008) are shown by a black dashed line. The standard deviations are depicted as bars above each data point.

sensitivity of the model at 1 kHz is due to a joint effect of the inhibitory delay, the first-order low-pass input filter, and the weighted moving average.

6.5 Discrimination of phase warp

The predicted maximally detectable phase-warp beat frequencies are presented in Table 6.1 along with psychophysical data collected by Dietz *et al.* (2008). The MSO model at the 500 Hz bandwidth exceeds the subjective results by about 50 Hz, but for the 1000 Hz bandwidth, the model shows extreme sensitivity that is almost double the values for the subjective data. However, the LSO shows less sensitivity than the subjective data at 500 Hz bandwidth, but fit the subjective data for 1000 Hz bandwidth. Although the two models do not agree with the data quantitatively, they show qualitatively the same increasing trend of the discrimination threshold with increasing bandwidth as the subjective data.

In addition, the temporal responses of both the MSO model and the LSO model to a phase warp with 8-Hz beat frequency are shown in Figure 6.5, panel A. Both models showed an apparent rotation of the sound image with the same frequency corresponding to the beat frequency of the phase warp. This behavior is in agreement with the subjective data of Siveke *et al.* (2008); Dietz *et al.* (2008). The visible rotation diminished at about 10 Hz and changed to noise-like output, which is again in agreement with the subjective data (Siveke *et al.*, 2008; Dietz *et al.*, 2008).

Table 6.1: The results of the first part of the phase-warp experiment. The table shows mean subjective, MSO, and LSO data with standard deviations for two different phase-warp bandwidths.

Phase warp	Mean sub. data	MSO model	LSO model
bandwidth	Dietz *et al.* (2008)		
550 Hz	96 ± 15 Hz	143.5 ± 15 Hz	61 ± 17.2 Hz
1100 Hz	219 ± 30 Hz	492 ± 22 Hz	211 ± 48 Hz

The maximum detectable phase-warp modulation depth for the MSO and LSO models is depicted in Figure 6.5, Panel B. The subjective data were reproduced from the study by Dietz *et al.* (2008). As in the first part of the experiment, the MSO shows better discrimination than the subjects, and the slope of the curve also does not correspond with the subjective data. The LSO model shows a similar slope as the subjective data but is about 2 dB less sensitive than the subjects at all of the three phase warp frequencies.

Chapter 7

DISCUSSION - BINAURAL MODELS DESIGN AND VERIFICATION

7.1 Lateralization of pure tones with IPD or ILD

The models of MSO and LSO showed a good match with human lateralization data for pure tones. The MSO model lateralization data deviated from the subjective data of Yost (1981) at IPDs between 120 and 180 degrees, which may be due to the ambiguity of the laterality of such stimuli. This behavior was reported by subjects who heard two sound images coming from opposite sides of the head (Yost, 1981). In their statistical analysis, these authors filtered correct output from the data using a mode instead of a mean value. A similar decrease in lateralization was reported by Sayers (1964), who used the mean of his subjective data. The MSO output at these extreme IPDs oscillates between maximum lateralization to the left and right sides, which ultimately decreases the mean output similarly to Sayers (1964) and may indicate the directional ambiguity of this type of stimulus. Another discrepancy is the slope of the lateralization as a function of IPD for a pure tone of 1500 Hz. The change in the slope is mainly due to the rate-code principle used in the MSO model.

7.2 Lateralization of narrow band noises with IPD or ILD

The performance of the model in the NBN stimuli around the center point (0 deg IPD or 0 dB ILD) in all tested setups is superior to the listeners' performance. After this initial discrepancy, the slope change (second derivative) lags back, and the performance of the model of more lateral localizations again matches the performance of the experimental subjects. As noise stimuli are more ecologically relevant than pure tones, this poses an open question: Why is our model performance around the center point better than the performance of the experimental subjects?

7.3 Discrimination of ILD

The ILD discrimination threshold of the LSO model shows a discrete reduction in sensitivity at 1 kHz, which is in line with the psychoacoustical data. This phenomenon is caused by a joint effect of inhibitory delay, the first-order input filter of LSO, and the weighted moving average in the model.

7.4 Discrimination of ITD

The ITD discrimination of the MSO fits the subjective data from Brughera *et al.* (2013) within their standard deviations. In contrast with the subjective data, however, the predicted data show no increase in the threshold at lower frequencies. The abrupt increase in the threshold with increasing frequency is mainly due to the low-pass filter in the MSO model. With the peripheral filter only, there would by a shallower slope rise, and the model would not be able to decode the ITD thresholds at higher frequencies.

7.5 Discrimination of phase warp

Both the MSO and LSO models show a trend with an increasing phase warp discrimination threshold that is qualitatively similar to the subjective data. Quantitatively, the MSO model is more sensitive than the human subjects, and the LSO model is less sensitive than human subjects in the case of 500 Hz bandwidth but shows a good fit for 1000 Hz bandwidth.

In the modulated phase warp discrimination task, the MSO model again shows better performance than is shown by the subjective data. Surprisingly, the LSO model matches the subjective data qualitatively but is about 2 dB less sensitive. This is in line with the neurophysiological data about LSO neurons being sensitive to envelope-ITDs in amplitude modulated signals (Joris and Yin, 1995). These two results indicate a possible theoretical contribution of MSO and LSO to the decoding of phase-warp stimuli.

7.6 Blind-spots of the presented models and possible solution

The information about lateralization is processed independently for the LSO and MSO models. However, it is unlikely that the processing in the inferior colliculus would be different for MSO and LSO. The overall structures of the central stages tend to follow the functional aspect rather

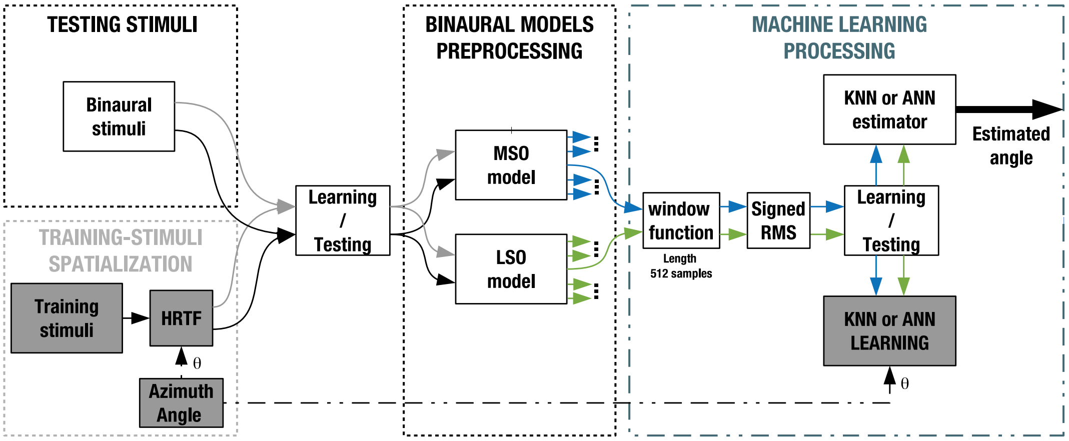

Figure 7.1: Schematic diagram of learning part of machine learning algorithm. It is composed of 3 parts, creation of spatialized learning stimuli, binaural preprocessing, and machine learning processing.

than the physiology. This also limits the possibility to carry out reliable experimental cue-trading tests in which ILD is compensated by ITD and vice versa. The performance of our LSO model is affected if the arrival time between the signals in the left and right ears is changed, as the different time delays between the neural signal in the contralateral and ipsilateral ears in the model affect the discrimination threshold for ILD. The effect of ILD on the outcome of the MSO model is minimal. Moreover, in the presented implementation of both models, one has to know beforehand the stimulus spectral properties and which binaural cue it contains to obtain proper results.

If these parameters were unknown, one possible solution was presented in pilot studies by Koshkina and Bouse (2016, 2017), which has shown that if the former versions of presented models (Bouse and Vencovsky, 2015) are combined with the K-nearest neighbour (KNN) or ANN learning algorithms, they can be used for localization tasks with a head-related transfer function (HRTF), which contains both ITD and ILD information. These pilot studies were further extended for localization in the vertical plane in a master thesis of Koshkina (2017), supervised by the author of this thesis. The algorithm was even-further extended for localization of drum-kits in Melechovský *et al.* (2018).

The overall structure of the algorithm is depicted in Figure 7.1. Mean values of the output signals from the MSO and LSO models in one time-frame (variable length) were processed separately by machine learning algorithms. Signals in each critical band were processed separately, and in the case of the MSO model only the critical bands below 1.5 kHz were used. The KNN and ANN were trained on speech samples filtered by HRTFs for different azimuths between -90 and 90 degrees. The algorithm was then tested by using another set of speech samples created

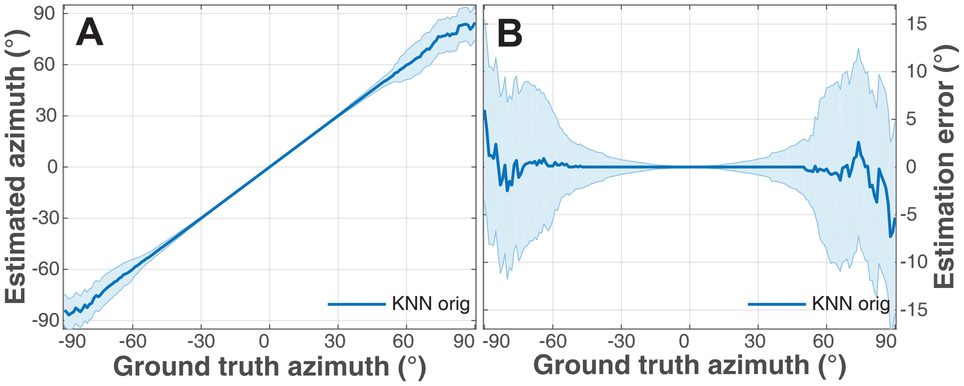

Figure 7.2: Example of localization results using the K-NN algorithm from Koshkina and Bouse (2017) using MSO and LSO models presented in this thesis. The localization estimation are depicted using solid blue line. The shaded blue area depicts the standard deviations of the estimation

by the same HRTFs. The decision device averaged the predicted azimuths from the MSO and LSO parts, and if the deviation between the MSO and LSO predictions was more than 20 degrees, only the prediction from the LSO was taken into account. The emphasis on the LSO predictions aroused from the simulations where the LSO predictions were more stable than the one from the MSO. The algorithm perfectly localized the azimuths near 0 deg (see Figure 7.2). The accuracy decreased with the increasing absolute value of the azimuth.

7.7 Design aims fulfillment

Two-channel binaural models of MSOs and LSOs were designed according to three criteria. The first criterion was that the models should take into account current neurophysiological findings (Brand *et al.*, 2002; Grothe, 1994, 2003; Joris, 1996; Joris and Yin, 1995; Roberts *et al.*, 2013; Tollin and Yin, 2005). Therefore, both models are based on the rate-code principle. In addition, the MSO model is most sensitive if a pure tone in the contralateral ear is delayed by 50- deg IPD (Figure 4.3 Panel A), which agrees with neurophysiological data of Grothe (2003). And with delayed broadband noise, the MSO output shows the broadest peak for low CFs at high ITDs, contrary for high CFs it shows a sharper peak at low ITDs (Figure 4.3 Panel B), which agrees with the neurophysiological data of McAlpine *et al.* (2001).

The second criterion was that the models should give a quantitative representation of the subjective lateralization. If complemented with phenomenological central stages of ITD and ILD, the models' predictions show a good match with subjective data from literature: pure-tone lateralization (Yost, 1981), ITD and ILD discrimination (Brughera *et al.*, 2013; Yost and Dye, 1988), and phase-warp discrimination (Dietz *et al.*, 2008). In addition, the models also predict

NBN lateralization data obtained by the author of this thesis. In the case of ILD discrimination, the LSO model predicts an experimentally observed (Yost and Dye, 1988) decrease in sensitivity around 1 kHz.

The third criterion was that everything mentioned in criteria 1 and 2 should be achieved with low structural and computational complexity. From the computational point of view, it is hard to make an evaluation without having all competing models in the same test scenarios. However, the computational performance of both models can be further increased by reducing the peripheral filters spacing to 1 ERB in a cost of reduced performance in the phase-warp detection task. Moreover, computational and structural complexity is further improved in Part III.

7.8 Comparison to other binaural models

The models of MSO and LSO presented here are the rate-code based models like, for example, the models of Dietz *et al.* (2008); Pulkki and Hirvonen (2009); Takanen *et al.* (2014); van Bergeijk (1962); von Békésy (1930); Encke and Hemmert (2018). The MSO model is most comparable with the models of Pulkki and Hirvonen (2009); Takanen *et al.* (2014). In line with physiology, these MSO models account for the shorter arrival time of the inhibitory signal from the contralateral ear. This is not incorporated in the models of Dietz *et al.* (2008); van Bergeijk (1962); von Békésy (1930). The presented LSO model assumes two inputs with slightly different time delays (shorter for the ipsilateral ear). These inputs are then subtracted. Generally, this approach is similar to the LSO model of Dietz *et al.* (2011); Pulkki and Hirvonen (2009); Takanen *et al.* (2014). The only difference is that the model of Dietz *et al.* (2011) does not incorporate the different delays of signals coming from the ipsilateral and contralateral ears.

Our models of MSO and LSO demonstrate that with a relatively simple neurophysiologically inspired signal processing design, it is possible to obtain performance comparable to human listeners in lateralization tasks. In addition, the ITD and ILD central stages' outputs give values directly representing the subjective lateralization, which is advantageous to the previous rate-code models (Pulkki and Hirvonen, 2009; Takanen *et al.*, 2014). The model further supports the hypothesis that the ITD and ILD information is coded with a rate-code instead of a place code. However, it should be noted that place–code models based on the Jeffress delay line can account for all of the phenomena shown in this thesis, and even for more complex phenomena (Braasch *et al.*, 2013; Breebaart *et al.*, 2001; Colburn, 1977; Faller and Merimaa, 2004; Lindemann, 1986; Prokopiou *et al.*, 2017).

Overall, both models presented here serve as a possible piece in the puzzle surrounding the processing of binaural hearing in the mammalian brain. And, due to their relatively low complexity and good performance, one can use the presented models in real-time applications.

Part III

Binaural Models Design - modifications

Chapter 8

MSO and LSO models with Bernstein's peripheral stage

In the previous part, novel binaural models were introduced in the same state as they were presented presented in Bouse *et al.* (2019). During the peer-review process of that article, several ideas to improve/change the models were proposed by reviewers. These ideas led to the use of the different monaural stage in the models and some follow-up changes in models' structure.

Furthermore, the performance of the former designed MSO model while compared to lateralization of pure tone with IPD subjective data showed decaying performance towards higher pure tone frequencies. This fact led to the idea to include in this comparison, pure tone subjective data from Zhang and Hartmann (2006). In this article, the authors challenged the results from Yost (1981) (data used in the verification of the original MSO design). Zhang and Hartmann (2006) had the main concerns towards the form of stimuli presentation in use in Yost (1981) article, i.e., the use of only one pure tone frequency during one trial. This presentation could lead to the adaptation of listeners' subjective scale and enhancement of less lateral stimuli. Which, in the end, became the case (Zhang and Hartmann, 2006).

In this chapter, we will discuss the proposed changes to the models; we will run the same simulations on them plus an additional pure tone lateralization experiment to accurately compare the performance of former and new design, and compare computational demands of the implementations.

8.1 Peripheral stage changes

In the original version of the model, we have been using Lopez-Poveda and Meddis (2001) DRNL cochlea model accompanied by a simple inner hair cell (IHC) stage, which was composed of

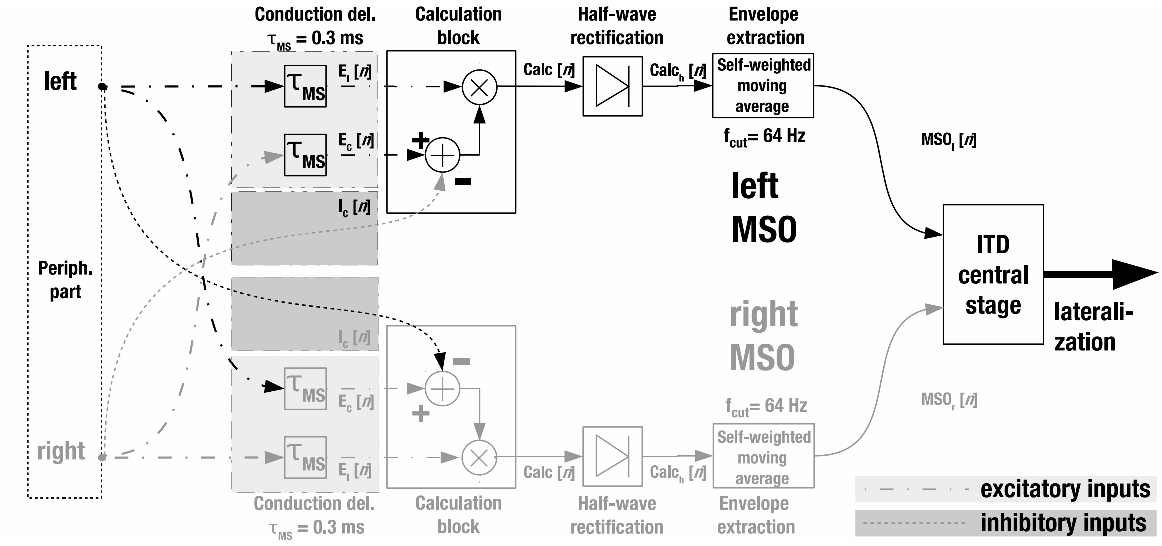

Figure 8.1: Schematic diagram of the modified MSO models for the left (black connection) and right (gray connection) peripheries with the ITD central stage. Each MSO model has three inputs, two excitatory inputs (light gray background) from both peripheries, and one inhibitory input (dark gray background) from the contralateral periphery.

the half-wave rectifier and low-pass filter only. Although the DRNL filter-bank is simulating cochlea non–linear compressibility, we were still forced to use another logarithmical compressive block (Figure 4.4) to further enhance the intensity-to-level transformation in the LSO model. Moreover, even with low–pass filtering (cut-off 770 Hz) in IHC block, additional low–pass filter had to be added into the MSO model to decrease model sensitivity at high frequencies in the ITD discrimination task. Both problems mentioned above were solved by replacing the cochlear selectivity and IHC stages by a gammatone filter–bank and Bernstein's IHC model (Bernstein *et al.*, 1999), respectively. The AMtoolbox (Søndergaard and Majdak, 2013) implementations of these models were used.

The gammatone-filter-bank simulating cochlear selectivity is a simplification of a DRNL, where it accounts only for linear frequency selectivity of the cochlea. It leads to a reduction of computational costs, but on the other hand, leave the signal not compressed. The compression is implemented in the next stage, the IHC model (Bernstein *et al.*, 1999). The compression ratio is set to raise the output signal about 0.2 dB when the input signal is raised by precisely 1 dB (Bernstein *et al.*, 1999). This can be achieved by multiplying the cochlear selectivity output (CS) with its Hilbert envelope raised by the power of -0.77. This processing, computationally–wise, with half-wave, and square-law rectification leads to equation:

$$\mathrm{IHC_c} = \max\left(0, \left|\mathcal{H}\left\{\mathrm{CS}^{-0.77}\right\}\right| \mathrm{CS}\right)^2, \tag{8.1}$$

where $\mathrm{IHC_c}$ is the output of the IHC compression stage, max is a maximum function, and

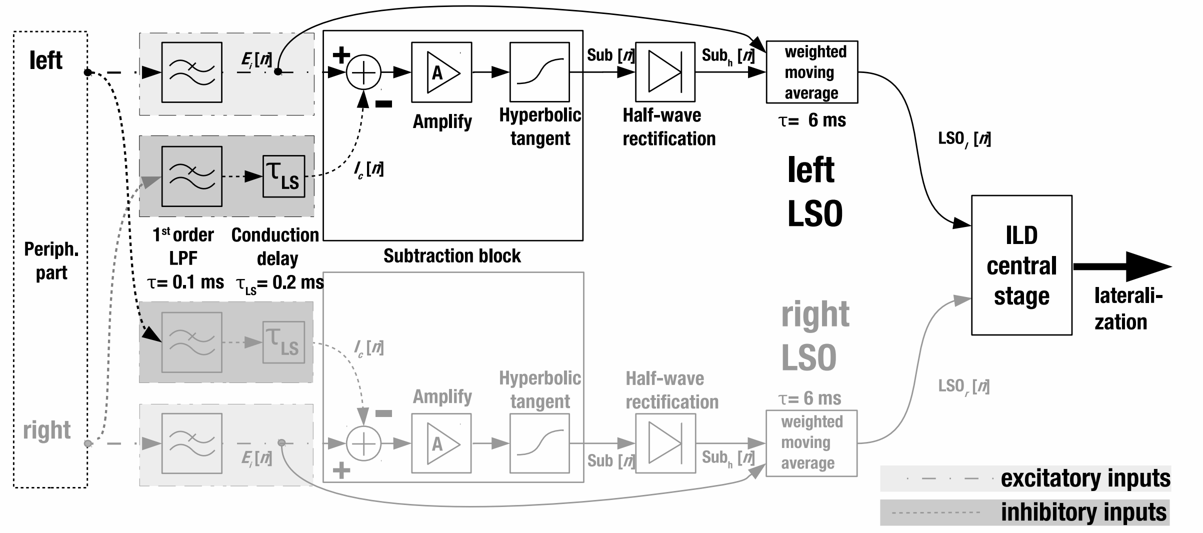

Figure 8.2: Schematic diagram of the modified LSO models for the left (black connection) and right (gray connection) sides of the brain with ILD central stage. The LSO model on one side has two inputs, an excitatory input (light gray background) from the ipsilateral periphery, and an inhibitory input (dark gray background) from the contralateral periphery.

$|\mathcal{H}\{\mathrm{CS}\}|$ is Hilbert envelope of the cochlear selectivity stage CS. The output of the IHC compression stage is then low-pass filtered by second-order Butterworth filter ($f_{\mathrm{cut}} = 425$ Hz) to simulate low–pass properties of IHCs.

8.2 MSO model modifications

The only difference in structure to the original MSO model is the omission of input low–pass filtering because IHC stage filtering is sufficient to follow subjective data in the ITD discrimination task. See the overall structure of the adjusted model in Figure 8.1.

8.3 LSO model modifications

In the LSO model, the compression block became redundant due to the compression in the IHC stage. Therefore the model was simplified by dropping it out (Figure 8.2). Due to the different input level from the peripheral stage, internal noise in ILD central stage was slightly reduced.

Chapter 9

VERIFYING MODEL BY SIMULATIONS - MODELS' MODIFICATIONS

The simulations were run with the same stimuli under the same conditions as in the previous part (Part II). Their detailed description can be found in Chapter 5.

To establish whether the MSO models could reflect that human listeners use ITD over IPD as the primary cue of lateralization (Zhang and Hartmann, 2006). The simulation pool was extended by experiment with lateralization of pure tone with ITD/IPD. The stimuli and statistical analysis of this experiment will be described in the subsection omitting a description of the experiments described in the previous part (Part II).

9.1 Stimuli and procedure - models modification

The stimuli parameters used for verification of the models are described in this section. The parameters were held to be the same as in the psychoacoustic experiments (Yost, 1981; Yost and Dye, 1988; Zhang and Hartmann, 2006; Brughera *et al.*, 2013; Dietz *et al.*, 2008), and Chapter 12.

All stimuli for verification were generated using 96 kHz sampling frequency.

9.1.1 Lateralization of pure tones with IPD or ILD

Stimuli used in lateralization of pure tones with IPD or ILD had the same parameters as in Yost (1981) article. Their full description can be found in Chapter 5.1.1.

9.1.2 Lateralization of pure tone ITD vs. IPD

The pure tones frequencies, IPDs, and ITDs were taken from Zhang and Hartmann (2006). The first part of the whole stimuli set can be seen in Table 9.1, and the other part consisted of the same stimuli but with reversed plus and minus signs. The frequencies, IPDs, and ITDs were initially chosen in the article (Zhang and Hartmann, 2006) to achieve equal distribution of all parameters over the ecological relevant range, where humans can lateralize pure tones fine-structure based of interaural time differences. For frequencies, it meant the distribution between 0.1 to 1.2 kHz, for IPD, distribution between ±150 degrees, and for ITD, distribution which spans between ± 1000 μs.

The simulation ran over whole stimuli set ten times following the subjective test procedure (Zhang and Hartmann, 2006), which gave five-hundred model responses overall. The mean values were calculated over the stationary part of the ITD central stage responses. Only a single band with the CF nearest to the pure tone frequency was taken into account. The filtering of raw responses to stimuli for IPDs ≥ 90, which had an opposite sign to the IPD (response in opposite hemifield to stimulus) in the case of subjective experiment (Zhang and Hartmann, 2006), was unnecessary for the models' predictions. The model's responses appeared in the correct sound-hemifield consistently. Therefore, all five-hundred responses were used for data analysis.

Mean values and their standard deviations as a function of IPD were calculated as follows. First, the mean response of each of 50 stimuli was calculated from all ten runs. The mean stimuli responses were then collected into eleven groups with other stimuli responses of the same IPDs. The mean and standard deviation of one IPD value, was then calculated by taking a mean and standard deviation of the corresponding group. The same processing was then applied to the ITD values. Zhang and Hartmann (2006) proposed a hypothesis; if either ITD or IPD is the most salient cue for the pure tones' lateral position, then the most salient cue's lateral position function should have had the smallest average standard deviation. Therefore, the average standard deviations from ten finite IPD and ITD model responses were calculated. Data from zero ITD and IPD were discarded for being from the same statistical population. A two-sample t-test based on the same ten finite values from IPD and ITD was calculated to validate that the difference in standard deviations averages was statistically significant.

9.1.3 Lateralization of narrowband noise with IPD or ILD

Stimuli used in the simulation with NBN with IPD or ILD experiment are described in Chapter 12.1.3.

9.1.4 Discrimination of ITD

Stimuli used in the ITD discrimination experiment had the same parameters as in Brughera *et al.* (2013) article. Their full description can be found in Chapter 5.1.3.

9.1.5 Discrimination of ILD

Stimuli used in the ILD discrimination experiment had the same parameters as in Yost and Dye (1988) article. Their full description can be found in Chapter 5.1.4.

9.1.6 Discrimination of phase warp

Stimuli used in the phase warp discrimination experiment had the same parameters as in Dietz *et al.* (2008) article. Their full description can be found in Chapter 5.1.5.

Table 9.1: Half of the stimuli set for ITD vs. IPD simulation. The other half consisted of the same stimuli but with reversed plus and minus signs.

Stim. No.	IPD (°)	ITD (μs)	Freq. (Hz)
1	-30	-200	417
2	+30	+400	208
3	-30	-600	139
4	+30	+800	104
5	-60	-200	833
6	+60	+400	417
7	-60	-600	278
8	+60	+800	208
9	-60	-1000	167
10	+90	+200	1250
11	-90	-400	625
12	+90	+600	417
13	-90	-800	313
14	+90	+1000	250
15	-120	-400	833
16	+120	+600	556
17	-120	-800	417
18	+120	+1000	333
19	-150	-400	1042
20	+150	+600	694
21	-150	-800	521
22	+150	+1000	417
23	0	0	167
24	0	0	333
25	0	0	694

Chapter 10

SIMULATION RESULTS - MODELS' MODIFICATIONS

The results of the simulations were obtained for the stationary part of the output signals only. The input transient was omitted. Therefore, the models are not intended to be used for studying the onset dominance of human sound localization or precedence effect.

The former MSO and LSO model (Part II) data are represented by blue diamonds and green triangles, respectively. Their modified MSO and LSO counterparts are represented by purple squares and yellow stars, respectively. The subjective data are in the figures interconnected by the lines in various shades of grey.

10.1 Lateralization of pure tones with IPD or ILD

In Figure 10.1 are depicted lateral position prediction of both original and modified MSO models for lateralization of pure tones with IPD experiment. Subjective data used to calculate lateralization error were re-sampled from Yost (1981) article. The original version of the MSO model performs better for all pure tone frequencies, except the 1.5 kHz one, where the modified model shows less deterioration of results towards higher IPDs.

The lateral position estimation for pure tones with ILDs can be seen in Figure 10.2. The subjective data were re-sampled from the same article as in previous case (Yost, 1981). Both versions of the LSO model perform with almost the same accuracy for pure tones with frequencies of 0.2 and 2 kHz. For the rest, the modified LSO outperforms the original.

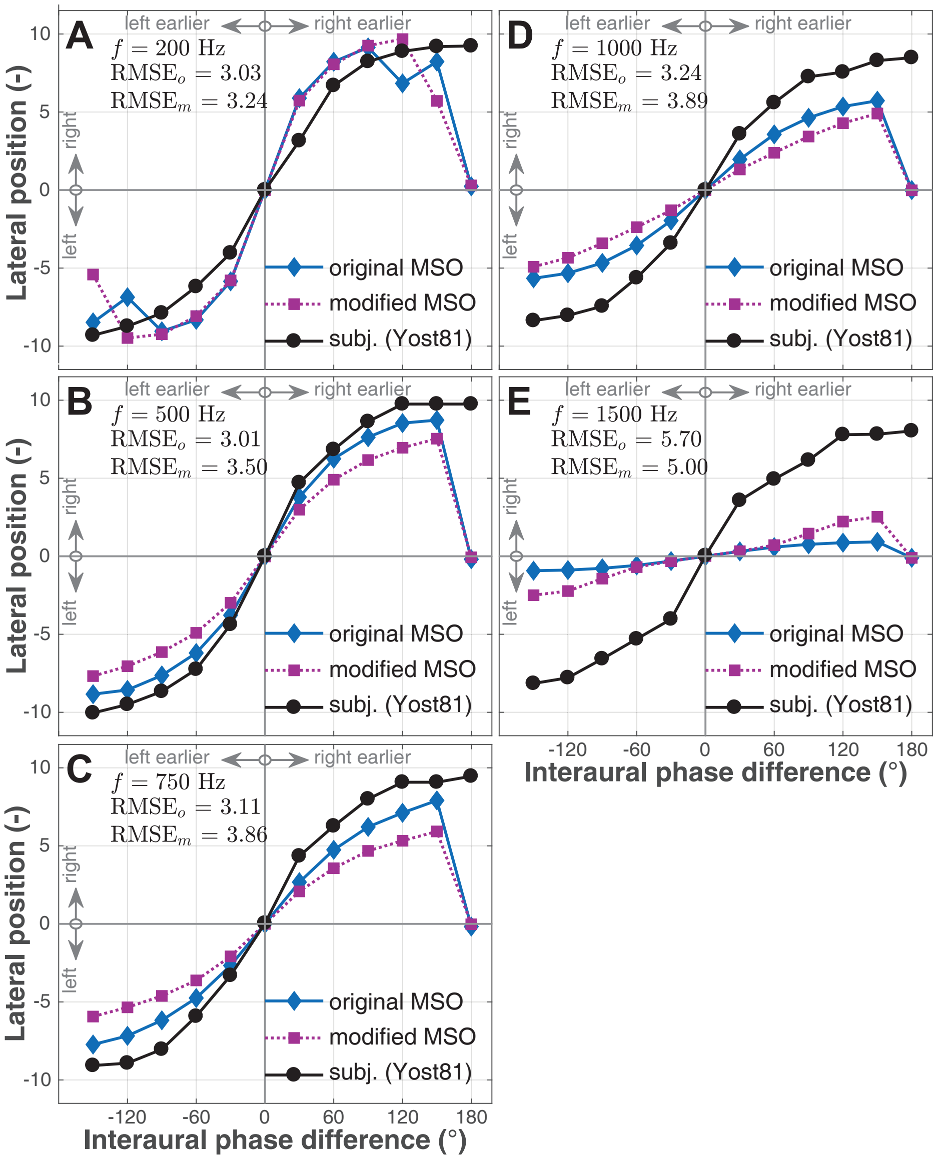

Figure 10.1: Results of the lateralization experiment with pure tones with interaural phase differences (IPD). The subjective data (Yost, 1981) are represented by circles (mode values), and are connected by a solid black line. The responses of the original MSO model are represented by diamonds connected by a solid blue line, and responses of modified MSO model by the squares connected by a dashed purple line. Panels A–E show the data for pure-tone frequencies of 0.2, 0.5, 0.75, 1, and 1.5 kHz, respectively.

10.2 Lateralization of pure tones ITD vs. IPD

The lateral position of pure tones as a function of IPD and ITD is depicted in Figure 10.3, Panel A and Panel B, respectively. The subjective data were re-sampled from Zhang and Hartmann (2006), and only their mean values are depicted for the sake of readability of the plot. The subjective data were linearly re-scaled from a range of ±40 used in Zhang and Hartmann

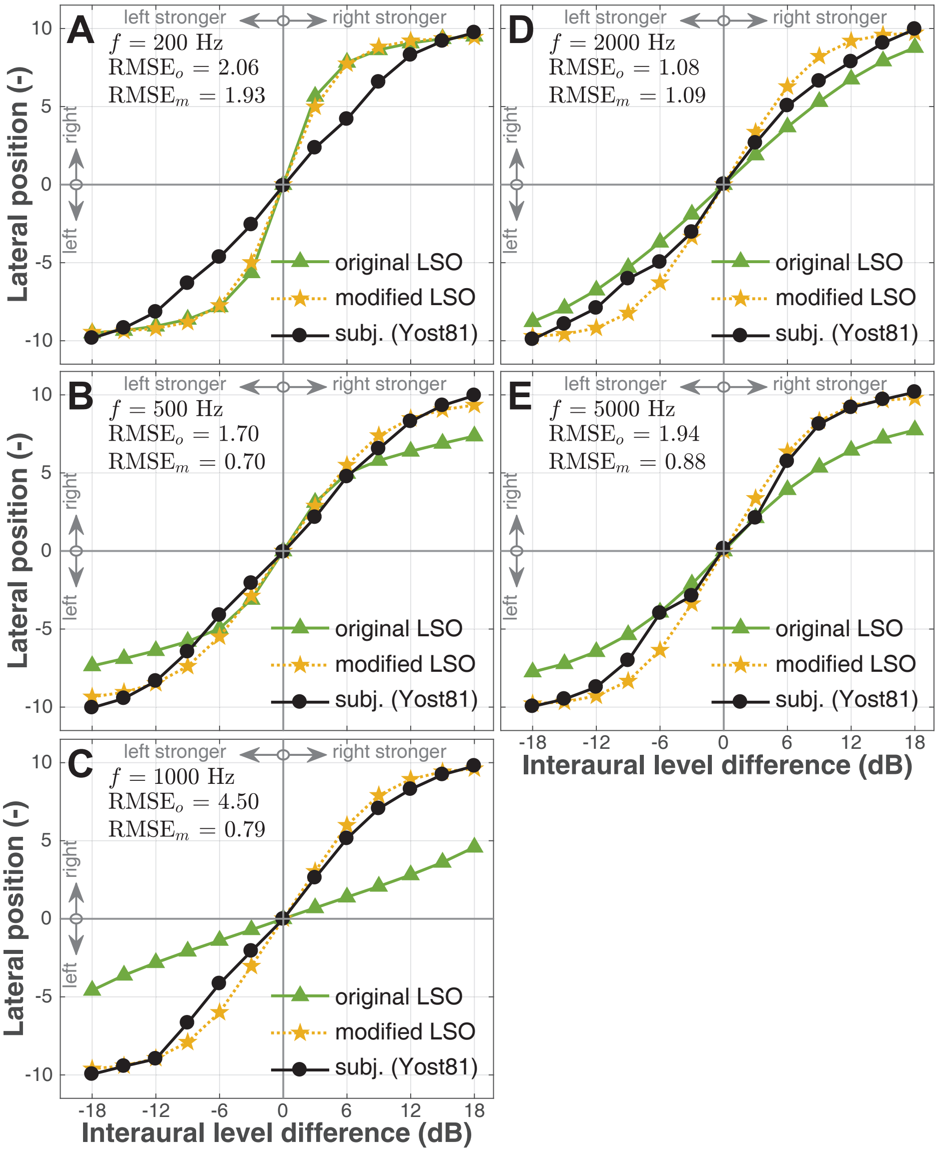

Figure 10.2: Results of the lateralization experiment with pure tones with interaural level differences (ILD). The subjective data (Yost, 1981) are represented by circles (mode values), and are connected by a solid black line. The responses of the original LSO model are represented by triangles connected by a solid green line, and responses of the modified LSO model by the stars connected with dashed orange line. Panels A–E show data for pure-tone frequencies of 0.2, 0.5, 1, 2, and 5 kHz, respectively.

(2006) to ±10 used in our other simulations. Both versions of MSO models fit the subjective data both qualitatively and quantitatively; only subject X deviated from models predictions quantitatively. Subject X exhibited large scale-wise-compression, which is not accountable in the current implementations of the models. In Figure 10.3, RMSE values between the models' estimations and mean subjective data responses over subjects are shown to assess a rough estimation of models' performance.

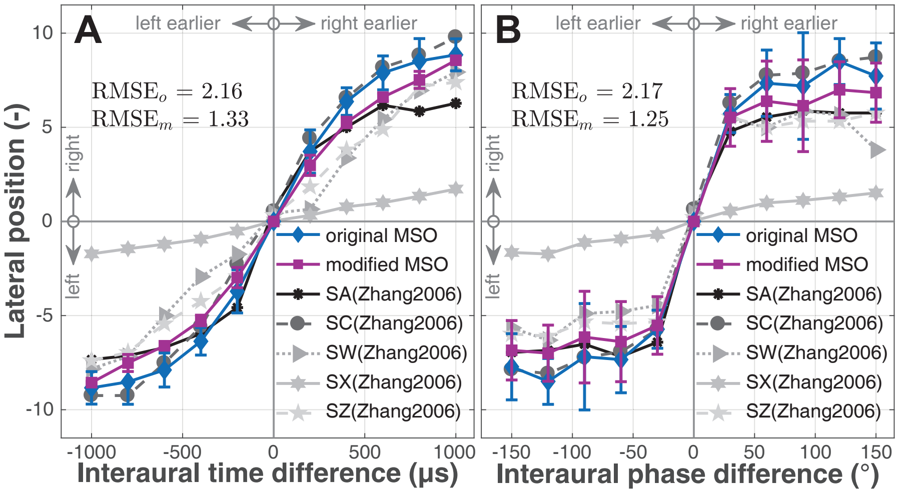

Figure 10.3: Results of the lateralization of pure tones ITD vs. IPD experiment. The means of original and modified MSO models data are represented by a blue diamonds and purple squares with whiskers, respectively. The whiskers represent the standard deviations. The mean subjective data from five subjects (Zhang and Hartmann, 2006) are depicted as stars, circles, right arrows, hexagons, and pentagrams. They are interconnected by lines in shades of grey. **Panel A** represents the data as a function of ITD. **Panel B** represents the data as a function of IPD.

In Table 10.1, the standard deviation of each listener Zhang and Hartmann (2006) and models' responses averaged over non-zero ITDs, or IPDs are shown. The two-sample t-test p-values accompany these values. Both former and modified models show a significant difference in standard deviation means, following the hypothesis by Zhang and Hartmann (2006) that the ITD is the primary cue for lateralization of pure tones fine structure. Moreover, they show the same trend as human listeners (Zhang and Hartmann, 2006).

10.3 Narrow band noises lateralization

For the NBN experiment, a comparison between model revisions and subjective data collected by the author of this thesis (Chapter 12) has been made. The comparison is depicted in Figure 10.4. Both MSO versions were similar performance-wise, with a slight advance of the modified MSO model. In the case of LSO models, the original showed better performance in case of NBN with 350 Hz f_c. With 760 Hz f_c NBNs, the original LSO model responses undershoot the subjective data; in contrast, the modified LSO model responses overshoot the data. In both cases, the RMSE value for the original LSO model was lower.

Table 10.1: Standard deviation of each listener/model responses averaged over non-zero ITD or IPD. P-value represents a probability of being the IPD and ITD means equal based on the result of two-sample t-test.

Listener/Model	Mean STD ITD	Mean STD IPD	p-value
Sub. A (Zhang and Hartmann, 2006)	0.550	0.925	0.01
Sub. C (Zhang and Hartmann, 2006)	0.550	1.825	$\ll$0.01
Sub. W (Zhang and Hartmann, 2006)	0.975	2.125	<0.01
Sub. Z (Zhang and Hartmann, 2006)	0.725	1.725	$\ll$0.01
Sub. X (Zhang and Hartmann, 2006)	0.275	0.350	0.05
Original MSO model	0.878	1.560	0.0186
Modified MSO model	0.277	1.665	2.029e-06

10.4 ITD discrimination

The ITD threshold for both revisions of MSO models is depicted in Figure 10.5 Panel A. Subjective data used for comparison were taken from Brughera *et al.* (2013). For the RMSE calculations the mean of the subjective data was calculated. The original MSO model responses fit the subjective data better than the modified MSO, which has worse ITD thresholds for all frequencies. However, both models approximate the subjective data qualitatively.

10.5 ILD discrimination

In Figure 10.5 Panel B, the results of the ILD threshold experiment, can be seen. The subjective data was reproduced from Yost and Dye (1988). While the RMSE error shows better performance for the modified LSO model, the overall shape of the subjective results are better qualitatively fitted by the original model.

10.6 Phase warp discrimination

The results of the first part of the phase warp experiment for both original and modified MSO and LSO models are shown in Table 10.2. In the case of the MSO model, the original model shows a better fit for lower bandwidth of phase warp, while for the higher bandwidth, it outperforms subjects for over 200 Hz in discrimination. While the modified MSO model, even-tho still over-performing human subjects, shows prediction more aligned to the subjective data. Both

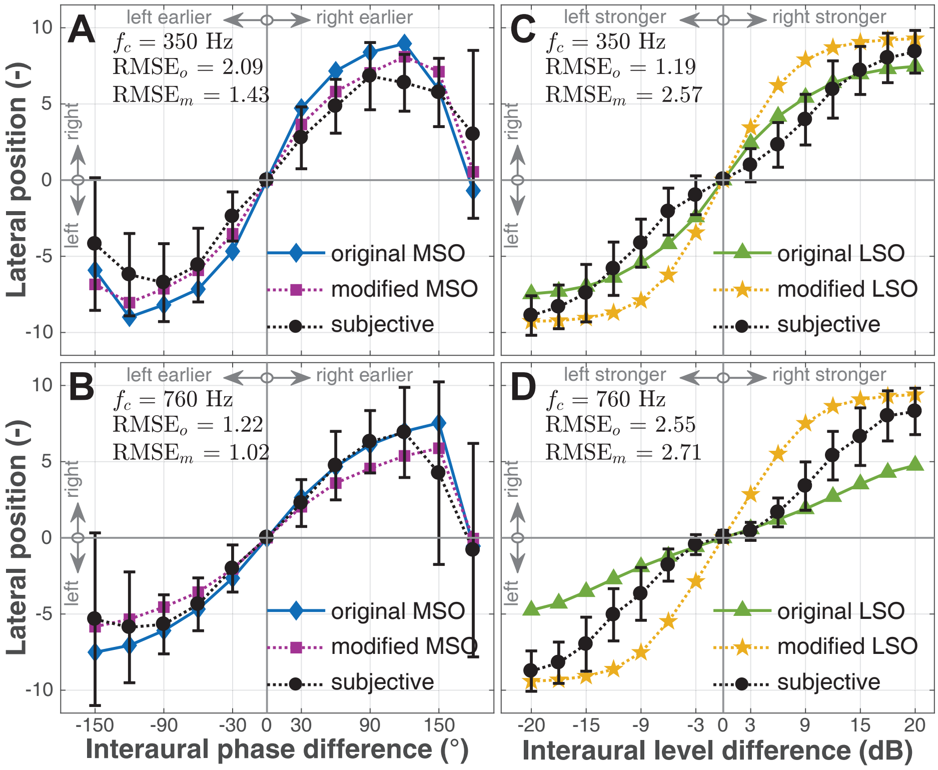

Figure 10.4: Results of the lateralization experiment with narrow band noise. The subjective data are represented by a circle (mean values) with a whisker (standard deviation) connected by a solid black line. The responses of the original and modified MSO models are represented by a diamond connected by a solid blue line and by a square connected by a dashed purple line. The responses of the original and modified LSO models are represented by a triangle connected by a solid green line and by a star connected with dashed orange line. Panels A and B show the results of the NBN experiment with IPD, in comparison with the response of the original and modified MSO models, while panels C and D show the NBN experiment with ILD in comparison with the response of the original and modified LSO models. The top row shows the results for NBNs with 350 Hz f_c and the bottom row shows the results for NBNs with 760 Hz f_c.

revisions of LSO models show a good fit for subjective data with minor differences. The only deviation is in the case of the original LSO model and phase warp bandwidth of 550 Hz, where the prediction is worse than the subjective data.

Figure 10.6 Panel A shows the experiment results with phase warp modulation depth for both MSO models. The subjective data were reproduced from the study by Dietz *et al.* (2008). Eventho a modified MSO model is quantitatively around 8 dB more sensitive than human subjects. It qualitatively fits the slope of the subjective response with rising phase warp frequency almost entirely. The original MSO model shows better discrimination than the subjects, and the slope of the curve also does not correspond with the subjective data.

Figure 10.6 Panel B shows the maximum detectable phase-warp modulation depth for both

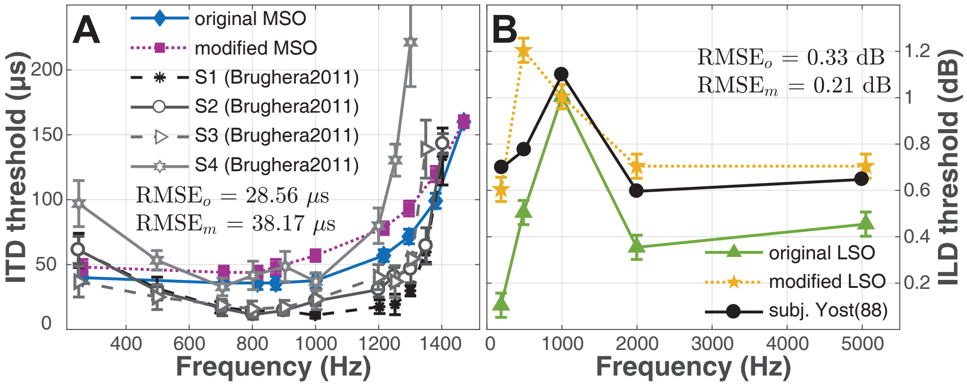

Figure 10.5: Panel A shows the results of the ITD discrimination experiment; the original and modified MSO models data are represented by a blue diamond and purple square with a whiskers, which represents the standard deviations, and are interconnected by a solid blue and dashed purple line. The mean subjective data from four subjects from Brughera *et al.* (2013) are depicted as stars, circles, right arrow, and hexagons. **Panel B** shows the results of the ILD discrimination experiment, where green triangles and orange stars represent the original and modified LSO models data, and the subjective data reproduced from Yost and Dye (1988) are represented as black dots.

LSO models and subjective data from Dietz *et al.* (2008). The modified LSO model predictions show similar properties to the modified MSO, it means that the model predictions are around 4 dB better than subjects, but it again fit the data qualitatively. Those results might indicate that the optimal detector used in the simulations is "too optimal" for this use-case, at least for the modified models. The original LSO model shows a similar slope as the subjective data but is about 2 dB less sensitive than the subjects at all three phase warp frequencies.

Table 10.2: The results of the first part of the phase warp experiment. The table shows mean subjective, original or modified MSO, and LSO data with standard deviations for two different phase warp bandwidths.

Phase warp bandwidth	Mean subjective data (Dietz *et al.*, 2008)	MSO original	MSO modified	LSO original	LSO modified
550 Hz	96±15 Hz	143.5±15 Hz	154±12 Hz	61±17 Hz	105±15Hz
1100 Hz	219±30 Hz	492±22 Hz	305±33 Hz	211±48 Hz	237±26 Hz

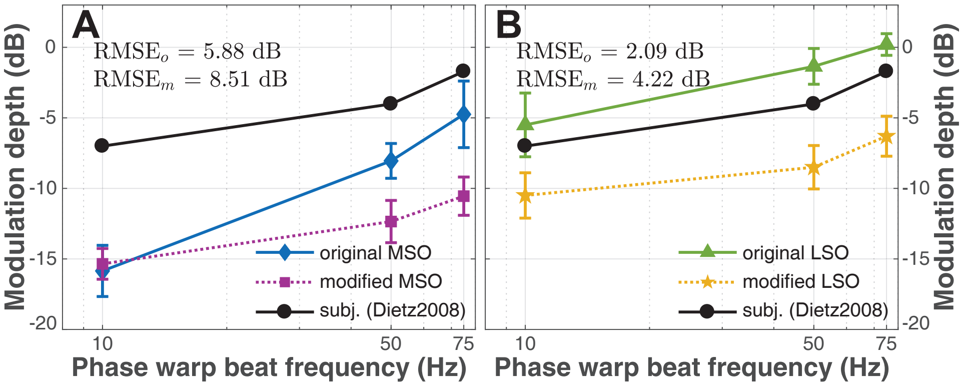

Figure 10.6: Panel A shows the results of the experiment with a phase warp with variable modulation depth. The response of the original MSO model is represented by diamonds connected by a solid blue line, and the response of the modified MSO model is represented by squares connected by a purple dotted line. **Panel B** shows the results of the experiment with a phase warp with variable modulation depth. The response of the original LSO model is represented by triangles connected by a solid green line, and the response of the modified LSO model is represented by stars connected by a orange dotted line. In both panels, subjective data reproduced from Dietz *et al.* (2008) are shown by a black dashed line. The standard deviations are depicted as bars above each data-point.

10.7 Computational demands

Another aspect of the models' performance that needs to be taken into account is computational demands. While one model might be perfectly accurate in predicting subjective data, it still does not mean that we would not need to sacrifice accuracy to a speedup of the calculations for some applications. Therefore, a simple test to estimate overall computational time was run on three different PCs, each with different processors and counts of logical cores [1]. The one-second pre-calculated binaural broadband stimulus with both ILD and ITD was fed into MSO and LSO model for a thousand times.

The computational times were normalized by the time of the original MSO or LSO models. The results in Table 10.7 indicates that with modified models, we save between 30-to-40 percent of the time to process the binaural data.

Table 10.3: The normalized computational demands of original or modified MSO and LSO models.

		original MSO	modified MSO	original LSO	modified LSO
Relative	PC1	1	0.59	1	0.59
computational	PC2	1	0.57	1	0.56
time	PC3	1	0.7	1	0.7

[1] PC1: processor Intel Core i5-3570K@4.4 GHz (4 logical cores), PC2: processor AMD Ryzen 7 1700X@3.4 GHz (16 logical cores), PC3: Intel Core i7-920@2.8 GHz (8 logical cores)

Chapter 11

DISCUSSION - MODELS' MODIFICATIONS

The proposed changes in MSO and LSO models were presented in this part. The modified versions of the models were then compared to the original ones in the same simulations, as in the previous chapter. Furthermore, simulations of lateralization of pure tone ITD vs. IPD experiment of Zhang and Hartmann (2006) was added. This simulation results explain the decrease in MSO models' performance in Yost (1981) pure tone IPD experiment on higher pure tone frequencies. Both original and modified MSO models fit the Zhang and Hartmann (2006) data and support the hypothesis that the data from Yost (1981) were biased due to the adaptation-to-scale of the listeners, which is something the MSO models cannot reflect at their outputs.

The open-ended question which version of models to use still stays. Both versions have similar qualitative results in predicting subjective data; in some experiments, the original version is better, in some the modified. It is worth to note here that there are two exceptions, the first modified version of MSO doesn't hold for neurophysiological data with either pure tones with IPDs or with noises with ITDs. On the other hand, both modified MSO and LSO outperform the originals in the phase warp experiment by qualitatively fitting the subjective data.

The quantitative differences in models' revisions accuracies by their mean RMSE values are shown in Table 11.1. However, they are inconclusive as in the qualitative case, where the performance changes experiment to experiment.

The only case where we have a clear performance difference is in computational demands. It is not surprising that after simplifying peripheral and binaural parts, the modified models need less computational power. Nonetheless, when the computational power is not the main criteria for the choice of the model, one has to experiment which version of the models perform better in his application.

Table 11.1: The overall RMSE per experiment comparison between original and modified models. Lower RMSE value of either original or modified model is in bold font. *Note: RMSE values across experiments are not comparable due to different scales used (with the exception of lateralization experiments).

Experiment	MSO original RMSE	MSO modified RMSE	LSO original RMSE	LSO modified RMSE
Pure tones with IPD	**18.10**	19.50	-	-
Pure tones with ILD	-	-	11.27	**5.39**
Pure tones ITD vs. IPD f(IPD)	2.16	**1.33**	-	-
Pure tones ITD vs. IPD f(ITD)	2.17	**1.25**	-	-
NBNs with IPD	3.31	**2.45**	-	-
NBNs with ILD	-	-	**3.74**	5.28
ITD discrimination	**28.65 μs**	38.20 μs	-	-
ILD discrimination	-	-	0.33 dB	**0.21 dB**
Phase–warp discrimination	195.94 Hz	**73.35 Hz**	25.39 Hz	**14.23 Hz**
Phase–warp modulation discr.	**5.88 dB**	8.50 dB	**2.09 dB**	4.22 dB

Part IV

Psychoacoustic Experiments

Foreword for the psychoacoustic experiments

The functional binaural models and psychoacoustic experiments are tightly connected. Therefore, in the process of designing binaural models, we have designed and organized several psychoacoustic experiments with binaural stimuli: lateralization of dichotic pitches (presented as a part of Bouse and Vencovsky (2015)), lateralization of narrow band noises (part of Bouse *et al.* (2019)), lateralization of pure tones ITD vs. IPD (Bouse and Schimmel, 2017), and audible quality assessment of DHRTF artifact reduction methods (part of Storek *et al.* (2016)).

However, for this thesis purpose, only experiments published in IF journals will be further described, i.e., lateralization of narrow band noises, and audible quality assessment of DHRTF artifact reduction methods.

Chapter 12

LATERALIZATION OF NARROW BAND NOISES WITH IPD OR ILD

This chapter focuses on the experimental methods used to evaluate the subjective lateralization of narrow band noises with IPD or ILD, and the result of the experiment. This experiment was already published as a part of Bouse *et al.* (2019).

The primary motivation behind this experiment was to obtain subjective data, which can then be used to compare the stability of subjective response and responses of the designed models' in Part II. We choose the narrow band noises with a bandwidth equal to one ERB (Equation (1.1)) with a hypothesis that they should be audibly lateralized similarly to the pure tones of similar frequency with the same interaural parameters and therefore directly comparable to the data from the literature (Yost, 1981).

12.1 Experimental methods - NBN experiment

The experimental method, the stimulus details, and the subjects participating on the NBN experiment are described below.

12.1.1 Apparatus and procedure - NBN experiment

The NBN listening experiment took place in a sound-insulated booth. The subjects were seated in front of a computer monitor and listened to the stimuli through headphones (in the IPD experiment, Sennheiser HD 595; in the ILD experiment, Sennheiser HD 650) connected to the sound card output (RME Fireface UC). The headphones were calibrated to maintain the same SPL in both channels of the test material. The experimental procedure was similar to that used by Yost (1981). The listeners indicated a perceived lateral position of the sound on a graphic

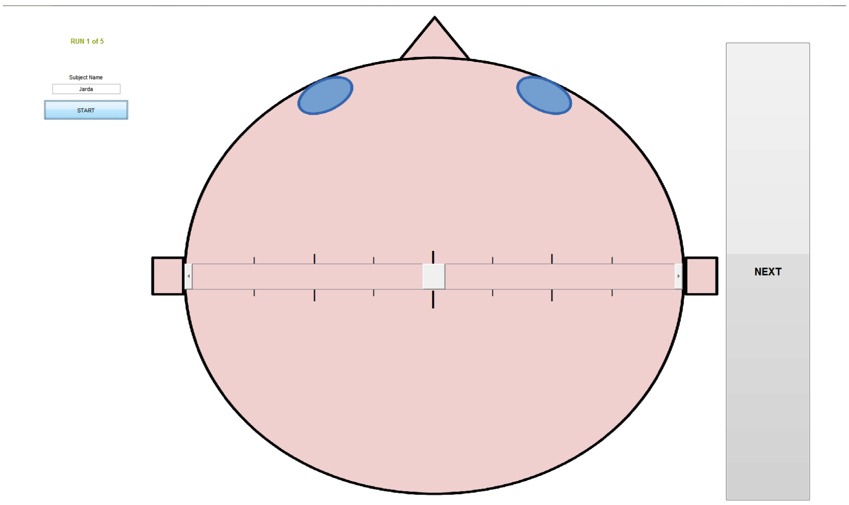

Figure 12.1: Screenshot of Matlab GUI for the NBN experimental procedure.

scale (Figure 12.1). The scale was represented as a movable pointer on a drawing of a head. The position of the pointer represented the perceived lateralization of the sound inside the head. Each slider position was linearly mapped to a numerical value ranging from -10 to 10, where -10 corresponded to the maximum lateral displacement to the left, and 10 corresponding to the maximum lateral displacement to the right. The experiment was organized in four listening sessions, two for each central frequency (350 and 760 Hz). Listeners were required to take a minimum break of five minutes after each session. All stimuli were presented five times to a listener during a given session, in random order. From each of the listening sessions, we obtained 10 complete sets of subjective data per central frequency. In the first listening session for each central frequency, the first two sets were discarded to provide enough time for the listeners to adjust to the procedure. In the second session for each central frequency, the first set of data was discarded. Overall, 7 sets of subjective data for each central frequency were used in the evaluation. The stimuli were presented to the listener in the form of a repetitive stimuli train (see Figure 12.2).

12.1.2 Subjects - NBN experiment

Seven subjects (including one female) participated in the NBN with IPD experiment, and eight subjects (including two females) participated in the NBN with IPD experiment. The participants were aged between 20 and 46 (all except for one were aged below 40). One subject participated in both experiments. Subjects had no or little prior experience with this type of listening test. Their pure-tone hearing thresholds were within a range of 15 dB hearing level for frequencies between 0.25 and 8 kHz (normal hearing). The subjects voluntarily took part in

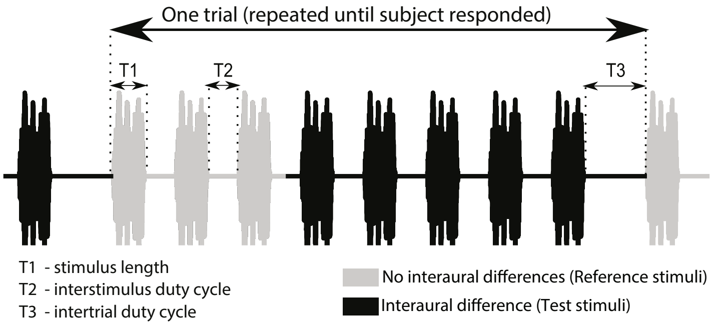
One trial (repeated until subject responded)
T1
T2
T3
T1 - stimulus length
T2 - interstimulus duty cycle
T3 - intertrial duty cycle
No interaural differences (Reference stimuli)
Interaural difference (Test stimuli)

Figure 12.2: Time diagram of one stimuli train. The train was composed of three reference narrowband noises without interaural differences, followed by five stimuli with the same testing interaural difference. All eight stimuli were pulsed with a 50-percent duty cycle. The stimuli train was repeated after 700 ms of silence until the listener responded.

the experiments, and all procedures were following the current ASA Ethical Principles (ASA, 2019).

12.1.3 Stimuli - NBN experiment

Narrowband noise (NBN) with a bandwidth equal to 1 ERB (see Equation (1.1)) was used as a stimulus in the experiment. Two distinct central frequencies f_c (350 and 760 Hz) of NBN were used. NBNs were generated in each listening session in the frequency domain with random amplitude and phase in the passband frequencies. The stimuli were 100-ms long with an 8-ms-long on- and off-raised-cosine ramps. The level was set to 50 dB relative to the subject's hearing level at the central frequency f_c.

In the IPD experiment, the IPDs were varied from -150 to 180 degrees with steps of 30-degrees. The IPD between the left and right ear channels was created in the frequency domain: the desired IPD was imposed onto the phase spectrum of the to-be-delayed signal, after which the signal was transformed back to the time domain.

In the ILD experiment, the ILDs were varied within the range from the interval from -18 to 18 dB with steps of 3 dB with -20 and 20 dB ILD in addition. The ILD was imposed on the stimuli by amplifying one channel by ILD/2 and attenuating the other channel by the same amount.

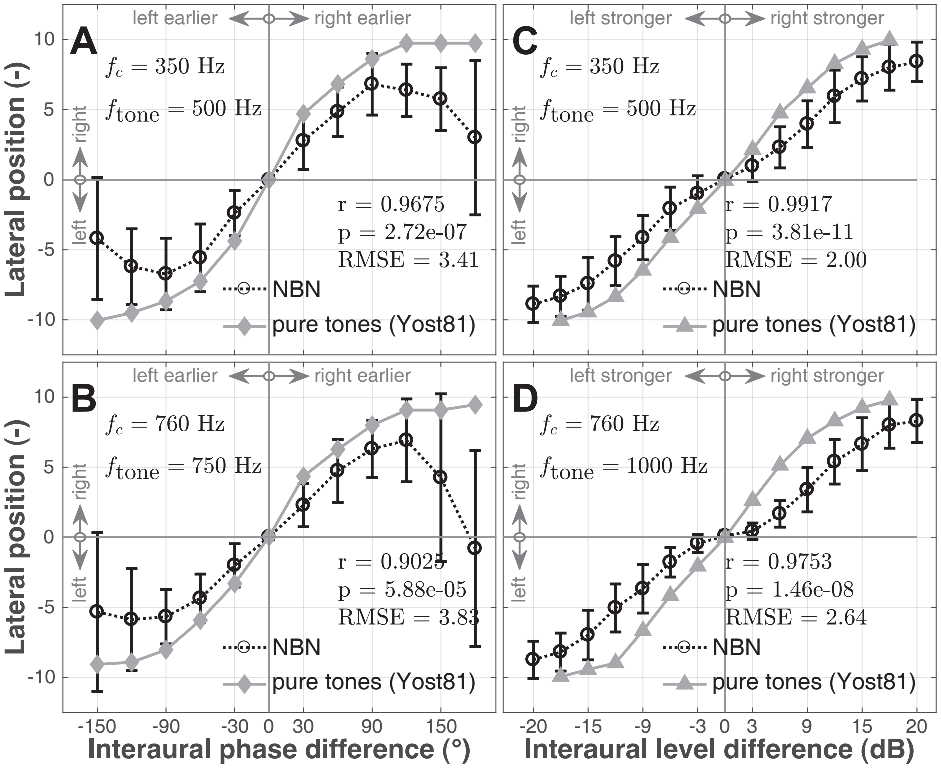

Figure 12.3: Results of the lateralization experiment with narrow-band noise. The NBN subjective data are represented by a circle (mean values) with a whisker (standard deviation) connected by a black dashed line. The subjective responses to pure tones with IPD or ILD from Yost (1981) are represented by a diamond or a triangle connected by a solid gray line. Panels A and B show the results of the NBN experiment with IPD, in comparison with the subjective response to the pure tone data with IPD (Yost, 1981), while panels C and D show the NBN experiment with ILD in comparison with the subjective response to the pure tone data with ILD (Yost, 1981). The top row shows the results for NBNs with 350 Hz f_c and the bottom row shows the results for NBNs with 760 Hz f_c.

12.2 Results - NBN experiment

The mean subjective data and their standard deviations from the experiment with narrow band noise with IPD or ILD are shown in Figure 12.3. The subjective data for the NBNs with IPD are compared to the subjective data for pure tones with IPD (Yost, 1981) in panels A and B (with the closest frequency to the central frequency of the NBN); and the subjective data for the NBNs with ILD are compared with the subjective data for pure tones with ILD (Yost, 1981) in panels C and D(with the closest frequency to the central frequency of the NBN).

The NBN with IPD data are qualitatively comparable with the pure tones' subjective results of Yost (1981). The quantitative differences are mainly because of the different statistics used to calculate the data points (modes vs. mean values). However, in case of IPD equal to 180 deg, subjects in the NBN experiment responded preferably to the center of the head (even with modes statistic), while in case of pure tones the subjects preferred the maximal lateral position (Yost, 1981). This can result from the fact that we have changed the IPD of each spectral component in the noise, which could be unnatural to the subjects. Therefore, the listeners could not use changes in the time domain envelope.

While the NBN IPD data fit the pure tones equivalents qualitatively, the NBN ILD data deviates in the shape of the response to the pure tones (Yost, 1981), as is discussed below (Chapter 12.3).

12.3 Discussion - NBN experiment

The subjects' performance in the NBN with IPD stimuli around the center point (0 deg IPD) lagged to the performance of the subjects with pure tones. After this initial discrepancy, the slope change (second derivate) and match the pure tone responses, however, at the extreme points it begins to decay. This discrepancy around extreme points is most probably caused by the artificial nature of the NBN noise with IPD to the listeners; in the natural condition, those will be located only by using ITD. This could further amplify the ambiguity of location, which was reported by Yost (1981) for pure tones at this extreme IPDs.

Strange behavior is evident in the ILD experiment for low ILDs (range -3 to 3 dB), where the subjective lateralization remains around zero. These results disagree with Yost's pure tone data (Yost, 1981). As noise stimuli are more ecologically relevant than pure tones, this poses an open question: Why is the pure tones performance around the center point better than the performance of our subjects in NBN experiments? We investigated whether this stickiness to zero was caused by the way the ILD was induced to the stimuli. In an experiment, the ILD was added to one channel, while the other channel was left with the same amplitude. Five subjects participated. The results showed slightly higher stickiness to zero for both CFs of NBN. With two of the subjects, we also conducted several reduced experiments (fewer runs per test) in order to exclude any systematic error of the equipment or test methodology. Here is a full list of changes (each change was evaluated separately):

- the sound card was changed to RME Fireface 800,
- the headphones were changed to Sennheiser HD 280 Pro,
- the stimuli train was changed by removing the middle head reference increasing the stimuli length and reducing the range of the ILD to 18 dB.

In all cases, the overall shape of the lateralization curve remained the same. We hypothesized that the reason for the discrepancy might be in the envelope fluctuation of the NBN noise, which could deteriorate the hearing system's ability to map small changes in the ILD to the lateral displacement. To test this hypothesis, we organized a quick informal test with a 760 Hz pure tone. Six subjects participated, three of which had particular knowledge from earlier experiments, while the other three were naive. The ILD range was reduced to ±6 dB. The subjective responses scaled linearly with increasing ILDs, which follows the results of Yost (1981) and supports our hypothesis.

Chapter 13

SUBJECTIVE EVALUATION OF DHRTF ARTIFACTS

This chapter focuses on the experimental methods used to subjectively evaluate artifact reduction algorithms used for DHRTF virtual sound positioning. This topic was thoroughly described in Storek *et al.* (2016). First author, Dominik Storek, developed the methods to reduce artifacts occurring in DHRTF positioning, like audible hissing, overall noise, while the task of the author of this thesis was to design, perform, and evaluate subjective experiments. Therefore only part covering the experiments will be described in this chapter.

13.1 Experimental methods - DHRTF experiment

The experimental method, the stimulus details, and the simulation parameters are described below. The experiment's main goal was to address whether proposed methods were able to reduce audible artifacts in the positioned sound and assess their quality on the subjective scale. Two experiments were performed (Experiment A and B). In the Experiment A subjects listened to the sound positioned dynamically across the whole front horizontal plane $\vartheta \in (0, 180)$. While in Experiment B, the scope was reduced to the angles with the highest probability of the artifact occurrence, i.e., angles ranging in $\vartheta \in (0, 30)$.

13.1.1 Apparatus and procedure - DHRTF experiment

The experiments took place in a sound-insulated booth. The subjects were seated in front of a computer monitor and listened to the stimuli through headphones (Sennheiser HD650) connected to the sound card output (RME Fireface UC). The headphones were calibrated to maintain the same sound pressure level in both the left and right channels at 1 kHz. Artificial ear (Bruel & Kjær 4153) connected to the microphone conditioner (Bruel & Kjær Nexus) were used. Each headphone-can was measured 25 times, re-positioned after each measurement, using

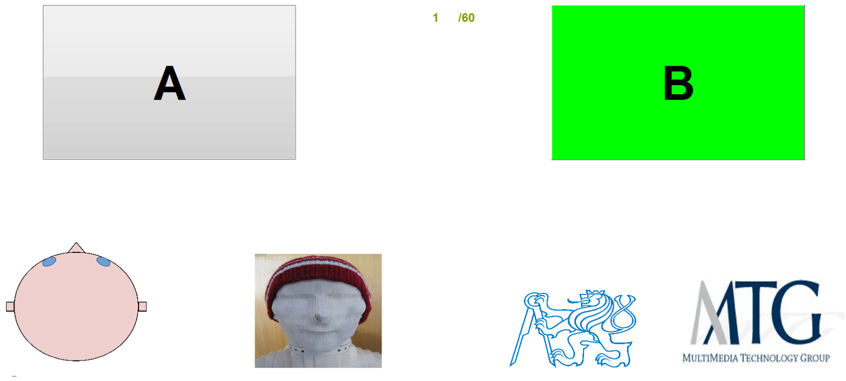

Figure 13.1: Screenshot of Matlab GUI for the DHRTF experimental procedure.

logarithmic sweep harmonic function as a measuring signal (Farina, 2007). Force equal to 5 N was added to each can to hold their position on the artificial ear to simulate the force of headband on the listener's head. All measurements were Fourier transformed in the frequency domain, and mean amplitude spectra value of all measurement per can were calculated. Detailed information regarding the calibration procedure can be found in Bouse (2015).

Listeners responded in a Matlab GUI (Figure 13.1) displayed on the computer screen in the sound-insulated booth. The GUI consisted of two large buttons, labeled A and B. During the playback of the stimuli pair, the color of the buttons turned green to emphasize the currently reproduced stimulus. After the stimuli pair presentation, the listeners were instructed to choose subjectively preferred stimulus by clicking on the corresponding button. The listeners were able to replay the stimuli pair multiple times by clicking on the right mouse button.

The experiments were organized in two listening sessions, one for each experiment. The session was divided into a learning and testing phase. In the learning phase, the subjects listened to three stimuli (mono reference, HRTF-positioned, and DHRTF-positioned with artifacts). Subjects had the opportunity to replay each of the stimuli to get accustomed to the form of artifacts presented in the recordings.

In the testing phase, five stimuli (guitar chord, snare drum phrase, white noise, singing, and speech segment) were positioned by four methods (untreated DHRTF, low-pass filtered DHRTF, spectral limited DHRTF, and DHRTF smoothed by moving average). To acknowledge variation in HRTF/DHRTF person to person, we have used three different HRTF set were chosen for positioning of the signal: two from the CIPIC HRTF Database (KEMAR manikin with small and large ears) (Algazi *et al.*, 2001) and one measured on the first author of the original paper (Storek *et al.*, 2016) at Lodz University of Technology, Poland (Dobrucki *et al.*, 2010).

With five stimuli, four positioning methods, and three HRTF sets to obtain DHRTF; we had sixty virtually positioned sounds to be evaluated by listeners. The stimuli were grouped as such: one group contained only one type of stimulus positioned using one particular set of HRTF, giving fifteen stimuli groups overall. In case we would have chosen full pair comparison design to be used in the experiment we would get ninety stimuli pairs to evaluate, according to the equation:

$$n_p = n_s \cdot n_{\mathrm{HRTF}} \cdot \frac{n_{\mathrm{stim}} \cdot (n_{\mathrm{stim}} - 1)}{2}\,, \tag{13.1}$$

where n_p is a sum of pairs to evaluate, n_s a number of stimuli types (five), n_{HRTF} is a number of HRTF sets used (three) to obtain DHRTF, and n_{stim} is a sum of all virtually positioned sounds in one stimuli group, which is equal to a number of methods to make virtual sound used in the experiment (four). With this amount of stimuli pairs (ninety), we were concerned about the length of the listening session. Which could result in possible bias due to the fatigue of the listeners and affect their overall comfort. Therefore, we used an adaptive square design instead. The method, which was capable of reducing the number of all possible stimuli pairs, and maintaining the accuracy of subjective results (Li *et al.*, 2013). The number of stimuli pairs needed for the adaptive square design was calculated as follows (Li *et al.*, 2013):

$$n_{\mathrm{asp}} = n_s \cdot n_{\mathrm{HRTF}} \cdot n_{\mathrm{stim}} \cdot (\sqrt{n_{\mathrm{stim}}} - 1), \tag{13.2}$$

giving sixty stimuli pairs to evaluate. For each listener the stimuli pair order was randomized within and also inside the pair.

After each listener session, the pair square matrix was updated according to the scores calculated by the Bradley-Terry (B-T) model implemented in Matlab (Wickelmaier and Schmid, 2004). A new arrangement of the square matrix allowed the next listener to compare stimuli pairs, which are similar in quality, following the adaptive square design principle (Li *et al.*, 2013).

13.1.2 Subjects - DHRTF experiment

DHRTF - Experiment A

Twenty (including seven females) participated in the Experiment A. The participants were aged between 19 and 46 years. Subjects had no or little prior experience with this type of listening test.

DHRTF - Experiment B

Eleven (including four females) participated in the Experiment B. Three subjects participated in both experiments; the rest had no or little prior experience with this type of listening test. The participants were aged between 21 and 46 years.

In both experiments, listeners had their pure-tone hearing thresholds within a range of 15 dB hearing levels for frequencies between 0.25 and 8 kHz (normal hearing).

13.1.3 Stimuli - DHRTF experiment

DHRTF - Experiment A

Five different types of stimuli were chosen, with their length ranging from 1.7 to 2.4 s, including guitar chord, snare drum phrase, white noise, singing segment, and speech segment. The selection was made to provide different spectral and temporal content within each stimulus. The DHRTF artifacts are sparse across the horizontal plane. Therefore, we choose to spatialize each stimulus dynamically across the whole front horizontal plane. Thus, the stimulus was split by an overlap-add method using the Hamming window (Bosi and Goldberg, 2002) into frames. Each frame was convolved with DHRIR corresponding to the desired virtual sound source angle and then merged with the next frames whose virtual direction changed slightly and therefore implied moving sound source illusion to the listener. For more details regarding the dynamic sound source positioning, please refer to Storek (2014).

The use of different DHRTF sets and stimuli types can introduce loudness differences between each test stimuli, which can lead to a bias in subjective preference of each. Therefore, each test signal's loudness was matched according to ITU-R BS.1770-2 recommendation (ITU-R, 2011) to -23 dB LUFS in Adobe Audition program environment.

DHRTF - Experiment B

The results from Experiment A showed little significant differences between the DHRTF artifact methods. Therefore, in Experiment B the number of virtual positions of each signal was reduced to emphasize ones with occurring artifacts. Since the artifacts in DHRTF occur most frequently for positions ranging between 0° and 30°, this range was used while preparing stimuli for the Experiment B. The Lodz HRTF set was chosen as it shows the highest artifacts occurrence. Stimuli from Experiment A were used except for the guitar record, which was replaced by a record of the saxophone. This change was motivated by the results of Experiment A.

The use of different DHRTF sets and stimuli types can introduce loudness differences between each test stimuli, which can lead to a bias in subjective preference of each. Therefore, the loudness of each test signal was matched according to ITU-R BS.1770-2 recommendation (ITU-R, 2011) to -23 dB LUFS in Adobe Audition program environment.

13.2 Results - DHRTF experiment

The results from the subjective evaluation of Experiments A and B were contained in the form of pair preference matrix for each type of stimuli and HRTF set used, giving us fifteen and five matrices for each experiment. However, it is more desirable to have the results in the form of a continuous scale than in preference matrices. For this purpose, a Matlab implementation of the Bradley-Terry model (Wickelmaier and Schmid, 2004) was utilized. The output of the model (B-T score) is a maximum likelihood estimation of the scale parameters, i.e., the preference scales for the four different artifact reduction methods with 95 % confidence intervals. The scale is logarithmic, and the output values are negative; therefore, the least preferred method has the largest absolute value. Since the B-T parameters are unique up to multiplication by a positive constant (Wickelmaier and Schmid, 2004), they were normalized in order to achieve a B-T score of -10 for unprocessed (pure) DHRTF in all cases.

The goodness of fit of the choice models was evaluated (H_0) that the B-T model holds the data), and the p-value was calculated. The Bradley-Terry model also allowed us to determine whether the subjective response to different DHRTF positioning methods was statistically different from uniform. The test T_U is distributed approximately χ^2 with degrees of freedom equal to three in our case, i.e. $\chi^2(3)$. With given degrees of freedom, the data is considered with 95 % statistical confidence non-uniform when condition $T_U > 7.82$ is fulfilled (Handley, 2004).

13.2.1 DHRTF Experiment A

In the Table 13.1 statistics for all stimuli group responses in Experiment A are summarized, the statistics include the goodness of fit of the B-T model and test for uniformity of predicted scores. The B-T model was able to fit all the presented data. However, the test of uniformity showed that data for guitar stimuli (including all HRTF sets used) and for speech stimulus in case of CIPIC KEMAR Large HRTF set, do not hold for the 95 % significance condition ($T_U > 7.82$). It shows that the stimuli in each group, no matter which DHRTF conditioning method was used, are statistically from the same uniform distribution, and there is no statistically significant difference between them within the group. In other words, the listeners were unable to perceive an audible difference between those stimuli.

In Figure 13.2, the B-T scores (bars), and their 95 % confidence intervals (whiskers) for all stimuli and HRTF sets (one subplot each) evaluated in Experiment A are shown. The B-T score is a logarithmic value; therefore, in our case, the higher the absolute value of the B-T score, the lower the perceived quality by listeners. Additional symbols under whiskers indicate whether the DHRTF conditioning method in use is significantly different from: ★ – Pure DHRTF, ♦ – Spectral limitation, ● – Low-Pass filtering, ▲ – Moving average.

Table 13.1: Experiment A: Evaluated B-T goodness of fit test statistics and corresponding p values, and results of whether the obtained data statistically significantly differ from uniform. Note: $^{*}T_U <$ 7.82 .

Stimulus	HRTF set	$\chi^2(3)$	p	T_U
Guitar	KEMAR Small	2.14	0.544	1.78*
	KEMAR Large	3.69	0.296	2.93*
	Lodz	0.35	0.840	1.14*
Noise	KEMAR Small	0.38	0.945	14.45
	KEMAR Large	0.65	0.884	18.30
	Lodz	1.92	0.589	16.48
Singing	KEMAR Small	0.61	0.895	22.82
	KEMAR Large	0.24	0.626	18.38
	Lodz	1.78	0.620	25.37
Snare	KEMAR Small	3.37	0.338	26.08
	KEMAR Large	4.22	0.239	10.92
	Lodz	0.18	0.981	24.23
Speech	KEMAR Small	1.15	0.764	20.11
	KEMAR Large	1.36	0.714	1.38*
	Lodz	1.46	0.692	8.34

13.2.2 DHRTF - Experiment B

The test statistics of the goodness of fit and uniformity of the B-T model scores for Experiment B are summarized in Table 13.2. B-T model fitted all the subjective data. Also, the test for uniformity T_U holds in all cases 95 % significance condition (T_U > 7.82). It concludes that the data are statistically different from a purely random sample from a uniform distribution. Therefore, there exists a clear audible preference over the presented methods by listeners. Figure 13.3 shows the B-T scores (bars) with 95 % confidence intervals (whisker) for all stimuli positioned by DHRTF derived from Lodz HRTF set used in Experiment B. The B-T score is a logarithmic value; therefore, in our case, the higher the absolute value of the B-T score, the lower the perceived quality by listeners. Additional symbols under whiskers indicate whether the DHRTF conditioning method in use is significantly different from: ★ – Pure DHRTF, ♦ – Spectral limitation, ● – Low-Pass filtering, ▲ – Moving average.

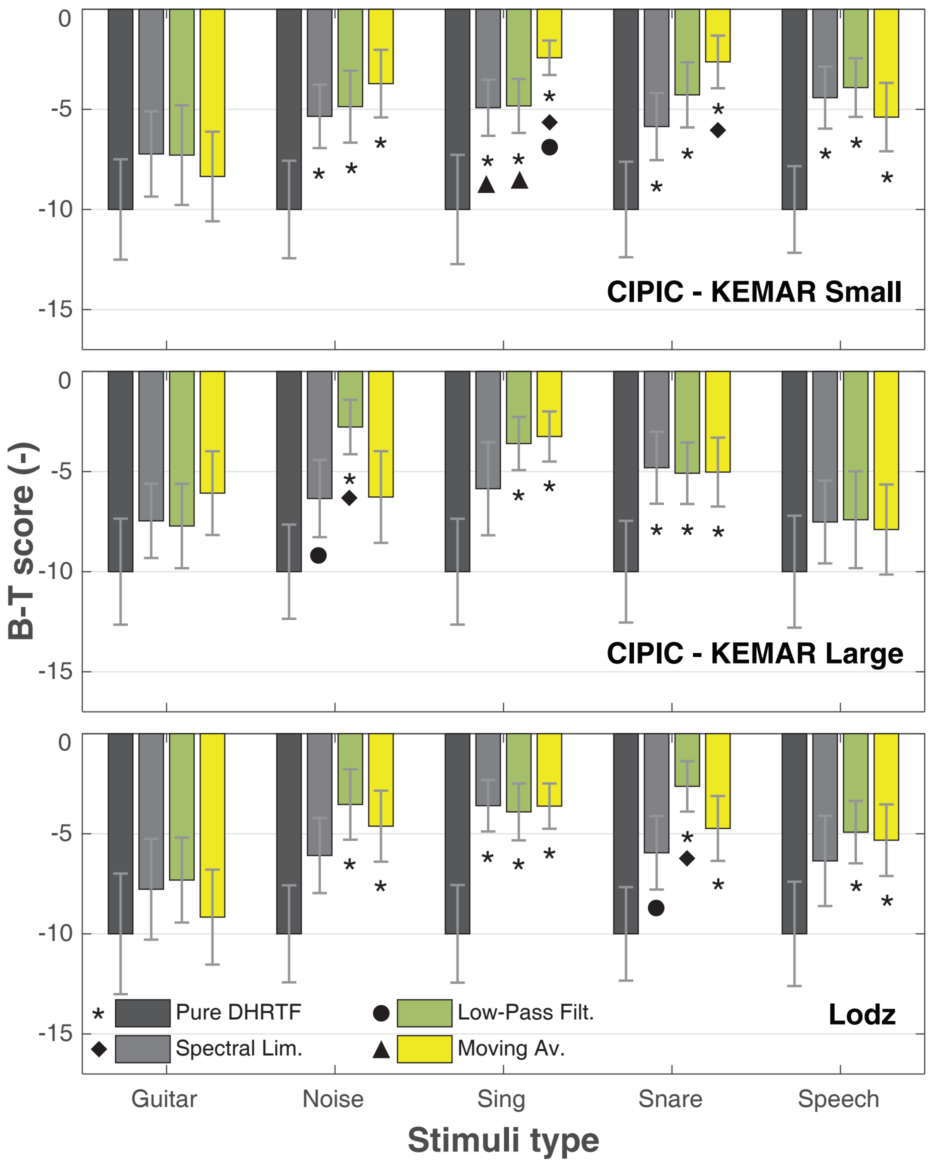

Figure 13.2: Results of the DHRTF Experiment A. Each subplot corresponds to specific HRTF set denoted in the lower left corner. X axis represents particular stimuli, while Y axis denotes B-T score (bars). The whiskers correspond to 95 % confidence intervals. Symbol under the whisker indicates whether the method is significantly different from: ★ – Pure DHRTF, ◆ – Spectral limitation, ● – Low-Pass filtering, ▲ – Moving average.

13.3 Discussion - DHRTF experiment

The results from the DHRTF Experiment A showed that each of the DHRTF conditioning methods, in most cases, improved perceived audible quality significantly over untreated DHRTF. Even the most straightforward method, hard limiting, provided a robust basis of the artifact

Table 13.2: Experiment B: Evaluated B-T goodness of fit test statistics and corresponding p values, and results of whether the obtained data statistically significantly differ from uniform. Note: $^*T_U <$ 7.82 .

	$\chi^2(3)$	p	T_U
Sax	1.17	0.760	28.9580
Noise	1.74	0.627	17.3764
Singing	2.10	0.552	25.9340
Snare	1.40	0.705	20.7524
Speech	1.63	0.654	18.3303

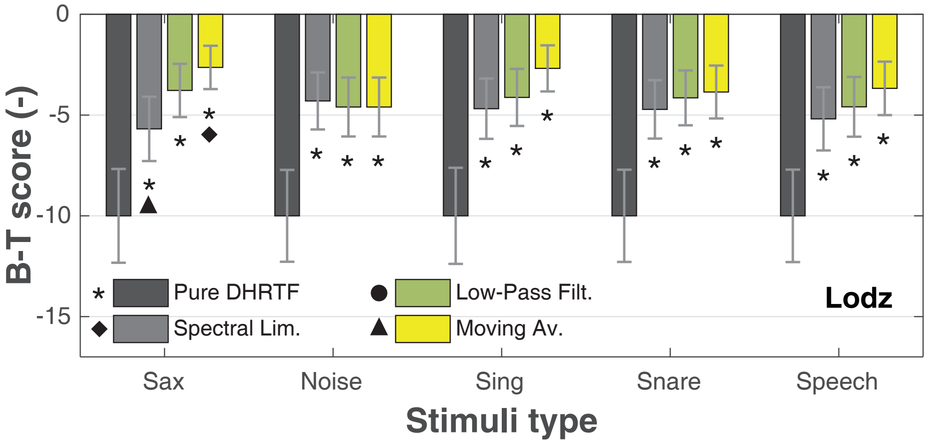

Figure 13.3: Results of the DHRTF Experiment B. They corresponds to Lodz HRTF set used to generate DHRTF. X axis represents particular stimuli, while Y axis denotes B-T score (bars). The whiskers correspond to 95 % confidence intervals. Symbol under whisker shows whether the method is significantly different from: ★ – Pure DHRTF, ◆ – Spectral limitation, ● – Low-Pass filtering, ▲ – Moving average.

reduction. However, for some combination of the HRTF sets and stimulus, the artifacts remained still audible (see Figure 13.2 for reference). The additional smoothing of DHRTF function by either low-pass filtering or spectral moving average improved the perceptible quality even further.

In the DHRTF Experiment B, we focused on emphasizing the artifacts occurring in DHRTF processing. Therefore, we reduced a range of possible sound source positions to the ones where artifacts occurred the most. As in the case of the Experiment A, all conditioning methods outperformed untreated DHRTF significantly by audible quality. However, except for the results for sax stimulus, there was no statistically significant preference between the methods themselves. This might have caused by a smaller pool of listeners than in the Experiment A, and HRTF set used.

In general, the results are highly stimuli and HRTF set dependant. The stimuli had both temporal and spectral different content, which could audibly mask some of the DHRTF induced artifacts. For instance, the tone-like guitar stimulus shows a tiny difference across the artifact reduction methods. This probably results from the discrete character of its tonal spectrum. Therefore, emphasizing the higher harmonics (timbre affection) need not be perceived as unpleasant distortion in this case. Stimuli, with a wider range of higher frequency content (snare, singing), refers to a more distinguishable difference among the methods since the artifacts occur in form of well noticeable birdies.

The dependence on the HRTF set is primarily caused by the specific position of the negative ILD in each particular set. The cause of the spike occurrence may result from uncertainties in HRTF measurement; preserving equal gains of the both measuring microphones is particularly important.

Although the use of a DHRTF might be considered as an obsolete topic, because modern computational systems are powerful enough to process HRTF without a problem, the method to evaluate subjectively perceived quality mentioned in this thesis can be used even in the new systems, for example, to decide which of the ”universal” HRTF set produce least spatial and quality distortion in video games applications.

Part V

Conclusions

CONCLUSIONS

In the presented thesis, novel models of binaural interactions were presented based on the author's publication (Bouse *et al.*, 2019). The models' outputs were compared with subjective data from literature and authors's own experiment. Besides, improvements to these models were also presented. These models can predict the human lateralization data of pure tones, NBNs, ITD thresholds, ILD thresholds, and phase warp thresholds with sufficient accuracy. Furthermore, psychoacoustic experiments conducted by the author were presented, the lateralization of narrow band noises with IPD or ILD (published as a part of Bouse *et al.* (2019)) and subjective evaluation of DHRTF artifacts (published as a part of Storek *et al.* (2016)).

Summary

The thesis is divided into five parts, covering the theoretical basis essential to the thesis aims, the proposal of a binaural model, the binaural model modifications, and presentation of results of own psychoacoustical experiments.

Part I introduces known facts about binaural hearing from both neurophysiological and psychoacoustical point of view. In this part's last chapter, two major binaural model families are presented, containing their own blind spots, which have been addressed in this thesis.

Part II introduces the novel models of binaural interactions based on the author's publication (Bouse *et al.*, 2019). The binaural models sufficiently simulate both neurophysiological data from animals, and the psychoacoustic data of humans.

In Part III, modifications of the models from Part II are introduced. The modified models are verified against the same data as the originals. Furthermore, the differences in the performance of original and modified models are discussed.

Part IV introduces psychoacoustic experiments designed and organized by the author of this thesis. In the first chapter of this part, lateralization experiment with narrow band noises with IPD or ILD is introduced. This experiment was organized in order to validate models from

Part II and was included in the author's publication (Bouse *et al.*, 2019). In the second chapter, Subjective evaluation of DHRTF artifacts is presented. This experiment was part of Storek *et al.* (2016) article, the contribution of the author of this thesis was a design, execution, and evaluation of the experiment.

Contribution of the Thesis

- The functional models of medial and lateral superior olives based on Bouse *et al.* (2019) are presented in this thesis. They fulfill all three main criteria determined in the aims of the thesis:

  1. They are based on the current neurophysiological findings.
     - They are not only based on the neurophysiological findings (Brand *et al.*, 2002; Grothe, 1994, 2003; Joris, 1996; Joris and Yin, 1995; Roberts *et al.*, 2013; Tollin and Yin, 2005) but can reproduce some of them as well (Grothe, 2003; McAlpine *et al.*, 2001), see Chapter 4 for details.
  2. Their outputs give a quantitative representation of subjective lateralization based on ITD or ILD at a corresponding frequency.
     - The models' predictions show a good match with subjective data from literature: pure tone lateralization (Yost, 1981), ITD and ILD discrimination (Brughera *et al.*, 2013; Yost and Dye, 1988), and phase warp discrimination (Dietz *et al.*, 2008). Besides, the models also predict NBN lateralization data obtained by the author of this thesis. In the case of ILD discrimination, the LSO model predicts an experimentally observed (Yost and Dye, 1988) decrease in sensitivity around 1 kHz. See Chapter 6 for further details.
  3. They have low structural complexity and low computational complexity for low power applications.
     - The computational performance of both models can be further increased by reducing the peripheral filters spacing to 1 ERB in a cost of reduced performance in the phase warp detection task. The computational and structural complexity is further improved in the modified MSO and LSO models (see Part III).

- Improvements to the models mentioned above are proposed as well:
  - Simplified peripheral stage and changes in the MSO and LSO models structure further improve computational performance, mentioned in one of the criteria from the aim of this thesis. In some cases, almost twice (see Chapter 10).

  - The improved models are compared with the originals using the same subjective data from the literature (Yost, 1981; Brughera *et al.*, 2013; Yost and Dye, 1988; Dietz *et al.*, 2008). In addition to the test-pool, both revisions of the MSO models predict lateralization data from Zhang and Hartmann (2006). Overall from the models' results performance of both revisions was observed to be similar. However, in the ILD discrimination, the improved LSO model no longer predicts experimentally observed (Yost and Dye, 1988) decrease in ILD sensitivity around 1 kHz.

- Results of two binaural psychoacoustic experiments are presented.
  - The lateralization of narrow band noises with IPD or ILD
    - The subjective results of the NBN experiment were used to validate our MSO and LSO models. Furthermore, they showed a remarkable difference to pure tones' lateralization data with ILD, as discussed in Chapter 12.3.
  - The subjective evaluation of differential head-related transfer function artifacts
    - The subjective experiment showed that even the simplest method to reduce the DHRTF artifacts produced audible improvements to the listeners, which improved the positioning algorithm developed in our department.

Future Work

The research results presented in this thesis are theoretical, and some of the aspects of binaural modeling and hearing are here just briefly covered. The author proposes a brief summary of possible paths worth further exploration:

- The information about lateralization is processed independently for the MSO and LSO models presented in this thesis. Furthermore, the lateralization is calculated in structures of the central stages, which tend to follow the functional aspect rather than the physiology. It is highly unlikely that the processing in inferior colliculus would be different for MSO and LSO. Therefore, a path to design one central stage for both MSO and LSO models based on neurophysiology should be the next step in the development.

- In Koshkina and Bouse (2016, 2017); Koshkina (2017); Melechovský *et al.* (2018), the localization algorithms using the presented models accompanied by a simple machine learning stages were presented. These algorithms showed a similar decay in performance degradation towards extreme azimuths to subjects in Mills (1972). Even with those unchanged algorithms, it might be worth constructing a humanoid-like robot, which would turn to the sound source with its head in real-time with similar performance to humans. Alternatively, from an audio-engineering point of view, they can be used for objective–evaluations of virtual sound positioning algorithms.

- In Chapter 12, we showed that the response of the listeners to narrow band noises had strange 'stickiness' to zero for low ILDs, while the responses for pure tones with similar ILDs do not (Yost, 1981). We proposed a hypothesis that this 'stickiness' might be caused by the envelope fluctuation of the NBN noise, which could deteriorate the hearing system's ability to map small changes in the ILD to the lateral displacement. However, more experiments should be done on more subjects and different laboratories to prove this hypothesis.

Bibliography

Alberti, P. (**2001**). "The anatomy and physiology of the ear and hearing," in *Occupational exposure to noise: evaluation, prevention and control*, edited by B. Goelzer, C. H. Hansen, and G. A. Sehrndt (World Health Organization, Geneva), pp. 53–62.

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

ANSI S3.6-1996 (**1996**). *American national specifications for audiometers (ANSI S3.6-1996)* (American National Standard Institute, New York).

ASA (**2019**). *Code of Ethics* (American Sociological Association), http://www.asanet.org/code-ethics.

Batteau, D. W. (**1967**). "The role of the pinna in human localization," Procceeding of the Royal Society, B, Biological Sciences **168**(1011), 158–180.

Bernstein, L. R., and Trahiotis, C. (**1985**). "Lateralization of low-frequency complex wave-forms: The use of envelope-based temporal disparities," The Journal of the Acoustical Society of America **77**, 1868–1880.

Bernstein, L. R., van de Par, S., and Trahiotis, C. (**1999**). "The normalized interaural correlation: Accounting for Nosπ thresholds obtained with Gaussian and "low-noise" masking noise," The Journal of the Acoustical Society of America **106**(2), 870–876.

Bilinski, P., Ahrens, J., Thomas, M. R., Tashev, I. J., and Platt, J. C. (**2014**). "HRTF magnitude synthesis via sparse representation of anthropometric features," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 4468–4472.

Blauert, J. (**1997**). *Spatial Hearing - The psychophysics of human sound localization* (MIT Press, Cambridge).

Blauert, J. (**2013**). *The technology of binaural listening* (Springer, Berlin).

Bosi, M., and Goldberg, R. E. (**2002**). *Introduction to digital audio coding and standards* (Kluwer Academic Publishers, Norwell, MA, USA).

Bouse, J. (**2015**). "Headphone measurement tool implemented in Matlab," in *Proc. of 19th International Scientific Student Conference POSTER 2015*, Czech Technical University in Prague, p. 5.

Bouse, J., and Schimmel, J. (**2017**). "Lateralization of pure tones: ITD vs. IPD salience" Presented at: The 3rd Workshop on Cognitive neuroscience of auditory and cross-modal perception, Institute of Computer Science at Faculty of Science, Safarik University.

Bouse, J., and Vencovsky, V. (**2015**). "Two-channel models of medial and superior olive based on psychoacoustics," BMC Neuroscience **16(Suppl 1)**.

Bouse, J., Vencovsky, V., Rund, F., and Marsalek, P. (**2019**). "Functional rate-code models of the auditory brainstem for predicting lateralization and discrimination data of human binaural perception," The Journal of the Acoustical Society of America **145**(1), 1–15.

Braasch, J., Clapp, S., Parks, A., Pastore, T., and Xiang, N. (**2013**). "A binaural model that analyses acoustic spaces and stereophonic reproduction systems by utilizing head rotations," in *The technology of binaural listening*, edited by Blauert (Springer, Berlin, Heidelberg), pp. 201–223.

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (**2002**). "Precise inhibition is essential for microsecond interaural time difference coding," Nature **417**, 543–547.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**). "Binaural processing model based on contralateral inhibition. I. model structure," The Journal of the Acoustical Society of America **110**, 1074–1088.

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

Bures, Z. (**2012**). "The stochastic properties of input spike trains control neuronal arithmetic," Biological Cybernetics **106**, 111–122.

Bures, Z., and Marsalek, P. (**2013**). "On the precision of neural computation with interaural level differences in the lateral superior olive," Brain Research **1536**, 16–26.

Campbell, K. B., and Bartoli, E. A. (**1986**). "Human auditory evoked potentials during natural sleep: The early components," Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section **65**(2), 142 – 149.

Campbell, R. A. A., King, A. J., Nodal, F. R., Schnupp, J. W. H., Carlile, S., and Doubell, T. P. (**2008**). "Virtual adult ears reveal the roles of acoustical factors and experience in auditory space map development," Journal of Neuroscience **28**(45), 11557–11570.

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

Cherry, E. C., and Sayers, B. M. (**1956**). "Human cross-correlator – A technique for measuring certain parameters of speech perception," The Journal of the Acoustical Society of America **28**, 889–895.

Colburn, H. S. (**1977**). “Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise,” The Journal of the Acoustical Society of America **61**, 525–533.

Colburn, H. S. (**1978**). “Models of binaural interaction,” in *Handbook of perception, vol. IV*, edited by E. Carterette and M. Friedman (Academic Press, San Diego, CA, USA), pp. 467–518.

Dietz, M., Ewert, S. D., and Hohmann, V. (**2009**). “Lateralization of stimuli with independent fine-structure and envelope-based temporal disparities,” The Journal of the Acoustical Society of America **125**, 1622–1635.

Dietz, M., Ewert, S. D., and Hohmann, V. (**2011**). “Auditory model based direction estimation of concurrent speakers from binaural signals,” Speech Communication **53**, 592–605.

Dietz, M., Ewert, S. D., Hohmann, V., and Kollmeier, B. (**2008**). “Coding of temporally fluctuating interaural timing disparities in a binaural processing model based on phase differences,” Brain Research **1220**, 234–245.

Dobrucki, A., Plaskota, P., Pruchnicki, P., Pec, M., Bujacz, M., and Strumillo, P. (**2010**). “Measurement system for personalized head-related transfer functions and its verification by virtual source localization trials with visually impaired and sighted individuals,” Journal of the Audio Engineering Society **58**(9), 724–738.

Encke, J., and Hemmert, W. (**2018**). “Extraction of inter-aural time differences using a spiking neuron network model of the medial superior olive,” Frontiers in Neuroscience **12**, 140.

Faller, C., and Merimaa, J. (**2004**). “Source localization in complex listening situations: Selection of binaural cues based on interaural coherence,” The Journal of the Acoustical Society of America **116**, 3075–3089.

Farina, A. (**2007**). “Advancements in impulse response measurements by sine sweeps,” Journal of the Audio Engineering Society .

Gaik, W. (**1993**). “Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling,” The Journal of the Acoustical Society of America **94**, 98–110.

Goode, R. L., Killion, M., Nakamura, K., and Nishihara, S. (**1994**). “New knowledge about the function of the human middle ear: Development of an improved analog model,” American Journal of Otolaryngology **15**, 145–154.

Grantham, D. W. (**1984**). “Interaural intensity discrimination: Insensitivity at 1000 Hz,” The Journal of the Acoustical Society of America **75**, 1191–1194.

Grothe, B. (**1994**). “Interaction of excitation and inhibition in processing of pure tone and amplitude-modulated stimuli in the medial superior olive of the mustached bat,” Journal of Neurophysiology **71**, 706–721.

Grothe, B. (**2003**). “New roles for synaptic inhibition in sound localization,” Nature Reviews Neuroscience **4**, 540–550.

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

Hacker, M. J., and Ratcliff, R. (**1979**). “A revised table of d’ for M-alternative forced choice,” Perception & Psychophysics **26**, 168–170.

Handley, J. (**2004**). “Comparative analysis of Bradley-Terry and Thurstone-Mosteller paired comparison models for image quality assessment,” Proceedings of the IS&T PICS Conference .

Hartmann, W. M., Rakerd, B., and Crawford, Z. D. (**2016**). “Transaural experiments and a revised duplex theory for the localization of low-frequency tones,” The Journal of the Acoustical Society of America **139**, 968–985.

ITU-R (**2011**). *Recommendation ITU-R BS.1770-2 (Algorithms to measure audio programme loudness and true-peak audio)*, `http://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1770-2-201103-S!!PDF-E.pdf`.

Jeffress, L. A. (**1948**). “A place theory of sound localization,” Journal of Comparative and Physiological Psychology **41**, 35–39.

Johannesma, P. I. M. (**1972**). “The pre-response stimulus ensemble of neurons in the cochlear nucleus.,” pp. 58–69.

Joris, P. X. (**1996**). “Envelope coding in the lateral superior olive. II. Characteristic delays and comparison with responses in the medial superior olive,” Journal of Neurophysiology **76**, 2137–2156.

Joris, P. X., and Yin, T. C. T. (**1995**). “Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences,” Journal of Neurophysiology **73**, 1043–1062.

Klumpp, R. G., and Eady, H. R. (**1956**). “Some measurements of interaural time difference thresholds,” The Journal of the Acoustical Society of America **28**, 859–860.

Koshkina, E. (**2017**). “*Utilization of machine learning in binaural hearing model* .,” Master’s thesis, CTU in Prague, Faculty of Electrical Engineering (Advisors: Rund, F. and Bouse, J.).

Koshkina, E., and Bouse, J. (**2016**). “Lazy learning sound localization algorithm utilizing binaural auditory model,” in *Proc. of 20th International Scientific Student Conference POSTER 2016*, Czech Technical University in Prague, pp. 1–4.

Koshkina, E., and Bouse, J. (**2017**). “Localization in static and dynamic hearing scenarios: Utilization of machine learning and binaural auditory model,” in *Proc. of 21th International Scientific Student Conference POSTER 2017*, Czech Technical University in Prague, pp. 1–5.

Kumar, R. (**2020**). “The Truth about binaural hearing” `https://www.hearingsol.com/articles/binaural-hearing/`, accessed: 2020-02-01.

Laback, B., and Majdak, P. (**2008**). “Binaural jitter improves interaural time-difference sensitivity of cochlear implantees at high pulse rates,” Proceedings of the National Academy of Sciences **105**(2), 814–817.

Langendijk, E. H. A., and Bronkhorst, A. W. (**2002**). "Contribution of spectral cues to human sound localization," The Journal of the Acoustical Society of America **112**(4), 1583–1596.

Larsen, C. H., Lauritsen, D. S., Larsen, J. J., Pilgaard, M., and Madsen, J. B. (**2013**). "Differences in human audio localization performance between a HRTF-and a non-HRTF audio system," in *Proceedings of the 8th audio mostly conference*, pp. 1–8.

Levitt, H. (**1971**). "Transformed Up-Down methods in psychoacoustics," The Journal of the Acoustical Society of America **49**, 467–477.

Li, J., Barkowsky, M., and Le Callet, P. (**2013**). "Boosting paired comparison methodology in measuring visual discomfort of 3DTV: performances of three different designs," Proceedings of SPIE, Stereoscopic Displays and Applications XXIV **8648**.

Lindemann, W. (**1986**). "Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wave front," The Journal of the Acoustical Society of America **80**, 1623–1630.

Lopez-Poveda, E. A., and Meddis, R. (**1996**). "A physical model of sound diffraction and reflections in the human concha," The Journal of the Acoustical Society of America **100**(5), 3248–3259.

Lopez-Poveda, E. A., and Meddis, R. (**2001**). "A human nonlinear cochlear filterbank," The Journal of the Acoustical Society of America **110**, 3107–3118.

Louden, R. B., and Kuehn, M., eds. (**2006**). *Kant: Anthropology from a pragmatic point of view* (Cambridge University Press), cambridge Books Online.

Lundqvist, L.-M., and Eriksson, L. (**2019**). "Age, cognitive load, and multimodal effects on driver response to directional warning," Applied ergonomics **76**, 147–154.

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

McAlpine, D., and Grothe, B. (**2003**). "Sound localization and delay lines – Do mammals fit the model?," Trends in Neurosciences **26**, 347–350.

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

Melechovský, J., Bouše, J., Rund, F., and Koshkina, E. (**2018**). "Drum kit sound localization tests on binaural hearing model with ANN," in *2018 28th International Conference Radioelektronika (RADIOELEKTRONIKA)*, pp. 1–5.

Mills, A. W. (**1960**). "Lateralization of high frequency tones," The Journal of the Acoustical Society of America. **32**, 132–134.

Mills, A. W. (**1972**). "Auditory localization," Foundations of modern auditory theory 303–348.

Moore, B. C. J. (**2003**). *An Introduction to the psychology of hearing*, 5th ed. (Academia, San Diego).

Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," The Journal of the Acoustical Society of America **74**, 750–753.

Pec, M., Bujacz, M., Strumillo, P., and Materka, A. (**2008**). "Individual HRTF measurements for accurate obstacle sonification in an electronic travel aid for the blind," in *2008 International Conference on Signals and Electronic Systems*, pp. 235–238.

Pecka, M., Brand, A., Behrend, O., and Grothe, B. (**2008**). "Interaural time difference processing in the mammalian medial superior olive: The role of glycinergic inhibition," The Journal of Neuroscience **28**, 6914–6925.

Pralong, D., and Carlile, S. (**1996**). "The role of individualized headphone calibration for the generation of high fidelity virtual auditory space," The Journal of the Acoustical Society of America **100**, 3785–3793.

Prokopiou, A., Moncada-Torres, A., Wouters, J., and Francart, T. (**2017**). "Functional modelling of interaural time difference discrimination in acoustical and electrical hearing," Journal of Neural Engineering **14**, 1–21.

Pulkki, V., and Hirvonen, T. (**2009**). "Functional count-comparison model for binaural decoding," Acta Acustica united with Acustica **95**, 883–900.

Rayleigh, O. M. (**1907**). "On our perception of sound direction," Philosophical Magazine **13**, 214–232.

Roberts, M. T., Seeman, S. C., and Golding, N. L. (**2013**). "A mechanistic understanding of the role of feedforward inhibition in the mammalian sound localization circuitry," Neuron. **78**, 923–935.

Sakitt, B. (**1973**). "Indices of discriminability," Nature **241**, 133–134.

Salminen, N. H., Tiitinen, H., Yrttiaho, S., and May, P. J. C. (**2010**). "The neural code for interaural time difference in human auditory cortex," The Journal of the Acoustical Society of America **127**, EL60–EL65.

Sayers, B. M. (**1964**). "Acoustic-image lateralization judgments with binaural tones," The Journal of the Acoustical Society of America **36**(5), 923–926.

Silbernagl, S., and Despopoulos, A. (**2009**). *Color atlas of physiology* (Thieme).

Siveke, I., Ewert, S. D., Grothe, B., and Wiegrebe, L. (**2008**). "Psychophysical and physiological evidence for fast binaural processing," The Journal of Neuroscience **28**, 2043–2052.

Slattery, W. H., and Middlebrooks, J. C. (**1994**). "Monaural sound localization: Acute versus chronic unilateral impairment," Hearing Research **75**(1), 38 – 46.

Søndergaard, P., and Majdak, P. (**2013**). "The auditory modeling toolbox," in *The technology of binaural listening*, edited by Blauert (Springer, Berlin, Heidelberg), pp. 33–56.

Steinhauser, A. (**1879**). "XLII. the theory of binaural audition. A contribution to the theory of sound," The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **7**(43), 261–274.

Stern, R. M., and Colburn, H. S. (**1978**). "Theory of binaural interaction based in auditory-nerve data. IV. A model for subjective lateral position," The Journal of the Acoustical Society of America **64**, 127–140.

Stevens, S. S., and Newman, E. B. (**1936**). "The location of actual sources of sound," The American Journal of Psychiatry **48**, 297–306.

Storek, D. (**2014**). "Rendering moving sound source for headphone-based virtual acoustic reality aspects of signal processing implementation," in *Applied Electronics (AE), 2014 International Conference on*, IEEE, pp. 271–276.

Storek, D., Bouse, J., Rund, F., and Marsalek, P. (**2016**). "Artifact reduction in positioning algorithm using Differential HRTF," Journal of the Audio Engineering Society **64**(4), 208–217.

Takanen, M., Santala, O., and Pulkki, V. (**2014**). "Visualization of functional count-comparison-based binaural auditory model output," Hearing Research **309**, 147–163.

Thompson, S. K., von Kriegstein, K., Deane-Pratt, A., Marquardt, T., Deichmann, R., Griffiths, T. D., and McAlpine, D. (**2006**). "Representation of interaural time delay in the human auditory midbrain," Nature Neuroscience **9**, 1096–1098.

Tianyi Yan, and Jinglong Wu (**2007**). "The contribution of pinna to the discriminate the vertical angle for virtual reality technology," in *SICE Annual Conference 2007*, pp. 3080–3083.

Tollin, D. J. (**2003**). "The lateral superior olive: A functional role in sound source localization," Neuroscientist **9**, 127–143.

Tollin, D. J., and Yin, T. C. T. (**2002**). "The coding of spatial location by single units in the lateral superior olive of the cat. I. Spatial receptive fields in azimuth," Journal of Neuroscience **22**(4), 1454–1467.

Tollin, D. J., and Yin, T. T. (**2005**). "Interaural phase and level difference sensitivity in low-frequency neurons in the lateral superior olive," The Journal of Neuroscience **25**, 10648–10657.

van Bergeijk, W. A. (**1962**). "Variation on a theme of Bekesy: A model of binaural interaction," The Journal of the Acoustical Society of America **34**, 1431–1437.

Vetter, D. E. (**2015**). "The mammalian olivocochlear system —a legacy of non-cerebellar research in the mugnaini lab," The Cerebellum **14**(5), 557–569.

von Békésy, G. (**1930**). "Zur Theorie des Hörens. Über das Richtungshören bei einer Zeitdifferenz oder Lautstärkenungleichheit der beiderseitigen Schalleinwirkungen," Physikalische Zeitschrift **31**, 824–835.

Wallach, H. (**1939**). "On sound localization," The Journal of the Acoustical Society of America **10**(4), 270–274.

Watanabe, K., Ozawa, K., Iwaya, Y., Suzuki, Y., and Aso, K. (**2007**). "Estimation of interaural level difference based on anthropometry and its effect on sound localization," The Journal of the Acoustical Society of America **122**(5), 2832–2841.

Weiss, T. F., and Rose, C. (**1988**). “A comparison of synchronization filters in different auditory receptor organs,” Hearing Research **33**, 175–179.

Wickelmaier, F., and Schmid, C. (**2004**). “A Matlab function to estimate choice model parameters from paired-comparison data,” Behavior Research Methods, Instruments, & Computers **36**(1), 29–40.

Wiener, F. M., and Ross, D. A. (**1946**). “The pressure distribution in the auditory canal in a progressive sound field,” The Journal of the Acoustical Society of America **18**, 401–408.

Wightman, F. L., and Kistler, D. J. (**1992**). “The dominant role of low-frequency interaural time differences in sound localization,” The Journal of the Acoustical Society of America **91**, 1648–1661.

Yost, W. A. (**1981**). “Lateral position of sinusoids presented with interaural intensive and temporal differences,” The Journal of the Acoustical Society of America **70**, 337–409.

Yost, W. A., and Dye, Jr., R. H. (**1988**). “Discrimination of interaural differences of level as a function of frequency,” The Journal of the Acoustical Society of America **83**, 1846–1851.

Zhang, P. X., and Hartmann, W. M. (**2006**). “Lateralization of sine tones–interaural time vs phase,” The Journal of the Acoustical Society of America **120**(6), 3471–3474.

Zwislocki, J., and Feldman, R. S. (**1956**). “Just noticeable differences in dichotic phase,” The Journal of the Acoustical Society of America **28**, 860–864.

PUBLICATIONS OF THE AUTHOR RELEVANT TO THE THESIS

The author has participated in these grant projects:

- SGS20/180/OHK3/3T/13
  - Audio Signal Processing Related to its Perception.
- SGS17/190/OHK3/3T/13
  - Sound Quality in Connection with Characteristics of Human Hearing System.
- ASA International Student Grant to assist the research of promising graduate students in acoustics
  - Computational models of binaural interaction in sound localization.
- SGS14/204/OHK3/3T/13
  - Binaural models and measurements.
- SGS11/159/OHK3/3T/13
  - Personal Navigation System using Artificial Audio Signals for Assistive Technology.

Journals with Impact Factor:

Bouse, J., Vencovsky, V., Rund, F., and Marsalek, P. (**2019**). "Functional rate-code models of the auditory brainstem for predicting lateralization and discrimination data of human binaural perception" The Journal of the Acoustical Society of America **145**(1), 1–15

Shares: **25**/25/25/25

Storek, D., Bouse, J., Rund, F., and Marsalek, P. (**2016**). "Artifact reduction in positioning algorithm using Differential HRTF" Journal of the Audio Engineering Society **64**(4), 208–217

Shares: 35/**35**/20/10

Indexed in ISI:

Melechovský, J., Bouše, J., Rund, F., and Koshkina, E. (**2018**). “Drum kit sound localization tests on binaural hearing model with ANN” in *2018 28th International Conference Radioelektronika (RADIOELEKTRONIKA)*, pp. 1–5

Shares: 25/**25**/25/25

Other Relevant Publications - honored:

Koshkina, E., and Bouse, J. (**2017**). “Localization in static and dynamic hearing scenarios: Utilization of machine learning and binaural auditory model” in *Proc. of 21th International Scientific Student Conference POSTER 2017*, Czech Technical University in Prague, pp. 1–5.

Shares: 50/**50**

Koshkina, E., and Bouse, J. (**2016**). “Lazy learning sound localization algorithm Utilizing binaural auditory model” in *Proc. of 20th International Scientific Student Conference POSTER 2016*, Czech Technical University in Prague, pp. 1–4.

Shares: 50/**50**

Bouse, J. (**2015**). “Headphone measurement tool implemented in Matlab” in *Proc. of 19th International Scientific Student Conference POSTER 2015*, Czech Technical University in Prague, p. 5.

Shares: **100**

Other Relevant Publications:

Bouse, J., and Vencovsky, V. (**2015**b). “Two-channel models of medial and superior olive based on psychoacoustics” BMC Neuroscience **16(Suppl 1)**.

Shares: **50**/50

Supka, O., Rund, F., and Bouse, J. (**2014**). “Automatized HRTF measurement system implemented in MATLAB” in *Proc. of 22nd Annual Conference Proceedings Technical Computing Bratislava.*

Shares: 33/33/**33**

Lindner, T., and Bouse, J. (**2014**). “Optimization in measuring and analysis of head-related transfer function” in *Proc. of 18th International Student Conference on Electrical Engineering POSTER.*

Shares: 50/**50**

Rund, F., Bouse, J., and Barath, T. (**2013**). “Comprehensive Matlab tool for HRTF measurement and virtual auditory space testing” in *Proc. of 21th Annual Conference Proceedings Technical Computing Prague.*

Shares: 33/**33**/33

Bouse, J., and Vencovsky, V. (**2013**). “The Matlab implementation of binaural orocessing model simulating lateral position of tones with interaural time differences” in *Proc. of 21th Annual Conference Proceedings Technical Computing Prague.*

Shares: **50**/50

Bouse, J., and Vencovsky, V. (**2012**). "Implementation of binaural processing model" in *Proc. of 16th International Student Conference on Electrical Engineering POSTER.*

Shares: **50**/50

Bouse, J., and Vencovsky, V. (**2011**). "Matlab implementation of the count-comparison lso model" in *Proc. of 19th Annual Conference Proceedings Technical Computing Prague.*

Shares: **50**/50

Vencovsky, V., and Bouse, J. (**2011**). "Binaural processing model simulating the lateral position of tones with interaural time differences" in *Proc. of 15th International Student Conference on Electrical Engineering POSTER.*

Shares: 50/**50**

Not published presentations given on conferences or workshops:

Bouse, J., and Schimmel, J. (**2017**). "Lateralization of pure tones: ITD vs. IPD salience" Presented at: The 3rd Workshop on Cognitive neuroscience of auditory and cross-modal perception, Institute of Computer Science at Faculty of Science, Safarik University.

Shares: **90**/10

Bouse, J., and Vencovsky, V. (**2015**a). "Functional two channel models of medial and lateral superior olive" Presented at: The Auditory Model Workshop, Carl Ossietzky University Oldenburg.

Shares: **50**/50

Bouse, J. (**2014**). "Binaural model of lateralization" Presented at: Workshop of Cognitive neuroscience of auditory and cross-modal perception, Institute of Computer Science at Faculty of Science, Safarik University.

Shares: **100**

Bouse, J. (**2013**). "Binaural auditory model" Presented at: 3rd SPLab Workshop, Signal Processing Laboratory, Brno University of Technology.

Shares: **100**

REMAINING PUBLICATIONS OF THE AUTHOR

Indexed in ISI:

Rund, F., Vencovský, V., and Bouše, J. (**2016**). "Detection of clicks in analog recordings using peripheral-ear model" in *19th International Conference on Digital Audio Effects (DAFx16)*, Brno University of Technology, Brno, CZ

Shares: 33/33/**33**

Rund, F., Khaddour, H., Schimmel, J., and Bouše, J. (**2015**). "Objective quality assessment for the acoustic zoom" in *38th International Conference on Telecommunications and Signal Processing*, IEEE, Piscataway, US

Shares: 25/25/25/**25**