Charles University in Prague

First Faculty of Medicine

Study Program: Biomedical Informatics

Mgr. Peter Tóth

Modeling of Binaural Hearing

Modelování binaurálního slyšení

Doctoral Thesis

Thesis Supervisor: prof. MUDr. RNDr. Petr Maršálek, PhD.

Thesis Consultant: doc. Ing. Zbyněk Bureš, Ph.D.

Prague 2020

Declaration

I declare that the thesis has been composed by myself and that I have provided references on all supporting literatures and resources. I also declare that this work has not been submitted for any other or the same degree.

I agree with the permanent depository of the electronic version of my thesis in the database of interuniversity project Theses.cz in order to allow permanent inspection of the resemblance of theses.

Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem řádně uvedl a citoval všechny použité prameny a literaturu. Současně prohlašuji, že práce nebyla využita k získání jiného nebo stejného titulu.

Souhlasím s trvalým uložením elektronické verze mé práce v databázi systému meziuniverzitního projektu Theses.cz za účelem soustavné kontroly podobnosti kvalifikačních prací.

In Prague/V Praze, 30.06.2020

Peter Tóth

Identification record

TÓTH, Peter. Modeling of binaural hearing. [Modelování binaurálního slyšení]. Prague, 2020. 52 pages, 3 appendices. Doctoral Thesis. Charles University, First Faculty of Medicine, Institute of Pathological Physiology. Thesis supervisor Maršálek, Petr.

Identifikační záznam

TÓTH, Peter. Modeling of binaural hearing. [Modelování binaurálního slyšení]. Praha, 2020. 52 stran, 3 přílohy. Disertační práce. Univerzita Karlova, 1. lékařská fakulta, Ústav patologické fyziologie. Vedoucí práce Maršálek, Petr.

Acknowledgements

I would like to thank my supervisor, prof. Petr Maršálek, for his patient and caring guidance and support through my postgraduate studies, his shared wisdom, ideas, comments and suggestions.

I would also like to express my gratitude to doc. Norbert Kopčo for the opportunity to participate in international projects and for the welcoming and inspirational environment in his lab.

My special thanks go to co-authors of our publications and other colleagues at the Institute of Pathological Physiology in Prague, the Department of Radio Engineering in Prague, the Perception and Cognition Laboratory in Košice and the Center for Computational Neuroscience and Neural Technology in Boston for their cooperation and for their friendship.

My most sincere gratitude goes to my wife and family for their endless support and care.

I would like to dedicate this work to my grandfather Vasiľ.

Abstract

The central theme of this thesis is a description of information processing in the sound localization circuit of the auditory pathway. The focus is on principal neurons of the medial superior olive (MSO), the first major convergence point for binaural information. Selected properties and relations of MSO neurons are derived and expressed through models.

In the thesis we present three modeling studies. The first one clarifies a relationship between biophysical parameters of the MSO neuron and its ability to detect coincidental spikes from the left and the right ear. The second study describes the statistical behavior of spike trains on the input and output of the MSO neuron. In the third work, we studied how interaural coherence could guide localization of sound sources in complex listening situations with multiple sound sources in reverberant environments.

The main results are analytical and numerical models describing the aforementioned relations and behaviors. Secondary results include that inhibitory input to the MSO neuron narrows and shifts the time range of coincidence detection, that ergodic assumption from statistical physics and circular statistics are beneficial in the description of spike trains in the auditory pathway, and that interaural level difference of parts of the signal with high interaural coherence could explain human localization performance in complex listening scenarios.

Keywords

binaural hearing, binaural neuron, model, sound localization

Abstrakt

Ústředním tématem této práce je popis zpracování informace v lokalizačním obvodu sluchové dráhy. Důraz je kladen na první místo konvergence binaurální informace, neurony mediální olivy superior (MSO). Vybrané vlastnosti a vztahy neuronů MSO jsou odvozeny a vyjádřeny prostřednictvím modelů.

V disertační práci uvádíme tři modelovací studie. První objasňuje vztah mezi biofyzikálními parametry neuronu MSO a jeho schopností detekovat simultánní signály z levého a pravého ucha. Druhá studie popisuje statistické distribuce vzruchů na vstupu a výstupu neuronu MSO. Ve třetí práci jsme studovali roli interaurální koherence při lokalizaci ve složitých poslechových situacích s více zdroji zvuku v dozvukovém prostředí.

Hlavní výsledky jsou analytické a numerické modely popisující výše uvedené vztahy a chování. Aplikací modelů jsme získali sekundární výsledky: (1) inhibiční vstup do neuronu MSO zužuje a posunuje časový rozsah detekce simultánních signálů, (2) ergodický předpoklad ze statistické fyziky a cirkulární statistika jsou vhodné nástroje při popisu vzruchů v sluchové dráze a (3) hlasitostní rozdíl v části signálu s vysokou interaurální koherencí může vysvětlit přesnost lidské lokalizace ve složitých poslechových situacích.

Klíčová slova

binaurální slyšení, binaurální neuron, model, lokalizace zvuku

Contents

List of Abbreviations

AP	Action Potential
BRIR	Binaural Room Impulse Response
CD	Coincidence Detector
DCN	Dorsal Cochlear Nucleus
ECD	Excitatory Coincidence Detector
EPSP	Excitatory Postsynaptic Potential
HRIR	Head-Related Impulse Response
HRTF	Head-Related Transfer Function
IC	Interaural Coherence
ICD	Inhibitory Coincidence Detector
ILD	Interaural Level Difference
ITD	Interaural Time Difference
IPD	Interaural Phase Difference
IPSP	Inhibitory Postsynaptic Potential
ISI	InterSpike Interval
LSO	Lateral Superior Olive
MNTB	Medial Nucleus of the Trapezoid Body
MSO	Medial Superior Olive
PDF	Probability Density Function
PSG	Postsynaptic Membrane Conductance
PSP	Postsynaptic Potential
RMSE	Root Mean Square Error
SRM	Spike Response Model
TMR	Target-to-Masker Ratio
VCN	Ventral Cochlear Nucleus

Introduction

The simple question "why do we have two ears?" may have a simple answer: "so we can hear better". However, if we want to know how this "better hearing" is achieved the answer is not that simple and may trigger many more questions.

This question was scientifically addressed with the advent of psychoacoustics in the nineteenth century. It did not take long to realize that the main advantage of a pair of ears is their different spatial position. A trivial observation — at least for high frequencies or complex sounds — was that a sound is perceived as louder with an ear closer to the source, thus, we know whether the source is on the left or right side.

In 1907, Lord Rayleigh demonstrated that the human brain can also detect disparities in the phase of sound arriving at individual ears. He stated that the brain utilizes differences in sound pressure level for high frequencies and differences in sound phase for low frequencies (Rayleigh, 1907). This theory was later named the "Duplex theory". For time differences the term Interaural Time Difference (ITD) was coinded, and differences in sound pressure levels were known as Interaural Level Difference (ILD).

After the measurement of human performance in sound localization and the precision in angles translated to corresponding ITD and ILD values (Mills, 1958), and the first electrophysiological measurements of auditory brainstem neurons with microsecond precision (Goldberg and Brown, 1969), an intriguing question arose. How can neural circuits, where the expression of information (action potential) lasts milliseconds, detect time differences as small as ten microseconds (two orders lower than the duration of action potential)?

Computational models may be useful tools for studying brain functions. A well-known example is the famous Jeffress model (Jeffress, 1948). In 1948, Jeffress proposed a neural circuit incorporating asymmetric delays produced by nerve fibers of different lengths (delay lines) and neurons performing time-limited spatial summation of inputs (coincidence detectors, CDs). This model, which could detect very small ITDs, was not based on any anatomical study. In fact, it was proved (in the case of birds) much later (Carr and Konishi, 1988). For mammals, the lack of

anatomical evidence of delay lines and the presence of inhibitory inputs prompted new theories of sound localization (Grothe, 2003). Nevertheless, it is still generally accepted that the first binaural neuron acts as a coincidence detector. The Jeffress model contributed not only to the field of neuroanatomy by proposing which circuit should be searched for, but also to the field of neurocomputing suggesting a new mechanism of how neurons can process information. Afterwards, multiple models were published, we can mention a quantification of Jeffress model (Colburn, 1973) and the equalization-cancellation model (Durlach, 1963), among others (Colburn et al., 1990; Batra et al., 1997; Brand et al., 2002; Hancock and Delgutte, 2004; Marsalek and Lansky, 2005; Zhou et al., 2005; Leibold, 2010; Franken et al., 2014; Ashida et al., 2015). More detailed descriptions of binaural models can be found in the classic review (Colburn and Durlach, 1978) or in more recent reviews (Braasch, 2005; Jennings and Colburn, 2010; Grothe et al., 2010).

A starting-point for the presented work is the model of binaural hearing developed by a group of people around prof. Marsalek over a period of around ten years (Marsalek and Kofranek, 2004; Marsalek and Lansky, 2005; Marsalek and Drapal, 2008; Sanda and Marsalek, 2012). The main focus of their study was the relationship between probability distributions of input and output spike trains, hence the model is named *"Probabilistic delay model"*. In this model, the principal neuron is reduced to a simple mathematical formulation — spikes are represented by point processes and the neuron detects coincidence if and only if spikes are closer in time than a certain constant (coincidence window). The category of models that do not deal with biophysical properties of neurons are also known as black-box models. Nowadays, there is an ongoing debate about how exactly principal neurons of the MSO are performing coincidence detection and which mechanisms are utilized to tune their function. Many extensive neurophysiological studies have been published recently trying to resolve these questions, but the issue is still not clear (Couchman et al., 2010; Jercog et al., 2010; Van der Heijden et al., 2013; Roberts et al., 2013; Myoga et al., 2014). From this point of view, it is important to question whether a black-box model has support in recent neurophysiological findings.

Regarding data from psychoacoustic studies, *Probabilistic delay model* was compared to the Mills study (Mills, 1958; Marsalek and Lansky, 2005) to show that

the model accounts for a dip in localization performance between low and high frequencies. Confrontation of the model with more complex psychoacoustic measurements including multiple speakers and reverberant environments (Kopco et al., 2010) showed that *Probabilistic delay model* alone is not able to explain localization with concurrent sound sources. A higher level model with the ability to separate cues from individual sound sources (Faller and Merimaa, 2004; Dietz et al., 2011) was needed to explain the results of Kopco et al. (2010).

After this introductory chapter, a short outline of principles of binaural hearing and a brief description of information processing in the auditory system follow. The rest of the thesis is our original work consisting of three modeling studies. The first two extend *Probabilistic delay model* of the MSO neuron: The first one clarifies a relationship between biophysical parameters of the MSO neuron and its ability to detect coincidental spikes from the left and the right ear. The second study describes the statistical behavior of spike trains on the input and output of the MSO neuron. In the third work, we studied how interaural coherence could guide localization of sound sources in complex listening situations with multiple sound sources in reverberant environments. The manuscripts of all three studies are enclosed in the appendix.

1. Binaural Hearing

Hearing is the ability to perceive and make sense of sound. Although hearing with one ear (monaural hearing) is possible and in many situations sufficient, hearing with two ears (binaural hearing) provides many advantages. Two main abilities in which binaural hearing has a prominent role are sound source localization and auditory scene analysis.

Precise information about sound source location can be essential. For many nocturnal animals, it is, in fact, a matter of survival. In order to avoid predators or catch prey, the brain utilizes subtle differences in sounds induced by specific location of the sound source.

Although people no longer have to rely often on sound localization to escape danger, communication and social interaction is an important part of our everyday life. Communication often takes place among multiple talkers with the presence of background noise, but the listener is still able to focus on one talker and suppress other sounds (so-called *"Cocktail party effect"*, Cherry, 1953). This effect is explained by auditory scene analysis (Bregman, 1990), in which binaural hearing allows better segregation of sounds from spatially separated sources (binaural release from masking) and suppression of reflected sounds (dereverberation).

This chapter outlines selected underlying principles of binaural hearing.

1.1 Auditory space

To localize an object in a three-dimensional environment, the estimation of three coordinates is needed. The head-centered spherical coordinate system has the origin in the middle of two ears, and these three coordinates sufficient to determine the location of the object are: 1. azimuth, 2. elevation, and 3. distance.

Azimuth is the angle between the medial sagittal plane and the object projected to a transverse (horizontal) plane. The changes in azimuth cause differences in sounds reaching the left and right ear. These differences can be used as cues for azimuth estimation. As they are based on information received from both ears, we call them binaural cues (section 1.2).

Elevation is the angle between the object and its orthogonal projection to a transverse (horizontal) plane. In case of organisms with symmetric ears, the changes in elevation while preserving the same azimuth and distance do not produce significant differences in sounds reaching the left and right ear. Instead, the shape of pinna serves as a filter and modifies the sound spectrum depending on the elevation. Modified spectrum of sound is compared to expected spectrum of sound to estimate elevation. This type of cue does not require two hearing ears, and so is known as a monaural cue.

The last coordinate is the distance between the origin and the object. Estimation of a distance is based mainly on monaural cues. The sound is being attenuated as it propagates through the environment. As this attenuation is frequency dependent, the spectral shape changes with the distance. In reverberant environments, distance can be estimated from the time lag between direct sound and reflection.

Sound sources have their spatial location in the environment determined by three coordinates. Nevertheless, an auditory space can be distorted by reflections, background noise, or interference of other sound sources due to the nature of the sound. Even if these phenomena may be helpful to some extent, generally, they pose a difficult task in the creation of a map of auditory space.

1.2 Binaural cues

Binaural cues are the basis of binaural hearing. The different spatial position of the ears means that the sound in the left ear is slightly different to the sound in the right ear.

As for sounds originating out of the midsagittal plane, the distance from the sound source to the left and right ears is not the same, and the time of arrival of sound differs in both ears. This disparity is called Interaural Time Difference (ITD). The sound is propagating in waves, thus, the temporal disparity after the time of arrival is preserved as Interaural Phase Difference (IPD).

Classical formula for ITD is (Woodworth, 1938):

$$\mathrm{ITD} = \frac{a}{c}(\theta + \sin\theta), \quad -\pi/2 \leq \theta \leq \pi/2, \tag{1.1}$$

where a is the radius of the head, c is the speed of the sound, and θ is the azimuth in radians. This formula assumes a planar wave front which approximately holds for sources at distances greater than $3a$ (Carlile, 1996).

For humans, the typical value of a is 8.75 cm (Hartley and Fry, 1921). With $c =$ 343 m/s this formula yields maximal ITD around 660 μs (for sounds at 90°). Small rodents, such as gerbils or kangaroo rats, have these values even smaller (120 μs or 80 μs, respectively) (Brand et al., 2002).

While the maximal ITD for humans is around 660 μs, detectable ITDs are even smaller. Experiments measuring human localization performance show that humans can achieve precision around 1° (Mills, 1958). This translates to the ability to detect time disparity in the order of tens of microseconds.

ITDs are only slightly frequency-dependent and the range of ITDs is smaller for higher frequencies due to the frequency dispersion of the diffracted waves (Schnupp et al., 2011).

IPD depends on the frequency of the sound and the distance between the ears: $\mathrm{IPD} = \mathrm{ITD}\omega$, where ω is angular frequency. To avoid ambiguity, the period of sound should be longer than ITD. Maximal ITD of 660 μs corresponds to a frequency of 1500 Hz. Consequently, for frequencies above 1500 Hz, IPD is not a reliable cue.

Interaural Level Difference (ILD) is due to the head acting as an acoustic obstacle, reflecting and diffracting the sound wave (Carlile, 1996). The interference of the head is highly frequency-dependent. For sounds with wavelengths greater than the size of the head, the interference is not significant. For wavelengths much smaller than the size of the head, one ear may appear in head shadow while the other can have amplified sound due to reflections from the head and pinna (near ear effect). The ILD for high frequencies can be up to 40 dB. Since the ILD is dependent on azimuth, frequency, distance and exact shape of head and pinnae, there is no simple formula as in case of ITD.

The fact that ITD/IPD is suitable only for low frequency sounds and ILD is suitable only for high frequency sounds led to the Duplex theory. This theory states that humans utilize ITD and IPD for localization of low frequency sounds and ILD for localization of high frequency sounds (Rayleigh, 1907). Experiments proved that duplex theory is accurate in relation to pure tones, but in the case of

more complex tones, both cues are used in a more complicated way. For instance, the auditory system can detect ITD of an envelope of amplitude-modulated high frequency signals.

1.3 Virtual auditory space

Virtual auditory space refers to special artificial stimuli presented over headphones creating the illusion that origins of stimuli are in the surrounding space.

Listening to acoustic stimuli over headphones usually leads to the perception that the sound is originating in the center of the head. If the sound presented over one headphone is louder than the sound in the other (i.e. it has nonzero ILD), listeners can observe lateralization, i.e. the perception that the sound originates inside the head closer to the ear in which the sound is louder. Generally, if the sounds presented over headphones are only modified in overall sound levels and timing (to achieve desired ILD and ITD), the sound is perceived as originating inside the head.

To achieve externalization, the perception that the sound originates outside the head, monaural cues need to be incorporated. The intent is to deliver to the ear tube a sound similar to a real situation where the sound is altered by the presence of head and pinnae.

The simple way to determine effects of head and pinnae is to record sounds using small microphones placed in the ear tubes. As the alterations are direction- and frequency-specific, a convenient way is to play brief broadband sounds (impulses) from different locations in space. The recorded sounds are responses to these impulses and are called head-related impulse responses (HRIRs). HRIRs are usually recorded in the anechoic chamber. When recordings are performed in a reverberant room, these responses also obtains various reflections and are called binaural room impulse responses (BRIRs).

The convolution of a sound with HRIR or BRIR related to specific location places the sound source to this location in virtual space.

Fourier transform of HRIR is called the head-related transfer function (HRTF). HRTF shows how spectrum is altered by the head and pinnae.

2. Information processing in the auditory system

The auditory system is a biological system for processing information contained in sound. It consists of two main parts. First, the periphery (the ears), that transduces mechanical vibrations to electrical impulses conducted by nerves. The second is the central auditory system containing neural circuits that process the information carried by the electrical impulses.

The auditory system varies across different species. Here, the focus is the manner of processing auditory information in mammals and, especially, humans. First, the function of each part of the ear (namely outer ear, middle ear, and inner ear) is described. It is followed by the description of neural pathways and circuits up to the nuclei where the pathways from the left and right ear meet, as the information processing in higher stages of the central auditory system is immensely complex and out of the scope of this work.

2.1 Outer ear

The outer ear is the external, visible part of the ear. It consists of the pinna and the ear canal. The main function of the outer ear is to collect sound and to funnel it to the eardrum. The secondary function is the frequency- and direction-specific amplification of the sound.

The pinna is a prominent fold of cartilage-supported skin. In the case of low frequencies (up to 500 Hz), sound reflected from the pinna amplifies direct sound. For higher frequencies, owing to the pinna's specific asymmetrical shape, amplification or attenuation depends on the vertical direction of sound. Maximal direction-specific modification is for frequencies close to 5000 Hz (Syka et al., 1981). The pinna also creates significant acoustic shade for sounds originating from behind.

The ear canal, or auditory canal, is a tube of approximately 2.5 cm length that ends with the eardrum. Closed tube resonance amplifies a relatively broad band of frequencies, from 2000 Hz to 6000 Hz, with peak amplification of 12 dB for

frequencies between 3000 Hz and 4000 Hz (Syka et al., 1981).

The modification of sound frequency spectrum by the outer ear, together with effect of the head, is usually described with HRTFs (section 1.3).

2.2 Middle ear

The part of the ear between the tympanic membrane and oval window is called the middle ear. Its main function is to efficiently transduce the (longitudinal) compressional waves in the air to (transverse) fluid-membrane waves in the cochlea (Voss et al., 2013). To efficiently transfer the acoustic energy, the system has to compensate for the impedance of cochlear fluid being 4000 times greater than the impedance of air. Impedance matching is secured by ossicles functioning as a lever and the membranes of different areas. This mechanical system is not able to perform with the same efficiency of transfer across the whole frequency spectrum. Transfer function favors frequencies around 1 kHz. During very loud sounds, the reflexive contraction of muscles, called stapedius reflex, reduces the transfer of acoustic energy and protects the inner ear from sound overexposure.

2.3 Inner ear

The two main parts of the inner ear are the cochlea and the vestibular system. The vestibular system provides the sense of balance and does not contribute to the processing of auditory information.

The cochlea is a spiral-shaped bone structure with three liquid-filled ducts separated by two membranes. This hydrodynamic apparatus is capable of mechanical analysis of sound frequency.

The basilar membrane is the main structural component of the cochlea that determines the propagation of mechanical vibrations. Oscillatory pressure differences produced by the motion of stapes propagate in the fluid along the cochlea and create a traveling wave in the basilar membrane. Because the basilar membrane does not have the same mechanical properties along its extent (it is narrow and stiff at the base and wide and floppy at the apex), its maximal displacement happens at a particular location depending on sound frequency, with lower frequencies having

locations further from the base. The base has maximal displacement at frequencies around 20 000 Hz while the apex at frequencies around 20 Hz. Logarithmic arrangement of "best frequencies" along the cochlea resembles a piano keyboard and is called tonotopy (Kandel et al., 2013). Alongside the basilar membrane is the organ of corti containing receptor cells called hair cells. Hair cells are mechanoelectrical transductors that react to mechanical movement in one direction with increased probability of neurotransmitter release. Because the mechanical movement at a particular site of the basilar membrane is proportional to the magnitude of corresponding frequency, each hair cell has its "best frequency" and the probability of response is bigger with bigger magnitude. However, hair cells react also to adjacent frequencies, but the response is smaller.

Although fourier transforms are usually used for spectral analysis, the specificity of spectral analysis of the cochlea is best modeled by set of gammatone filters (Schnupp et al., 2011). Hair cell transduction is usually modeled as a probability of auditory nerve excitation that is derived from gammatone-filtered signal after compression and rectification.

2.4 Auditory nerve

Primary sensory neurons in the auditory system have their bodies in the spiral ganglion and innervate cochlear hair cells. The auditory nerve conveys the information between the hair cells and the brain in a bidirectional manner. Afferent nerve fibers carry signals mostly from inner hair cells to the brain and efferent fibers carry signals from the brain to mainly outer hair cells that act as amplifiers.

At least 90 % of nerve cells innervate inner hair cells. From these, each nerve cell innervates only one inner hair cell, but one inner hair cell has connections to on average 10 nerve cells. Thus, information from one receptor is encoded in several parallel channels and the tonotopy is preserved, that is, each nerve cell can be linked to the frequency that excites it most (Kandel et al., 2013).

Because the hair cells have maximal probability of discharge during the maximal displacement, which happens during the amplitude of a sound wave, auditory nerve cells tend to discharge periodically with sound frequency. This phenomenon is called

phase-locking and, after tonotopy, it is the second way in which information about frequency is represented in the auditory nerve. However, this way of coding is limited to lower frequencies.

The auditory nerve cell does not have to fire with every sound cycle. The magnitude of each frequency is represented by the discharge rates of corresponding nerve cells. To better handle the range of possible sound intensities, there are several types of cells with respect to the threshold of excitation. One type of cells responds to very low intensities and usually have high spontaneous firing rates, and another type of cells responds only to sounds with higher intensities and usually have lower spontaneous firing rates.

2.5 Cochlear nucleus complex

The cochlear nucleus complex is the first relay station in the auditory system. Here, parallel ascending pathways with preserved tonotopy are created and projected to several higher areas. Neurons of the cochlear nucleus complex are the first to perform integration of information across auditory nerve cells.

Two main regions are distinguished in the cochlear nucleus complex, the dorsal cochlear nucleus (DCN) and the ventral cochlear nucleus (VCN).

The dorsal cochlear nucleus is believed to process spectral cues for localization (i.e. cues for elevation) and does not project to the superior olivary complex, thus is not considered to be directly involved in binaural hearing (Kandel et al., 2013).

The ventral cochlear nucleus enhances temporal and spectral information for further processing in other nuclei. It contains different types of cells with morphology specialized to particular functions. Octopus cells have dendrites with a large span summing smaller inputs from a large amount of auditory nerve fibers. Octopus cells are broadly tuned and detect onsets, transients and periodicity of sounds (Kandel et al., 2013). Stelate and bushy cells have short dendrites with a smaller number (up to 10) of strong inputs. They provide sharp tuned and temporally precise firing patterns. Outputs of bushy cells have better synchronization to the sound phase than the auditory nerve (Joris et al., 1994) and project to the superior olivary complex.

2.6 Superior olivary complex

Three primary nuclei of superior olivary complex are the medial superior olive (MSO), the lateral superior olive (LSO), and the medial nucleus of the trapezoid body (MNTB). LSO and MSO neurons are the first neurons where the information from both ears meets.

The function of MNTB neurons is to quickly and reliably switch excitatory inputs from VCN to inhibitory inputs for other nuclei, especially for LSO and MSO.

The LSO specializes in interaural level difference processing. Neurons of the LSO have excitatory inputs from ipsilateral VCN and inhibitory inputs from contralateral VCN through ipsilateral MNTB. LSO neurons perform subtraction of contralateral inputs from ipsilateral inputs.

MSO neurons are sensitive to time differences. They have excitatory and inhibitory inputs from both ipsilateral and contralateral VCN (inhibitory inputs are relayed through MNTB). MSO neurons perform coincidence detection, or multiplication, of contralateral and ipsilateral inputs.

There are two prominent theories how azimuth can be coded in the first binaural neurons. The first theory emerged with Jeffress's seminal work in which the place code was proposed (Jeffress, 1948). In the place code theory, individual neurons represent different azimuths and the array of such neurons creates a map of azimuthal space. It is assumed that this type of coding is utilized in avian auditory systems (Carr and Konishi, 1988). The rate code theory suggests that each neuron codes the azimuth with the actual spike rate. Averaging across the population of neurons or averaging the spike rate of one neuron in time yields the desired precision of azimuth estimation. This theory suits better to the mammalian auditory system in which inhibition plays an important role (Grothe, 2003).

Principal neurons of the MSO can be modeled on many levels, from highly abstract models to detailed physiological models. Studies that do not focus on modeling of MSO and only need ITD extraction often use cross-correlation (Faller and Merimaa, 2004) or complex-valued gammatone filters (Dietz et al., 2011) to analytically derive ITD. Works that pay attention mainly to spike trains tend to use black-box models where MSO neurons are represented by coincidence window (Marsalek and Kofranek, 2004; Franken et al., 2014). Studies that focus on MSO properties

use Hodgkin-Huxley models or formal neuron models with fixed thresholds (Colburn et al., 1990; Brand et al., 2002; Zhou et al., 2005) or models in which the probability of spiking correlates with the maximal membrane potential achieved by the summation of postsynaptic potentials (Batra et al., 1997; Leibold, 2010).

3. Goals

In our study of information processing in the auditory system, we focus on the sound localization circuit as it shows remarkable computational properties. Based on previous work in the field, we formulated several questions grouped into three research projects:

Analytical description of coincidence detection. In the black box model developed by our research group (Marsalek and Kofranek, 2004; Marsalek and Lansky, 2005), MSO neuron acts as a coincidence detector that reports coincidence if and only if excitatory inputs from both sides are closer in time than a constant Δ, or, excitatory input follows inhibitory input in a time shorter than Δ. This constant forms a time window called a coincidence window. However, the nature of a coincidence window is more complex, as was shown by recent physiological studies that explored the role of inhibition in mammalian MSO (McAlpine et al., 2001; Grothe, 2003; Pecka et al., 2008) and its importance in tuning of coincidence detection (Myoga et al., 2014; Franken et al., 2015). The research questions were:

- What is the relationship between biophysical neuronal parameters and the size and position of the coincidence window?
- How can inhibitory input influence parameters of the coincidence window?

Ergodicity and statistical properties in auditory circuits. One of the characteristic traits of low-frequency auditory nerve fibers is the periodicity of spike trains: neuronal discharges are phase-locked, i.e. the probability of spike is a function of the sound phase. As we measure spike timing on an angular scale, it is beneficial to use circular statistics to describe the stochastic processes in the auditory pathway. Another characteristic trait of the auditory nerve is high redundancy, as a large number of neurons carry information about the same sound feature. It was shown that a large number of auditory neurons converging in a subsequent nucleus resulted in increased precision (Joris et al., 1994; Marsalek et al., 1997). The precision in the auditory pathway also increases with a longer duration of stimulus. We borrowed the concept of ergodicity from statistical physics. It states that an average taken over a smaller set of units and longer time should equal a larger set of units and

shorter period of time. With the help of circular statistics and ergodicity concept, we ask:

- Can we calculate a vector strength of neural spike timing and spike train variability?
- How is the output spiking density of MSO neuron dependent on interaural time delay?

Sound localization in complex acoustic scenes. A psychoacoustical study showed that human localization is quite precise even in the presence of five concurrent talkers and reverberation (Kopco et al., 2010). In such complex environments, multiple sound waves reaching ears simultaneously results in unreliable binaural cues. One theory states that parts of the signal that are dominated by only one sound source have high interaural coherence. The auditory system thus can pick binaural cues from parts of the signal with high interaural coherence for precise localization of multiple sound sources (Faller and Merimaa, 2004; Dietz et al., 2011). Although a recent study claimed good results with ITD model that utilized interaural coherence in complex scenes (Josupeit et al., 2016), it did not explore the contribution of interaural coherence per se. A clarification of the following is needed:

- Could models that select binaural cues based on interaural coherence explain the level of precision of human sound localization in complex acoustic scenes with several concurrent talkers and reverberation?
- Could interaural level differences be more reliable than interaural time differences in complex acoustic scenes?

4. Methods and models

4.1 Analytical description of coincidence detection

To describe the relationship between biophysical neuronal parameters and the size and position of the coincidence window, we used a simplified version of the formal biological neuron model called the Spike response model (SRM_0), as defined by Gerstner and Kistler (2002). This model does not use differential equations for the description of the course of the neuronal membrane potential but is formulated using linear summations of the membrane potential responses on synaptic inputs. Postsynaptic potential (PSP) was modeled as a subtraction of two exponentials (Eq. 4.4). We created two models, ECD having one excitatory input from each side and ICD with one additional contralateral inhibitory input.

In SRM_0, the dependence of the response function on previous action potentials is removed. The time course of voltage u in time t on the membrane of a neuron i is defined by the equation:

$$u_i(t) = U_\mathrm{r} + \sum_j s_{ij} \sum_{t_j} \epsilon_0(t - t_j), \tag{4.1}$$

where j is an index of the presynaptic neurons, s_{ij} is the efficiency of synapse, $\epsilon_0(t)$ is the response function to a presynaptic AP (defined below), t_j are the spike times in presynaptic neurons, and U_r is the resting potential. The action potential is generated if the membrane potential $u_i(t)$ crosses the threshold ϑ

$$\max(u_i(t)) \geq \vartheta. \tag{4.2}$$

A natural response function following presynaptic action potential is the PSP caused by neurotransmitter-induced changes in postsynaptic membrane conductance (PSG). We model the course of the PSG g with exponential decay function using

the time constant τ. This exponential decay is expressed as

$$g(t) = G \exp\left(-\frac{t}{\tau}\right), t \geq 0, \tag{4.3}$$

where G is the maximum value of the postsynaptic conductivity. Substituting to the equations for postsynaptic currents and voltages adapted from the leaky integrate-and-fire model (Gerstner and Kistler, 2002; Tuckwell, 1988) leads to the normalized response function $\epsilon_0(t)$

$$\epsilon_0(t) = \frac{q}{C} \cdot \frac{1}{1 - \tau/\tau_{\mathrm{m}}} \left[\exp\left(\frac{-t}{\tau_{\mathrm{m}}}\right) - \exp\left(\frac{-t}{\tau}\right)\right] \frac{\mathrm{H}(t)}{U_{\mathrm{M}}}, \tag{4.4}$$

where q is the total charge passed through the membrane, τ_{m} is the membrane time constant and $\tau_{\mathrm{m}} = RC$. R and C are the resistance and trans-membrane capacitance, respectively. $\mathrm{H}(t)$ is the Heaviside step function: $\mathrm{H}(t) = 0$ for $t < 0$, otherwise 1. U_{M} is a normalization factor.

The model is illustrated in Fig. 4.1.

The coincidence window is defined as a range of ITDs that elicit coincidences. Within the duration of one sound period, the coincidence window is determined by the smallest and the largest ITD that evokes coincidence.

The analytical description of the relationship between neuron parameters and coincidence window is constructed through analyzing the local extremes of summed PSPs. We obtain the function m returning maximum total potential for the given ITD and neuron parameters. The ITD for which this reaches the threshold marks either the beginning or the end of the coincidence window. The inverse functions w to the monotonic parts of the function m thus form a relation that returns ITDs that mark the beginning and the end of the coincidence window for the given AP threshold and other neuron parameters.

The analytical description is constructed under the following assumptions: *First*, two equally large excitatory PSPs meet in the coincidence detection window. *Second*, one excitatory PSP is not enough to elicit an AP and two excitatory PSPs should be able to elicit an AP. *Third*, the inhibitory PSP precedes the contralateral excitatory PSP and the inhibitory PSP amplitude is not greater than the excitatory PSP amplitude. The *last* assumption needed for analytical description of the coincidence

detection is that τ_{m} is twice as fast as τ.

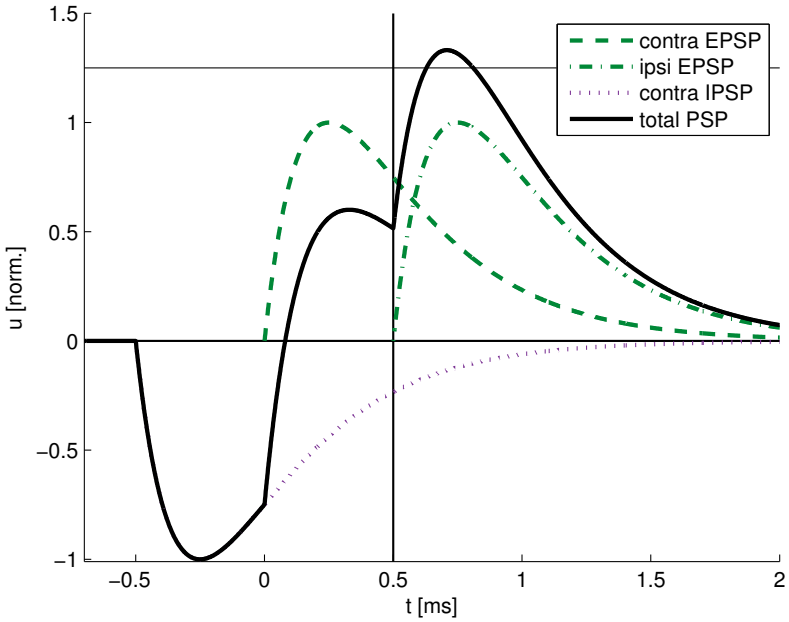

Figure 4.1: Individual PSPs with $\tau_{\mathrm{m}} = 0.180$ ms and $\tau = 0.360$ ms, are shown by green dashed, dash-dot and violet dotted lines. Inhibitory PSP precedes the first excitatory PSP by 0.5 ms. The second excitatory PSP follows the first one after ITD = 0.5 ms. The total potential (sum of the individual PSPs) is shown by solid line. In this example, the total potential crosses the threshold ϑ, shown by thin solid line, and AP is generated.

A straightforward numerical approach is used to explore the coincidence detection outside the restricted range of parameters posed by these assumptions and inherent limitations of the SRM_0. The coincidence window size is determined by evaluating the model equation across the range of parameters and ITDs and subsequently comparing this to the AP threshold. The numerical method allows differently parameterized response functions ϵ_0 to be used for each input.

4.2 Ergodicity and statistical properties in auditory circuits

Definitions, distributions and models used to derive statistical properties of spike trains are described in this section. Numerical computations with the descriptive circular statistics were realized with the help of a library package in MATLAB by

Berens (2009).

The periodicity of neural spikes are measured by vector strength defined for vector $\varphi_1, \ldots, \varphi_N$ as (Gumbel et al., 1953; Goldberg and Brown, 1969):

$$r(\varphi) = \frac{1}{N}\sqrt{\left(\sum_{i=1}^{N}\cos\varphi_i\right)^2 + \left(\sum_{i=1}^{N}\sin\varphi_i\right)^2}. \tag{4.5}$$

For continuous phase space φ that attains all values from an interval of $[0, 2\pi)$ with the probability density function $g(\varphi)$, we use the definition of continuous vector strength (van Hemmen, 2013):

$$r(\varphi) = \sqrt{\left(\int_0^{2\pi} g(\varphi)\cos\varphi \; d\varphi\right)^2 + \left(\int_0^{2\pi} g(\varphi)\sin\varphi \; d\varphi\right)^2}. \tag{4.6}$$

To calculate vector strength of spike train with the beta distribution, we introduce a three parameter version of the standard beta distribution (Mardia, 1972). This way we can use appropriate forms of the distribution regardless, whether the support is on the interval $[0, 1)$, or on the interval $[0, 2\pi)$. This can be decided by setting values of scale coefficient $s = (2\pi)^{-1}$ or $s = 1$. Continuous values of $s \in (\delta, 1]$ with small δ describe continuously parameterized phase locking input spikes. The beta distribution has then a probability density function written as:

$$f_B(t, a, b, s) = sB(a, b)^{-1}(t/s)^{a-1}(1 - t/s)^{b-1}\mathrm{H}(t/s)\mathrm{H}(1 - t/s), \tag{4.7}$$

with variable t and parameters $a, b \geq 1, s > 0$. The formula of the distribution is normalized by Euler beta function $B(a, b)$ to give unity integral (Cipra, 1994). Its range is cut off by Heaviside function $\mathrm{H}(t)$, $\mathrm{H}(t) = 1$ for $t > 0$, otherwise $\mathrm{H}(t) = 0$.

We calculate the output spike density of MSO neuron dependent on the time delay T between the right and left side as a difference of the two random variables, $T + Y$ and X. By the statistical calculus, the probability density function h of random variable $Z = (T + Y) - X$ is given by convolution

$$h(t) = g(t - T, a - T, b - T, u) * g(-t, a, b, u) \tag{4.8}$$

of probability densities of random variables $(T + Y)$ and X. We can use the op-

eration of convolution because under the ergodic assumption and the assumption that the resulting spikes form a renewal process, all the interspike intervals have the same probability distribution, and all are mutually independent. For the probability density of the right and left input X and Y we use a three-parameter probability distribution of the phase φ:

$$f(\varphi,\alpha,\beta,u) = \frac{1}{\beta-\alpha}\left[1-u\cos\left(2\pi\frac{\varphi-\alpha}{\beta-\alpha}\right)\right]\mathrm{H}(\varphi-\alpha)\mathrm{H}(\beta-\varphi). \tag{4.9}$$

It is a mixture of uniform and sine distribution and the parameter u is a weight of the sinusoidal part. Parameters α and β are bounds of the support of the distribution that satisfy $0 \leq \alpha \leq \beta \leq 2\pi$. The weight parameter, u, is limited to interval $[0,1]$. When $u = 0$, $f(\varphi)$ becomes the probability density of the uniform distribution on interval $[\alpha,\beta]$. Pure sinusoidal distribution is achieved for $u = 1$.

4.3 Sound localization in complex acoustic scenes

Our model of sound localization in complex acoustic scenes consisted of four stages: virtual auditory scene, auditory periphery, binaural processing and target detection. We implemented two variants of binaural processing: IC-based model and target-search-based model. Modeling work was realized in MATLAB with the AM toolbox.

Virtual auditory scene. Spatial configurations of speakers and room acoustics were modeled by convolving speech samples with binaural room impulse responses (BRIR) for desired positions. The speech samples were drawn from the same speech corpus of monosyllabic words recorded at Boston University's Hearing Research Center (Kidd et al., 2008) that was used in the experimental study (Kopco et al., 2010). The levels of samples were adjusted for the desired target-to-masker ratio. We used BRIR measured in a slightly reverberant room on a KEMAR manikin (Shinn-Cunningham et al., 2005) similar to the room in which the experimental study took place. To get rid of slight asymmetries in the BRIR we replaced recordings for positive azimuths with recordings of respective negative azimuth with switched left and right channels. An additional set of pseudo-anechoic BRIR was created by cutting off BRIR at 5.5 ms with 0.25 ms long cosine-square ramp.

Auditory periphery was modeled in a standard way (Faller and Merimaa, 2004).

The frequency analysis of the basilar membrane was modeled by a gammatone filterbank (Patterson et al., 1995) splitting left and right ear signals into twenty-nine frequency bands ranging from 200 Hz to 10 000 Hz. A model of basilar membrane compression and neural transduction (Bernstein et al., 1999) was applied to each frequency band: a Hilbert envelope was compressed by raising to a power of 0.23, then half-wave, square-law rectified, and filtered with a fourth-order low-pass filter with a cutoff frequency of 425 Hz. No additional internal noise was added.

IC-based localization. The binaural processor that computes ITD, ILD and IC was implemented as in (Faller and Merimaa, 2004). For every time-frequency bin, a normalized interaural cross-correlation was computed in a range of time shifts on defined exponential-decaying time window with time constant of 10 ms. The maximum of interaural cross-correlation was assigned as the IC value, and the lag of the maximum was assigned as the ITD. ILD was computed as a ratio of power of the left signal to the power of the right signal, over exponential-decaying time window lagged according to ITD. Only ITDs and ILDs of bins with IC higher than the threshold are selected for azimuth estimation. To combine localization information across critical bands, we first translated cues into azimuths and constructed PDF of individual azimuth estimations. A median of PDF is taken as a final azimuth estimation. Two azimuth estimations were computed - one from ITD cues of frequency bands below 1400 Hz and one from ILD cues of frequency bands above 1400 Hz.

Target-search-based localization. This model searches for a target in the temporal domain of both the left and right signals and estimates target sound source location from ITD as a difference of target timing between the left and the right signal. Search for a target was realized by computing cross-correlation function of a signal and a target template. A dot product on the window of target template size is computed for every possible shift in a temporal domain. Local maxima of resulting function were considered as candidates of the target position on time axis. An internal representation of a vowel from the target word "two" was used as a target template: a vowel was convolved with pseudo-anechoic BRIR at zero azimuth and preprocessed by auditory periphery.

Target detection. To distinguish between target and masker cues at IC-based localization, we computed a binary target mask that marked time-frequency bins in

which the target is expected to have some minimal energy. The mask was created from target-only scene. We utilized two types of masks - an ideal target mask, computed from true target position, and a generalized target mask, computed from the target at the position of 0° azimuth. Target-search-based localization had inherent target detection thus we did not use binary target mask in that model.

Simulations. Four basic types of model were tested: 1. low-frequency ITD only IC-based model with an ideal target mask, 2. high-frequency ILD only IC-based model with an ideal target mask, 3. high-frequency ILD only IC-based model with a generalized target mask, 4. target-search based model. The first three models were tested for a range of IC threshold settings. Each simulation consisted of multiple trials. In each trial, speech samples were presented from a subset of eleven evenly spaced loudspeakers ($-50°$ to $50°$ with $10°$ separation). Target sound could be presented from any of the loudspeakers, and maskers were arranged in one of five possible patterns. Maskers were presented either at the same level as the target or 5 dB louder, yielding target-to-masker ratio (TMR) of 0 dB, or, respectively, TMR of -5 dB. The target word was word "two" uttered by one of the female voices. Masker words were four different randomly drawn nondigit words spoken by four different male voices. Fifty different combinations of masker words were generated ensuring that individual variability in masker words combinations would not bias overall behavior.

5. Results

5.1 Analytical description of coincidence detection

The main function of the MSO neuron is coincidence detection. The range of ITDs that elicit a response of the neuron depends on its biophysical parameters. The analytical relation that determines the two boundaries of the coincidence window as a function of the relative threshold of a neuron with given parameters is expressed as two branches w_+ and w_- of the curve w,

$$w_{+,-}(\vartheta') = 2\tau_{\mathrm{m}} \log\left(\frac{a_2 \pm \sqrt{\vartheta'[a_2^2 - (\vartheta'-1)a_1]}}{\vartheta'-1}\right), \tag{5.1}$$

valid in the interval $\vartheta' \in (1, 1+a_2^2/a_1]$, where

$$a_1 = 1 + s_{\mathrm{I}} \exp\left(\frac{T_{\mathrm{INH}}}{\tau_{\mathrm{m}}}\right), \tag{5.2}$$

$$a_2 = 1 + s_{\mathrm{I}} \exp\left(\frac{T_{\mathrm{INH}}}{\tau}\right). \tag{5.3}$$

ϑ' is the threshold for generating the action potential relative to the size of normalized excitatory PSP, τ_{m} is the membrane time constant (Eq. 4.4) and τ is the time constant of postsynaptic membrane conductance exponential decay function (Eq. 4.3). The expressions a_1 and a_2 reflect the effect of the inhibitory input: s_{I} is the synaptic efficiency of the inhibitory input relative to the excitatory input and T_{INH} is the temporal difference between the inhibitory and the same-side excitatory input. In the case of the coincidence detection without inhibition $s_{\mathrm{I}} = 0$ and $a_1 = a_2 = 1$.

The two solutions correspond to two roots of quadratic equation, which can be clearly seen in the argument of logarithm in the formula. The function (5.1) splits at the ITD at which the highest total potential is achieved, called the best ITD, that is

$$T_{\mathrm{MAX}} = 2\tau_{\mathrm{m}} \log\left(\frac{a_1}{a_2}\right). \tag{5.4}$$

The center of the coincidence window w_C is at

$$w_C(\vartheta') = w_-(\vartheta') + (w_-(\vartheta') + w_+(\vartheta'))/2. \tag{5.5}$$

Two examples of the functions w_+ and w_- are shown in Fig. 5.1 below.

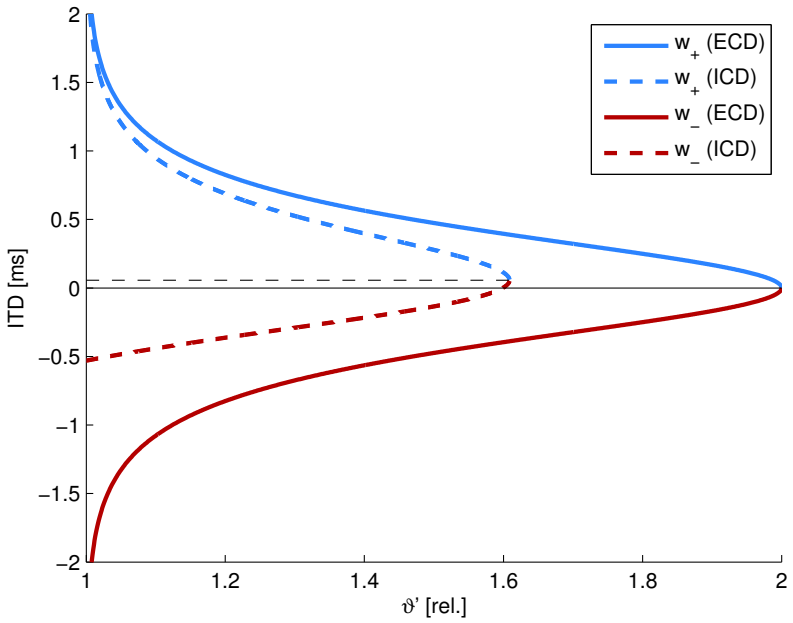

Figure 5.1: Branches of the curve w, function w_+, shown in blue, and w_-, shown in red, determine the boundaries of coincidence windows for the given relative threshold and neuron parameters. The curve w for ECD is shown by thick solid lines and the curve w for ICD is shown by thick dashed lines. The points where the corresponding branches meet are marked with thin lines. The parameters are τ_m = 0.180 ms, d = 2, T_{INH} = -0.2 ms, and s_I = -0.5.

5.2 Ergodicity and statistical properties in auditory circuits

Spike trains phase locked to sound phase are described by circular statistics. We present vector strengths of spike trains with various underlying probability densities.

Vector strength r of a spike train with the spike phases relative to sound phase distributed according to the beta distribution (4.7) with the lowest natural number non-trivial parameters $a = b = 2$ and the scaling parameter s attaining values from

$0 \leq s < 2\pi$ can be expressed as

$$r = 12s^{-3}|s\cos(s/2) - 2\sin(s/2)|. \tag{5.6}$$

For any given natural a and b, r can be expressed with the use of hyper-geometric function, however, a formula for all arbitrary values of a and b does not exist.

Vector strength r of a spike train with the spike phases relative to sound phase distributed according to the compound uniform and sine distribution (4.9) is

$$r_f = \frac{8\pi^2 + 2\delta^2(u-1)}{\delta(4\pi^2 - \delta^2)} \sin\frac{\delta}{2}, \tag{5.7}$$

dependent on the length of the support, δ, and the weight, u. In the temporal representation of the delay, expressed in terms of $t = \varphi T/2\pi$, where T denotes the period of the sound, and the length of the support is denoted analogously by $d = b - a$, it is

$$r_g = \frac{T^2 + d^2(u-1)}{\pi d(T^2 - d^2)} T \sin\frac{\pi d}{T}. \tag{5.8}$$

On the output of MSO neuron, the probability density function of the interspike intervals (ISIs) dependent on the time delay T between the right and left side is equal to

$$h(t) = \frac{d - |t-T|}{d^2} + \frac{u^2(d - |t-T|)}{2d^2}\cos\frac{2\pi|t-T|}{d} + \frac{u(4-u)}{4\pi d}\sin\frac{2\pi|t-T|}{d} \tag{5.9}$$

for $|t - T| \leq d$. This probability distribution has support $[T-d, T+d]$, and depends on the parameter d and the weight, u. The resulting probability density has zero skewness, its mean value is equal to

$$\mu_h = T \tag{5.10}$$

and variance equal to

$$\sigma_h^2 = \frac{d^2}{6\pi^2}(\pi^2 - 6u). \tag{5.11}$$

The coefficient of variation is often used for the description of the firing characteristics of a spike train. The derived distribution of ISI has coefficient of variation equal

to

$$C_{\mathrm{V}h} = \frac{d}{\pi T}\sqrt{\frac{\pi^2 - 6u}{6}}. \tag{5.12}$$

Expressing the weight, u, from (5.12) and substituting it into (5.8), we obtain the vector strength of the output spike train as a function of the coefficient of variation of the ISIs of the input spike train,

$$r_h = \frac{T}{6\pi d(T^2 - d^2)}\left[\pi^2(d^2 - 6T^2C_{\mathrm{V}h}^2) + 6(T^2 - d^2)\right]\sin\frac{\pi d}{T}, \tag{5.13}$$

with the parameter d controlling the length of both the supports of the delay and ISIs.

5.3 Sound localization in complex acoustic scenes

The aim of simulations of the complex acoustic environments was to find computational descriptors, which govern the subject's decisions of the target voice location in the complex auditory stimulus containing voice maskers.

Results are described individually for each of the simulations. For overall comparison we used root mean square error (RMSE) and a percentage of failed trials as the failure rate across the trials in one of the three runs - control run (target only), masker runs with TMR 0, and masker runs with TMR -5. For masker runs, RMSE were computed both relative to the true target position and relative to the model estimation in control run. For more detailed evaluation that includes RMSE, mean locations, and failure rate computed separately for each target-maskers spatial configuration, see appendix.

Simulation 1: IC-based, low-frequency ITD model with ideal target mask. RMSE and failure rate for IC-based, low-frequency ITD model with ideal target mask are shown on fig. 5.2 on the top row. For control run, although RMSE for low IC thresholds are higher than RMSE without cue selection ($\mathrm{RMSE} = 8.7^\circ$ for $\mathrm{IC}_0 = 0$ vs. $\mathrm{RMSE} = 14.5^\circ$ for $\mathrm{IC}_0 = 0.5$), increasing IC threshold leads to low RMSE in the range of $2.6 - 3.1$ for $\mathrm{IC}_0 \geq 0.985$. Failure rate is always zero, meaning that for each IC threshold there was at least one sample selected. Both masker conditions show similar trends: RMSE are relatively stable in a range of $25^\circ - 32^\circ$ for TMR 0

and $29° - 34°$ for TMR -5 up to $\mathrm{IC}_0 = 0.985$. Above that threshold we can observe slight decrease of RMSE, down to 21° for TMR 0 and 31° for TMR -5 connected with sharp increase of failure rate to 85% for both TMRs.

Simulation 2: IC-based, high-frequency ILD model with ideal target mask. RMSE is decreasing monotonically with higher IC thresholds for both control and masker conditions (fig. 5.2, middle row). For control condition, even low IC thresholds bring significant reduction in RMSE ($\mathrm{RMSE} = 23.8°$ for $\mathrm{IC}_0 = 0$ vs. $\mathrm{RMSE} = 8.7°$ for $\mathrm{IC}_0 = 0.5$). For highest $IC_0 = 0.995$ RMSE drops below 1° with failure rate around 18%. Masker conditions follow the same trend. Absolute RMSE are 13.8° (TMR 0), resp. 17.4° (TMR -5) for low IC_0, and decreasing to around 2° for highest $IC_0 = 0.995$. Above $IC_0 = 0.98$, the failure rate rises sharply. Masker conditions relative to control do not decrease up to $\mathrm{IC}_0 = 0.98$, above that value both decrease. Failure rate is zero for lower IC_0 and starts rising at around $\mathrm{IC}_0 = 0.98$ for simulations with maskers (both TMR 0 and TMR -5).

Simulation 3: IC-based, high-frequency ILD model with generalized target mask. Results of simulation 3 (fig. 5.2, bottom row) are similar to the results of simulation 2. Trends are the same, only values differ: Simulation 3 results show higher RMSE for control condition and slightly lower RMSE for masker conditions. Failure rate is higher compared to simulation 2.

Simulation 4: Target-search based model. Simulation 4 has overall RMSE of 2.6° and 0% failure rate for control condition, overall RMSE of 9.9° (9.1° re. control) and 1.5% failure rate for TMR 0 masker condition, overall RMSE of 17.2° (17.3° re. control) and 5.9% failure rate for TMR -5 masker condition.

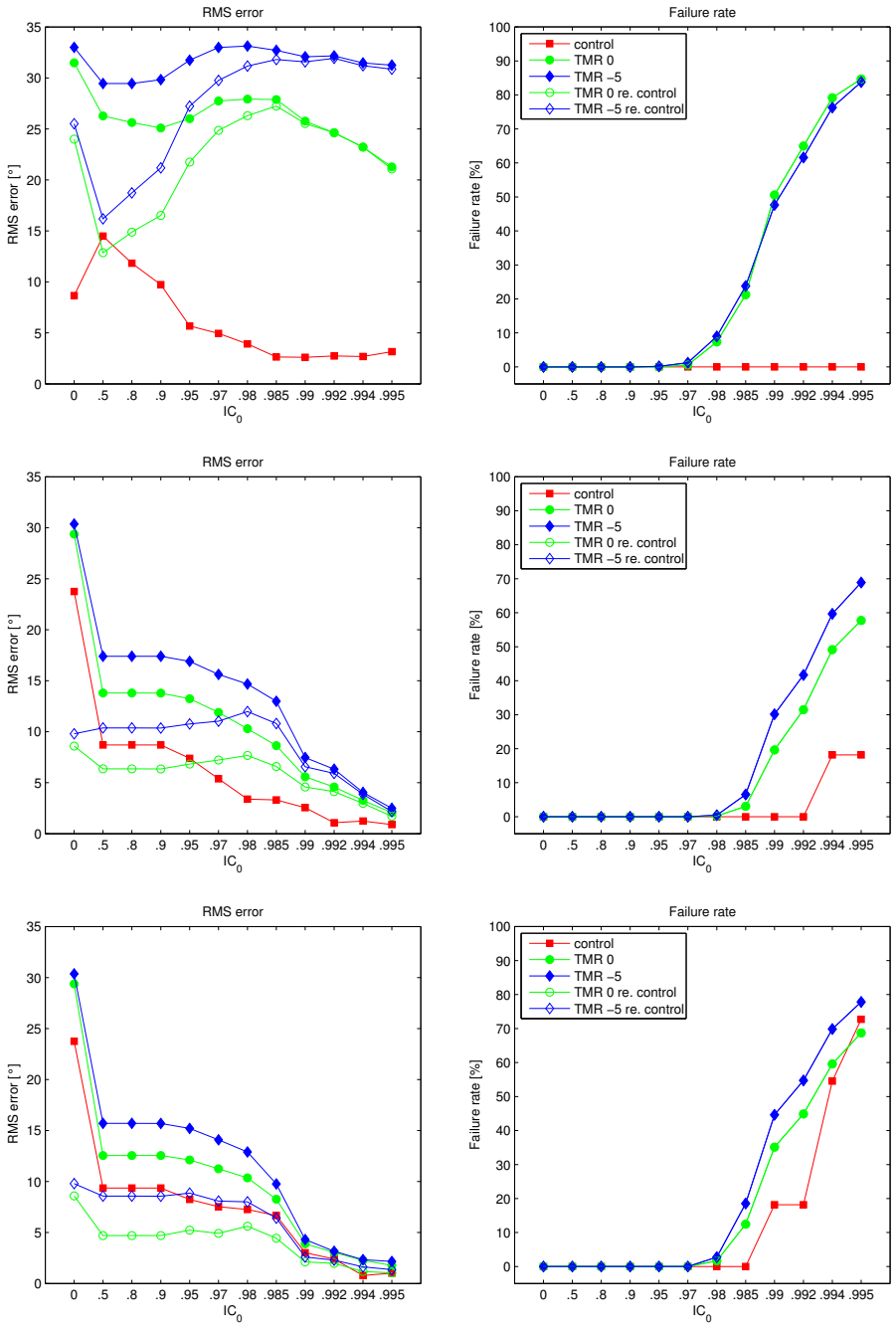

Figure 5.2: RMSE (left column) and failure rate (right column) of IC-based models as a function of IC threshold. Models showed are: low-frequency ITD model with ideal target mask (top row), high-frequency ILD model with ideal target mask (middle row), and high-frequency ILD model with generalized target mask (bottom row). Control runs are plotted with red squares, TMR 0 masker runs are plotted with green circles, and TMR -5 masker runs are plotted with blue diamonds. For masker runs, absolute errors are plotted with full marks and errors relative to control run are plotted with empty marks.

6. Discussion

6.1 Analytical description of coincidence detection

The Spike Response Model SRM$_0$ was chosen after the consideration of specific properties of MSO neuron and its function as a coincidence detector. We are not modeling the exact time or course of the AP since the question is whether or not the AP was generated. Synaptic interactions are examined within one sound period and we assume no interaction from previous sound cycles as inputs are phase-locked and sparse (Joris et al., 1994). Linear summation of inputs is in correspondence with electrophysiological measurements of MSO neuron (Van der Heijden et al., 2013; Roberts et al., 2013). However, spike generation is nonlinear (Van der Heijden et al., 2013) and that may affect the computation of the output spike rates based on the coincidence window in the similar way as the input jitter does (Marsalek and Lansky, 2005).

Postsynaptic potentials. The response function with parameters corresponding to that measured experimentally: $\tau_{\mathrm{m}} = 0.180$ ms, $\tau = 0.360$ ms, (Myoga et al., 2014), yields the half-width of 0.635 ms and rise time of 0.097 ms, which is on a par with experimentally measured data (Myoga et al., 2014). For slower $\tau = 1.440$ ms, as is the case for inhibitory synapses, the correct half-width of 1.531 ms is observed, but there is a faster rise time of 0.155 ms, while the desired rise time is around 0.3 ms. The way we model postsynaptic conductance yields very low values of the onset latencies, 0.019 ms and 0.029 ms versus the desired value of 0.1 ms. We argue that this small difference does not affect the model significantly as the relative timing of the response function peaks has been preserved. Analytical relation has a constraint $\tau = 2\tau_{\mathrm{m}}$. For excitatory postsynaptic potential (EPSP), this is biologically plausible, as mentioned above. For inhibitory postsynaptic potential (IPSP), this means an unrealistically fast τ for an inhibitory input. Although it is possible to find models using the same fast constants for excitatory and inhibitory inputs (Brand et al., 2002), we run numerical simulations with specific time constants for both excitatory and inhibitory inputs. The finding is that the ratio of time constants has

an influence on the coincidence window, but the main results discussed here have universal relevance regardless of the ratio of τ to τ_{m}.

Coincidence window duration. Experimental data suggest that the maximal duration of coincidence window is around 200 - 300 μs (Jercog et al., 2010; Roberts et al., 2013; Myoga et al., 2014). Computational studies also assume narrow coincidence windows (Sanda and Marsalek, 2012; Franken et al., 2014). In the presented model, with parameterization based on a recent electrophysiological experiment (Myoga et al., 2014), the window function w is relatively broad. Duration of the coincidence window around 200 - 300 μs is achieved with the relative threshold for the AP generation about 5% below the maximum of EPSP summation. This implies the necessity of precise setting of the AP generation threshold relative to the EPSP size. One way to adjust the duration of the coincidence window is to change the strengths of EPSPs or the threshold potential. However, if the developmental changes of the EPSP and AP generation kinetics have another aim, for example to attenuate AP backpropagation to soma or to achieve cycle-to-cycle independence of EPSP summation (Scott et al., 2005), there should be another mechanism capable of adjusting the duration of the coincidence window. Varying inhibition size shifts the window function w towards the lower values of ϑ', or, respectively, narrows the coincidence window for a fixed threshold value. We suppose that this mechanism does not need precisely delayed inhibition and could work even if IPSPs sum across cycles, as it happens with sounds of higher frequencies. Narrowing of coincidence window as an effect of inhibitory input is a general finding in many experimental studies (Brand et al., 2002; Pecka et al., 2008; Franken et al., 2015).

Coincidence window position. The window function for ECD is symmetrical and both the best ITD and the center of the coincidence window are at ITD = 0. In ICD, these two values show us the shifts and asymmetry in the coincidence detection induced by inhibitory input. In theory, lowering the thresholds could result in the center of the coincidence window shifted much further away from the midline than the best ITD for broad coincidence windows. In the parametrization we use, the difference between the coincidence window shift and the best ITD shift is negligible for the threshold within 5% range of its maximal values. The size and the nature of possible shift is in accordance with data from experimental studies (Brand et al.,

2002; Pecka et al., 2008). Specifically, we can observe very similar dependence of the best ITD shift on the relative inhibitory delay as in a recent study by Myoga et al. (2014). Shift of the best ITD depends on the exact delay of the inhibitory input relative to the excitatory input from the same side. Jitter in this relative timing may further reduce the size of this shift. We can hypothesize that this shift could be more reliably achieved with the delay of the excitatory input respective to the contralateral excitatory input.

Further considerations in relation to other studies. One of the aims of this study was to validate an asymmetric rule for coincidence detection in the probabilistic delay model. Using this rule, Sanda and Marsalek (2012) explored the duration of the coincidence window from 60 μs up to 2 ms and found that the best value is 600 μs. In this model, the shift of both coincidence window boundaries to positive values of ITDs is possible, but it is associated with a very narrow ICD coincidence window. Furthermore, in this case, the ECD coincidence window does not constitute a symmetrical complement to the ICD coincidence window. A comparison of coincidence windows with and without inhibition for the same relative threshold shows that inhibition narrows and shifts the coincidence window. However, the possible values of these effects are limited and some relations, for example, as in the asymmetric rule for coincidence detection, where inhibition halves and shifts the whole coincidence window to positive values, are excluded. In this model, we have determined the coincidence window for two coinciding excitatory inputs. There is evidence that multiple fibers innervate MSO cells (Couchman et al., 2010). A recent model based on coincidence window explored properties of complex coincidence detection (Franken et al., 2014) and confirmed that for low binaural thresholds, multiple fibers from each side are necessary, even though this number was lower for a broader coincidence window. Here, the decision to model only one input per side was based on the fact that in the probabilistic delay model (Marsalek and Lansky, 2005), multiple fibers can be considered in the input probability function. However, this approach could not account for monaural coincidences.

6.2 Ergodicity and statistical properties in auditory circuits

In our computations we sought a simple description of spike trains following a sound phase. The candidate functions we tried were circular normal and circular beta distribution functions. These are circular counterparts of normal and beta PDFs. The choice of the PDFs (normal, circular normal, beta, sine) in the calculations was made based on the correspondence between PDFs on the aperiodic versus on the periodic supports. A sine is circular and is parameter-free, and it is the simplest of all the circular PDFs. When we use compound density, a weighted sum of uniform and of sine density, we arrive at an arbitrary value of vector strength. We can invert monotonous dependencies of vector strength on standard deviation and other parameters of circular densities. By using this procedure, we can fit the ratio of uniform and sine component to the vector strength of experimental spike trains.

Vector strength is an important parameter (Goldberg and Brown, 1969). It can be found virtually at all levels of spike train description, which takes into account spike timing relative to the stimulus phase of another spike train. Marsalek (2001) discusses vector strength in the MSO nucleus at low frequencies. Vector strength can be described at the moment of mechanical-to-electrical transduction in the cochlea (Camalet et al., 2000) and then at many places upwards in the auditory pathway and beyond.

Convolution is an important operator in signal processing. For the convolution calculation in the auditory pathway and comparison of beta and normal densities in this context, see Drapal and Marsalek (2010, 2011). The convolution of two spike time densities is used here for the ISI calculation. Another interpretation of the calculation and the parameters is possible - the output density also represents the probability of generating output spike in a neural circuit of the MSO as a function of the time delay imposed into the circuit by the spike arrival difference from the neurons of the peripheral pathways from the left and right ears. The ensemble code produced this way signals the azimuth of the sound source location.

Cochlear implants are successful in replicating a series of action potentials in the auditory nerve by imitation of the mechanical to electrical transduction in the

cochlea by contemporary electronics (Drapal and Marsalek, 2010). Constructing cochlear implants involves reverse biomimetic and neuromimetic engineering. The biomimetic and neuromimetic approaches are an engineering method to construct industrial sensors by mimicking nature's solution to the problem. Both reverse and neuromimetic engineering are very dependent on computational modeling, on the right choice of models and on their complexity. The knowledge that neural pathways calculate a quantity does not necessarily tell us how it is computed. Therefore, phenomenological models of the auditory brainstem computation are useful as a first approximation, which can be refined in subsequent research (Marsalek and Kofranek, 2005).

6.3 Sound localization in complex acoustic scenes

Simulations revealed that the IC-based model failed in our complex localization task using low frequency ITD cues. Errors remained high for every IC threshold setting in masker runs. A slight decrease in RMSE for IC thresholds above 0.985 could be attributed to increased model failure in lateral target positions that were the most erroneous ones. The IC-based model using low frequency ITD cues performed well in control runs, however, control runs are a much easier localization task with only one sound source in a reverberant room. This is in accordance with original papers presenting models, showing there was a possibility to localize two sound sources in a reverberant environment with some error (Faller and Merimaa, 2004) and concluding that localization of more sound sources in a reverberant room was not possible (Dietz et al., 2011). Josupeit et al. (2016) used an IC-based model using low frequency ITD cues for the same localization task and reported good results for TMR 0 condition. However, they combined the IC-based model with target masks that ensured certain minimal instantaneous target-to-masker ratio, meaning that target masks alone selected time-frequency bins having the target as a dominant sound source. In our study we explored the ability of the IC-model to select bins having one dominant sound source from a set of all target bins independent of instantaneous target-to-masker ratio.

IC-based model using high frequency ILD cues showed surprisingly good performance, comparable to that of humans. The precision of localization increased monotonically with IC threshold for control run and masker runs. RMSE of masker runs relative to control run shows the contribution of maskers presence to degradation of localization performance. ILD localization has very low values of RMSE re. control for any IC threshold setting compared to ITD localization, suggesting that the level of masking is much lower at high frequencies than at low frequencies. Moreover, RMSE re. control is relatively constant up to $IC_0 = 0.98$ suggesting that for $IC_0 \leq 0.98$ all the improvement in localization is based on target echo suppression. For $IC_0 > 0.98$, model was able to filter out bins with maskers energy, however, failure rates sharply increased mainly for target at lateral positions or TMR -5. Naturally, for more lateral positions left and right BRIR differ more, resulting in lower interaural coherence and higher probability that there will be zero bins with high enough interaural coherence to select.

7. Conclusion

We have presented three modeling studies of binaural hearing.

The first study describes a relationship between biophysical parameters of the MSO neuron and its ability to detect coincidental spikes from the left and the right ear. The most important of these parameters are membrane time constants, conductance decay constants, relative action potential threshold, and relative synaptic strengths. Analytical relation enables better understanding of possible parametrizations of coincidence windows for neurons with known biophysical properties. Comparison of coincidence windows of neurons with excitatory and inhibitory inputs and of neurons with excitatory inputs only shows the major impact of inhibition on the duration and position of the coincidence window. The presence of inhibitory inputs results in a shift of the coincidence window outside the axis of symmetry of ITD and in shortening its duration. Precise value of the threshold potential relative to the EPSP size is essential for a narrow coincidence window. This is achieved also by adjusting the strength of the inhibitory input.

The second study presents models of output spiking densities of the MSO neuron and provides formulas for vector strengths for various density functions of spike trains.

In the third study we model localization of sound sources in complex listening situations with multiple sound sources in a reverberant environment. We conclude that: highly correlated parts of the signal, if available, provide reliable ILD estimates sufficient for precise target localization comparable to that of human subjects; low frequency ITD cues of highly correlated parts of the signal alone are not sufficient to explain human performance; and localization based on ITD is possible in combination with accurate target template.

8. Souhrn

Představili jsme tři modelovací studie binaurálního slyšení. První studie popisuje vztah mezi biofyzikálními parametry neuronu MSO a jeho schopností detekovat simultánní signály z levého a pravého ucha. Nejdůležitější z těchto parametrů jsou membránové časové konstanty, konstanty poklesu vodivosti, prahová hodnota relativního akčního potenciálu a relativní synaptické síly. Analytický vztah umožňuje lepší pochopení možných parametrizací koincidenčních oken pro neurony se známými biofyzikálními vlastnostmi. Srovnání koincidenčních oken neuronů s excitačními a inhibičními vstupy a neurony s pouze excitačními vstupy ukazuje zásadní vliv inhibice na trvání a polohu koincidenčního okna. Přítomnost inhibičních vstupů má za následek posun koincidenčního okna mimo osu symetrie ITD a zkrácení doby jeho trvání. Přesná hodnota prahu vzhledem k velikosti EPSP je nezbytná pro úzké koincidenční okno. Toho se dosáhne také úpravou síly inhibičního vstupu.

Druhá studie představuje modely pravděpodobnostní hustoty výstupního neuronu MSO a poskytuje vztahy pro vektorovou sílu pro různé pravděpodobnostní rozdělení vzruchů.

Ve třetí studii jsme modelovali lokalizaci zdrojů zvuku ve složitých poslechových situacích s více zdroji zvuku v dozvukovém prostředí. Došli jsme k závěru, že: vysoce korelované části signálu, pokud jsou k dispozici, poskytují spolehlivé odhady ILD dostatečné pro přesnou lokalizaci srovnatelnou s lidskými subjekty; nízkofrekvenční ITD klíče vysoce korelovaných částí signálu samy o sobě nestačí k vysvětlení lidského výkonu; a lokalizace založená na ITD je možná v kombinaci s přesnou cílovou šablonou.

9. List of publications

List of publications related to thesis:

Toth, P., Marsalek, P. (2015). Analytical description of coincidence detection synaptic mechanisms in the auditory pathway. *Biosystems*, 136: 90-98.

Toth, P., Marsalek, P., Pokora, O. (2018). Ergodicity and parameter estimates in auditory neural circuits. *Biological Cybernetics*, 112(1-2): 41-55.

Toth, P., Kopco, N., Marsalek, P. (2020). Modeling speech localization in multitalker environment. In preparation for submission to *The Journal of the Acoustical Society of America*, 9 pages.

Bibliography

Ashida, G., Funabiki, K., and Kretzberg, J. (2015). Minimal conductance-based model of auditory coincidence detector neurons. *PLoS One*, 10(4).

Batra, R., Kuwada, S., and Fitzpatrick, D. C. (1997). Sensitivity to interaural temporal disparities of low-and high-frequency neurons in the superior olivary complex. I. Heterogeneity of responses. *Journal of Neurophysiology*, 78(3):1222–1236.

Berens, P. (2009). CircStat: A MATLAB toolbox for circular statistics. *Journal of Statistical Software*, 31(10):1–21.

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

Braasch, J. (2005). Modelling of binaural hearing. In Blauert, J., editor, *Communication Acoustics*, pages 75–108. Springer, Berlin, Heidelberg.

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.

Bregman, A. (1990). *Auditory Scene Analysis*. MIT Press.

Camalet, S., Duke, T., Julicher, F., and Prost, J. (2000). Auditory sensitivity provided by self-tuned critical oscillations of hair cells. *Proceedings of the National Academy of Sciences*, 97(7):3183–3188.

Carlile, S. (1996). *Virtual auditory space: Generation and applications*. RG Landes Austin TX, USA.

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

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, 25(5):975–979.

Cipra, T. (1994). Probability theory. In Rektorys, K., editor, *Survey of Applicable Mathematics*, pages 688–732. Springer Science and Business Media, New York, NY, USA. vol 2, chap 33, p 724.

Colburn, H. S. (1973). Theory of binaural interaction based on auditory-nerve data. i. general strategy and preliminary results on interaural discrimination. *The Journal of the Acoustical Society of America*, 54(6):1458–1470.

Colburn, H. S. and Durlach, N. I. (1978). Models of binaural interaction. *Handbook of perception*, 4:467–518.

Colburn, H. S., Han, Y., and Culotta, C. P. (1990). Coincidence model of MSO responses. *Hearing Research*, 49(1):335–346.

Couchman, K., Grothe, B., and Felmy, F. (2010). Medial superior olivary neurons receive surprisingly few excitatory and inhibitory inputs with balanced strength and short-term dynamics. *Journal of Neuroscience*, 30(50):17111–17121.

Dietz, M., Ewert, S. D., and Hohmann, V. (2011). Auditory model based direction estimation of concurrent speakers from binaural signals. *Speech Communication*, 53(5):592 – 605.

Drapal, M. and Marsalek, P. (2010). Stochastic model shows how cochlear implants process azimuth in real auditory space. *The Chinese Journal of Physiology*, 53(6):439–446.

Drapal, M. and Marsalek, P. (2011). Stochastic model explains role of excitation and inhibition in binaural sound localization in mammals. *Physiological Research*, 60(3):573–583.

Durlach, N. I. (1963). Equalization and cancellation theory of binaural masking-level differences. *The Journal of the Acoustical Society of America*, 35(8):1206–1218.

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(5):3075–3089.

Franken, T. P., Bremen, P., and Joris, P. X. (2014). Coincidence detection in the medial superior olive: mechanistic implications of an analysis of input spiking patterns. *Frontiers in Neural Circuits*, 8(42):1–21.

Franken, T. P., Roberts, M. T., Wei, L., Golding, N. L., and Joris, P. X. (2015). *In vivo* coincidence detection in mammalian sound localization generates phase delays. *Nature Neuroscience*, 18(3):444–452.

Gerstner, W. and Kistler, W. M. (2002). *Spiking neuron models. Single neurons, populations, plasticity*. Cambridge University Press, Philadelphia.

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

Grothe, B. (2003). New roles for synaptic inhibition in sound localization. *Nature Reviews Neuroscience*, 4(7):540–50.

Grothe, B., Pecka, M., and McAlpine, D. (2010). Mechanisms of sound localization in mammals. *Physiological Reviews*, 90(3):983–1012.

Gumbel, E., Greenwood, J. A., and Durand, D. (1953). The circular normal distribution: Theory and tables. *The Journal of the Acoustical Society of America*, 48(261):131–152.

Hancock, K. E. and Delgutte, B. (2004). A physiologically based model of interaural time difference discrimination. *Journal of Neuroscience*, 24(32):7110–7117.

Hartley, R. and Fry, T. C. (1921). The binaural location of pure tones. *Physical Review*, 18(6):431.

Jeffress, L. A. (1948). A place theory of sound localization. *Journal of Comparative and Physiological Psychology*, 41(1):35–39.

Jennings, T. and Colburn, H. (2010). Models of the superior olivary complex. In Meddis, R., Lopez-Poveda, E. A., Fay, R. R., and Popper, A. N., editors, *Computational Models of the Auditory System*, volume 35 of *Springer Handbook of Auditory Research*, pages 65–96. Springer US.

Jercog, P. E., Svirskis, G., Kotak, V. C., Sanes, D. H., and Rinzel, J. (2010). Asymmetric excitatory synaptic dynamics underlie interaural time difference processing in the auditory system. *PLoS Biology*, 8(6):e1000406, 1–9.

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

Josupeit, A., Kopco, N., and Hohmann, V. (2016). Modeling of speech localization in a multi-talker mixture using periodicity and energy-based auditory features. *The Journal of the Acoustical Society of America*, 139:2911–2923.

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., and Hudspeth, A. J. (2013). *Principles of Neural Science*. McGraw Hill Professional.

Kidd, G., Best, V., and Mason, C. (2008). Listening to every other word: Examining the strength of linkage variables in forming streams of speech. *The Journal of the Acoustical Society of America*, 124:3793–3802.

Kopco, N., Best, V., and Carlile, S. (2010). Speech localization in a multitalker mixture. *The Journal of the Acoustical Society of America*, 127(3):1450–1457.

Leibold, C. (2010). Influence of inhibitory synaptic kinetics on the interaural time difference sensitivity in a linear model of binaural coincidence detection. *The Journal of the Acoustical Society of America*, 127(2):931–942.

Mardia, K. V. (1972). *Statistics of Directional Data*. Academic Press, London, UK.

Marsalek, P. (2001). Neural code for sound localization at low frequencies. *Neurocomputing*, 38-40(1-4):1443–1452.

Marsalek, P. and Drapal, M. (2008). Mechanisms of coincidence detection in the auditory brainstem: Examples. *Mathematical Modeling of Biological Systems*, 2:245–253.

Marsalek, P., Koch, C., and Maunsell, J. (1997). On the relationship between synaptic input and spike output jitter in individual neurons. *Proceedings of the National Academy of Sciences*, 94:735–740.

Marsalek, P. and Kofranek, J. (2004). Sound localization at high frequencies and across the frequency range. *Neurocomputing*, 58(60):999–1006.

Marsalek, P. and Kofranek, J. (2005). Spike encoding mechanisms in the sound localization pathway. *Biosystems*, 79(1):191–198.

Marsalek, P. and Lansky, P. (2005). Proposed mechanisms for coincidence detection in the auditory brainstem. *Biological Cybernetics*, 92(6):445–451.

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

Mills, A. W. (1958). On the minimum audible angle. *The Journal of the Acoustical Society of America*, 30(4):237–246.

Myoga, M. H., Lehnert, S., Leibold, C., Felmy, F., and Grothe, B. (2014). Glycinergic inhibition tunes coincidence detection in the auditory brainstem. *Nature Communications*, 5(3790):1–13.

Patterson, R. D., Allerhand, M. H., and Giguère, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *The Journal of the Acoustical Society of America*, 98(4):1890–1894.

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. *Journal of Neuroscience*, 28(27):6914–6925.

Rayleigh, L. (1907). On our perception of sound direction. *Philosophical Magazine Series 6*, 13(74):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(5):923–935.

Sanda, P. and Marsalek, P. (2012). Stochastic interpolation model of the medial superior olive neural circuit. *Brain Research*, 1434(0):257—-265.

Schnupp, J., Nelken, I., and King, A. (2011). *Auditory neuroscience: Making sense of sound*. MIT Press.

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

Shinn-Cunningham, B. G., Kopco, N., and Martin, T. J. (2005). Localizing nearby sound sources in a classroom: Binaural room impulse responses. *The Journal of the Acoustical Society of America*, 117(5):3100–3115.

Syka, J., Voldrich, L., and Vrabec, F. (1981). *Fyziologie a patofyziologie zraku a sluchu*. Avicenum, Praha.

Tuckwell, H. C. (1988). *Introduction to theoretical neurobiology*. Cambridge University Press, New York, NY, USA.

Van der Heijden, M., Lorteije, J. A. M., Plauska, A., Roberts, M. T., Golding, N. L., and Borst, J. G. G. (2013). Directional hearing by linear summation of binaural inputs at the medial superior olive. *Neuron*, 78(5):936–948.

van Hemmen, J. L. (2013). Vector strength after Goldberg, Brown, and von Mises: Biological and mathematical perspectives. *Biological Cybernetics*, 107(4):385–396.

Voss, S. E., Nakajima, H. H., Huber, A. M., and Shera, C. A. (2013). Function and acoustics of the normal and diseased middle ear. In *The Middle Ear*, pages 67–91. Springer.

Woodworth, R. S. (1938). *Experimental Psychology*. Holt and Company, New York.

Zhou, Y., Carney, L. H., and Colburn, H. S. (2005). A model for interaural time difference sensitivity in the medial superior olive: interaction of excitatory and inhibitory synaptic inputs, channel dynamics, and cellular morphology. *Journal of Neuroscience*, 25(12):3046–3058.

Appendix