Neural Architecture for Echo Suppression during Sound Source Localization based on Spiking Neural Cell Models

Dissertation

zur Erlangung des akademischen Grades
Doktoringenieur (Dr.-Ing.)

an der Fakultät für Informatik und Automatisierung
der Technischen Unversität Ilmenau

vorgelegt am 1.Oktober 2003
verteidigt am 1.Dezember 2003

von Dipl.-Ing. Thomas Peter Zahn
geboren am 5. Juni 1965 in Reichenbach/Vogtl.

Gutachter:
Univ.-Prof. Dr.-Ing. Horst-Michael Gross, TU Ilmenau
Univ.-Prof. Dr. rer. nat. Benedikt Grothe, LMU München
Univ.-Prof. Dr. Thomas J. Park, University of Illinois Chicago

Summary

This thesis investigates the biological background of the psycho-acoustical precedence effect, enabling humans to suppress echoes during the localization of sound sources. It provides a technically feasible and biologically plausible model for sound source localization under echoic conditions, ready to be used by technical systems during man-machine interactions.

The model is based upon own electro-physiological experiments in the mongolian gerbil. The first time in gerbils obtained results reveal a special behavior of specific cells of the dorsal nucleus of the lateral lemniscus (DNLL) - a distinct region in the auditory brainstem. The explored persistent inhibition effect of these cells seems to account for the base of echo suppression at higher auditory centers. The developed model proved capable to duplicate this behavior and suggests, that a strong and timely precise hyperpolarization is the basic mechanism behind this cell behavior.

The developed neural architecture models the inner ear as well as five major nuclei of the auditory brainstem in their connectivity and intrinsic dynamics. It represents a new type of neural modeling described as Spike Interaction Models (SIM). SIM use the precise spatio-temporal interaction of single spike events for coding and processing of neural information. Their basic elements are Integrate-and-Fire Neurons and Hebbian synapses, which have been extended by specially designed dynamic transfer functions. The model is capable to detect time differences as small as $10\mu s$ and employs the principles of coincidence detection and precise local inhibition for auditory processing.

It consists exclusively of elements of a specifically designed Neural Base Library (NBL), which has been developed for multi purpose modeling of Spike Interaction Models. This library extends the commercially available dynamic simulation environment of MATLAB/SIMULINK by different models of neurons and synapses simulating the intrinsic dynamic properties of neural cells. The usage of this library enables engineers as well as biologists to design their own, biologically plausible models of neural information processing without the need for detailed programming skills. Its graphical interface provides access to structural as well as parametric changes and is capable to display the time course of microscopic cell parameters as well as macroscopic firing pattern during simulations and thereafter.

Two basic elements of the Neural Base Library have been prepared for implementation by specialized mixed analog-digital circuitry. First silicon implementations were realized by the team of the DFG Graduiertenkolleg GRK 164 and proved the possibility of fully parallel on line processing of sounds. By using the automated layout processor under development in the Graduiertenkolleg, it will be possible to design specific processors in order to apply the principles of distributed biological information processing to technical systems. These processors differ from classical von Neumann processors by the use of spatio temporal spike pattern instead of sequential binary values. They will extend the digital coding principle by the dimensions of space (spatial neighborhood), time (frequency, phase and amplitude) as well as the dynamics of analog potentials and introduce a new type of information processing.

This thesis consists of seven chapters, dedicated to the different areas of computational neuroscience.

Chapter 1 provides the motivation of this study arising from the attempt to investigate the biological principles of sound processing and make them available to technical systems interacting with humans under real world conditions. Furthermore, five reasons to use spike interaction models are given and their novel characteristics are discussed.

Chapter 2 introduces the biological principles of sound source localization and the precedence effect. Current hypothesis on echo suppression and the underlying principles of the precedence effect are discussed by reference to a small selection of physiological and psycho-acoustical experiments.

Chapter 3 describes the developed neural base library and introduces each of the designed neural simulation elements. It also explains the developed mathematical functions of the dynamic compartments and describes their general usage for dynamic simulation of spiking neural networks.

Chapter 4 introduces the developed specific model of the auditory brainstem, starting from the filtering cascade in the inner ear via more than 200 cells and 400 synapses in five auditory regions up to the directional sensor at the level of the auditory midbrain. It displays the employed parameter sets and contains basic hints for the set up and configuration of the simulation environment.

Chapter 5 consists of three sections, whereas the first one describes the set up and results of the own electro-physiological experiments. The second describes the results of 104 model simulations, performed to test the models ability to duplicate psycho-acoustical effects like the precedence effect. Finally, the last section of this chapter contains the results of 54 real world experiments using natural sound signals, recorded under normal as well as highly reverberating conditions.

Chapter 6 compares the achieved results to other biologically motivated and technical models for echo suppression and sound source localization and introduces the current status of silicon implementation.

Chapter 7 finally provides a short summary and an outlook toward future research subjects and areas of investigation.

This thesis aims to contribute to the field of computational neuroscience by bridging the gap between biological investigation, computational modeling and silicon engineering in a specific field of application. It suggests a new spatio-temporal paradigm of information processing in order to access the capabilities of biological systems for technical applications.

Contents

Acknowledgement

Over the years of this thesis there have been many friends, colleagues, advisors and relatives supporting my scientific work. This shall be the place to say thank you to all of you.

First of all, I have to mention my personal friend and advisor Prof. Horst Michael Gro"ss, who supported this thesis from the very beginning and provided me not just the funding but also many motivating discussions contributing to my ongoing enthusiasm for the field of computational neuroscience.
Together with him, the entire crew of the department of Neuroinformatics at the Technical University of Ilmenau has contributed to this result by providing me an scientific environment enabling this interdisciplinary work. Here I have to mention Ursula Körner once inducing my lifetime excitement for neurosciences, Klaus Debes and Hans-Joachim Böhme for long years of friendship and collaboration and Rolf Nestler and Carsten Schauer for exploring the field of audition together with me.

Of crucial importance to these results has been the inspiring environment of the Auditory Processing Group at the Max Plack Institute of Neurobiology in Munich. Here Prof. Benedikt Grothe was the one, who induced many of the ideas behind this work, provided me for more than 12 month a scientific home and gave me the extraordinary chance to conduct own animal experiments in order to get a real feeling of what the nervous system is about. The many discussions and the continuing support of Christoph Karpfer, Ursula Koch, Armin Seidl and Marianne Braun shall be rewarded with a warm thanks at this point. Special thanks goes to Sylvie Baudoux who took on the mammoth task of revising this thesis and discussing all the details with me as well as to Ida Kollmar for contributing to the experimental results. With both of you I hope to continue this way much further.

By now, there would be no chip design and even no model without the exciting team and the funding of the DFG Graduiertenkolleg GRK 164 led by Prof. Gerd Scarbata at the Department of Electrical Engineering of the Ilmenau University. Here I learned and understood what silicon engineering is about and capable of. Therefore, my personal thanks go to Prof. Scarbata for never loosing trust in the slowly emerging results, to Karten Trott and Richard Izak for taking on most of the work on circuitry and chip design and to all of my colleagues for accepting that unusual topic as a common goal of research.

Last but not least I need no mention the many nights, holidays and weekends my wife Sylvia has donated to this thesis. During all those years she has been standing tall at my side supporting me in any way possible.

Chapter 1

Problem and Motivation

This thesis models auditory brain functions for echo suppression based on computational models of dynamic spiking neurons. Since this is a true interdisciplinary task, combining elements of engineering, computer science and biology, the motivation for this study has been derived from three problems, each one out of these three fields of science.
First, the proposed model architecture aims to provide a useful tool for the mobile robot PERSES to localize sound sources in reverberating rooms in order to guide general attention during man machine interaction. Second, it aims to provide an easy programming tool to simulate dynamic neural networks based on Spike Interaction Models (SIM) with Integrate and Fire neurons. And third, it aims to provide a physiologically plausible explanation, for persistent inhibition of neurons in the Dorsal Nucleus of the Lateral Lemniscus (DNLL - a small region in the auditory brainstem) which has been measured by several biologists.
The sections of this introductory chapter have been dedicated to these three sources of motivation. The first section explains the importance and problems of echo suppression for mobile systems during the man-machine interaction. The second one introduces spike interaction models and provides 5 reasons to use them. Finally, the third one explains the persistent inhibition effect including an assumption on its functional role and a hypothesis of its root causes.

1.1 Echo Suppression during Man-Machine Interaction

At the beginning of the 21 century many technical and computer based machines penetrate almost all parts of human life. They have become daily communication partners to nearly everybody. Nevertheless, while communicating with those machines we are still limited to a narrow spectrum of artificial activities like moving mouse pointers, pushing keyboard buttons and speaking into closely attached microphones. Although, today's machines can display moving pictures and generate sound or speech fairly well, sensation and interpretation of human communication signals under natural conditions remains problematic.
Therefore, many engineers strive to understand the problems of natural signal processing and build models to overcome the current limitations. Attracted by the growing knowledge in neurophysiology, more and more engineers step back from common technical models and line up with biologists to model the proven capabilities of the nervous system during the processing of visual and acoustical signals in natural environments.

This study is concerned with a very specific feature of acoustical communication - the **localization of sound sources in reverberating environments**. In this context, it aims to add another piece to the challenging scientific task of understanding the principles of biological signal processing and to develop technical systems for effective processing of communication signals.

It has long been known that acoustical echoes, arriving from reflecting surfaces in closed acoustical environments, distort the physical phenomena of sound and confuse the technical localization of sound sources. However, natural auditory systems of many species easily suppress those echoes to a certain extend and obtain stable sound source localization to guide general and auditory attention.

During **acoustical Man-Machine Interaction**, a stable localization of sound sources, even under echoic conditions, is a critical prerequisite for the correct perception of acoustical communication signals. It enables the selection of auditory information from the auditory scenes as well as their interpretation at higher levels during speech recognition and the generation of visuo-motor reactions.
Typical examples of technical systems depending on echo suppression are Video Conferencing Systems or Voice Interaction Systems. For mobile technical systems like robots, echo suppression is also a critical ability, in order to separate and focus on specific sound sources within mixed acoustical scenes and unknown acoustic environments and to direct theirs sensors - i.e. microphones and cameras - toward the source of attention.

Especially for the combination of several senses in multimodal approaches to guide mobile systems, the auditory sound source localization provides a very useful cue to direct their initial attention. However, models of the Superior Olivary Complex, a higher brain stem structure combining the azimuthal maps of different modalities, need a stable source localization even under the reverberating conditions of closed rooms, most robots operate in. Such models have been developed for example by Gross et.al [BWH$^+$02], [SGB02]. The architecture presented here, aims to provide this stable sound source localization under echoic conditions to be used within the multimodal models for guidance of the mobile robot PERSES during navigation in a hardware store.

To obtain this stable localization of sound in reverberating rooms, any technical system has to cope with echoes, containing very similar information but arriving with significant delays up to $20ms$ from different directions in the horizontal and vertical plane. To exclude them from the evaluation of sound direction, a mechanism is needed to either separate or suppress them.

By now, there are technical solutions available to suppress echoes by inverse acoustical models or separate them by Independent Component Analysis (see section 6.2). However, a common prerequisite for their success is a stable spatial relationship between the acoustical sensor and the acoustical environment. The issue remains unsolved for mobile systems - like robots - constantly changing their position within unknown environments. Under these conditions, common computational models will lack essential information and fail to react within reasonable time.

The presented neural architecture provides *a physiologically plausible and technically feasible model to suppress echoes during sound source localization under natural conditions.* Exclusively driven by sensory information, it strives to duplicate specific features of auditory processing within the auditory brain stem as they are revealed by today's psychophysical and physiological experiments. To realize this task, the model does not need any prior knowledge on the acoustical scene or the environment.

Although the presented architecture does not cover the complex physiological mechanisms employed by nature during sound source localization, the obtained results parallel a number of important psychoacoustic experiments and could prove their stability and robustness under natural conditions. Since all parts of the proposed architecture are based on basic physiologically motivated models of auditory neurons and synapses, it may account as a reasonable base for further extension, adaptation and technical implementation.

1.2 Why to use a Spike Interaction Model?

One of the most prominent features of the proposed architecture is the attempt to model all levels of auditory processing based on a *dynamic Integrate and Fire (IF) model of nerve cells.* To distinguish networks of dynamic IF neurons from other modeling approaches, they shall be named as **Spike Interaction Model (SIM)** throughout this work. Spike Interaction Models significantly differ from common Neural Networks and communicate entirely by time-discrete, binary pulses known as *spikes.* In order to generate them at the right time and the right place, the inner- (soma-) potential of each cell has to exceed a specific firing threshold.
The time-continuous and analog valued soma potentials result from a spatial and temporal integration of post-synaptic potentials (PSP), generated at inter-neuron-link elements called *synapses.* These synapses function as unidirectional sender-to-receiver links between single cells and transform the uniform spike pulses into graded potentials with specific dynamic shapes depending on the specific dynamic transfer function of that synapse.
Each Cell contains specific internal parameters, describing the time course of the soma potential immediately after firing and the dynamic change of internal parameters depending on the firing history of that cell. For further details of the employed elements and their interactions see chapter 3 Neural Base Library and the description of Integrate and Fire Networks (IFN) below.

Based on these general principles of information processing, the simulation of Spike Interaction Models requires the dynamic solution of many coupled nonlinear differential equations and therefore, a significant computing power of conventional computers. Taking into account the engineering aim, to design on-line reactive systems, the question arises, why use this rather expensive dynamic IF cell model and *why to use spikes?*
Starting from a top level view on neural network modeling approaches, this section will describe 5 answers to this question and briefly postulate the immense computing power and technical relevance of *Spike Interaction Models.*

1.2.1 Neural Network Modeling Approaches

In today's literature the term *Neural Networks* refers to a wide variety of computational architectures and elements. It would be far beyond the scope of this work to provide a comprehensive overview on today's Neural Networks. For further insights the interested reader might refer to comprehensive books published by Zurrada [Zur92],Fausett [Fau94], Bishop [Bis96], Bulinaria [Bul02] or Bosque [Bos02]. However, it seems useful at this point, to provide a short overview on the evolution of neural network modeling approaches in order to display their historical development away and towards physiological compliance. Looking from above, two principal approaches can be observed - the *Statistical Modeling Approach* and the *Dynamic Modeling Approach.*

Statistical Modeling Approach

The roots of the *statistical approach* reach back to 1943 when McCulloch and Pitts [MP43] published their first formal model on neural functionality. It was based on the linear summation of static weight values and their comparison with constant firing thresholds. Since McCulloch and Pitts could realize the two basic logical operations NAND and NOR, they concluded, that any logical function could be computed based on this network type.

During the last 5 decades, this formal logical approach has been the base for a number of far more developed statistical models *concerned to find solutions for complex and high dimensional problems by geometric networks adapting to the inherent statistics of natural and technical signals.* Some of the most prominent and well known architectures of this type are the *Perceptron* [Ros58], the *Associative Memory* [Hop82], the *Error Backpropagation network*, the *Kohonen Maps* [Koh77], the *Neural Gases* [Mar91], the *Radial Basis Functions* [VC71], the *Support Vector Machines (SVM)* [Vap95] and the *Independent Component Analysis (ICA) Networks* [RE01][HK01].

During their course of development, the applied adaptation rules moved from strictly supervised learning - the target is known and used for error calculations - toward unsupervised learning - the target is unknown and the adaptation is obtained from the statistics of the presented input. Two of the most notable adaptation principles are the *Learning Vector Quantization (LVQ)* and the *Delta Rule.* In modern approaches, like SVM and ICA, adaptation rules become a very powerful but integrative part of the architecture itself (see [HK01]).

These Networks, and a wide variety of modifications, have proven significant power to solve a large number of technical and perceptual problems, based on software running on classical von Neumann computer architectures. Some of them are already commonly available as tool boxes in standard simulation software packages (i.e. MATLAB Neural Networks Toolbox) or stand alone applications.
By today, they have become very useful tools for engineers, but to most biologists they lack the ability to contribute to the understanding of the nervous system. The level of detail and the cell properties accessible in today physiology laboratories are barely part of these models. Therefore, they are of no significant help to bridge the gap between detailed knowledge on cell structures and the observed, but still not fully understood, functionality of nerve cells within their networks.

In general, statistical models of neural networks share the view on nerve cells and their connections as *computational elements, carrying and processing time continuous parameters describing the statistics of cell activity at the level of seconds or above* (i.e. cell spike rate per second, long term change of synaptic conductance, average dendritic potential per second etc.).

Dynamic Modeling Approach

The historical base of the *Dynamic Modeling Approach* is marked by the experiments and publications of Alan Hodgkin, A.F. Huxley and Bernard Katz in 1949 [HK49] [HH52]. They applied the Goldmann equation, describing ion currents trough a cell membrane, to Sodium, Potassium and Chloride Ions and were able to built an electrical circuit, reproducing the time course of physiological cell potentials, already postulated by Walter Nernst in 1888 [Ner88]. Three years later in 1952, they could prove the validity of their models by voltage clamp experiments at the giant axon of squid. Here, they observed congruent results on the dynamic of cell potentials in the range of *ms* from their technical model and the physiological experiment. For more detailed information see Kandel and Schwartz [KSJ91]S.81 ff.

Different from the statistical approach, Hodgkin and Huxley modeled the time dynamics of neural potentials in the millisecond range and directly compared it to the recordings of physiological experiments. In general, modeling of *sub-second time dynamics functioning as information parameter* can be seen as the principle characteristic of the Dynamic Modeling Approach for neurons and their networks.

Time dynamics has 3 principle degrees of freedom:

1. dynamic of magnitudes - the magnitude of each parameter changes over time
2. dynamic of phase - the timing relationship between different parameters changes over time
4. dynamic of frequencies - the frequency of periodic parameters changes over time

Based on the early work of Hodgkin and Huxley, a number of dynamic models have been developed using either single or several parameters of time dynamics for information representation. By example, one early work refers to the microscopic experiments of Woodhull [Woo73] and Sigworth [SN80] in the 70'th, concerned with the improvement of Hodgkin and Huxleys findings and their validation in other species, and the work of Dekin and Getting in the 80'th [DG87], explaining the mechanism behind the dynamic response properties and delayed cell responses in natural cells.

Another step toward physiological cell models has been the introduction of **dynamic neurons** in the late 80' th of the last century. Dynamic neurons do not generate spikes but use a general dynamic transfer function to model intrinsic cell functionality. Even though they use an entirely continuous simulation system with no discrete events, they are capable, to take advantage of phase and frequency as additional parameters for information coding. Typical examples have been published for example by the group of P. Tavan [KHRT94]. A common engineering approximation of dynamic neurons are the *complex neurons*, which use a rotating pointer as the model for dynamic cell potentials. However, the essential limitation of both models arises from the general assumption of oscillatory signals. Ignoring the effect of single events limits their time resolution to the range of ms and therefore, they lack the ability to explain physiological effects based on single spike events.

A separate type of dynamic networks to be mentioned are the **Time Delay Networks (TDNN)** originally developed by Alex Waibl et.al [WHL$^+$89]. TDNN introduce axonal and dendritic time delays to the interaction between dynamic neurons. Based on the modification of those time delays, they have proven significant capabilities, to solve time complex problems like speech recognition and time series approximation. However, although time delays are physiological plausible and well known, their modification is naturally limited to a narrow range of only a few *ms*, resulting from conductance changes of ion channels.

The detailed modeling of phase dynamics in TDNN marks a significant step towards physiological relevant modeling, but their nearly exclusive employment for information processing can't claim to cover all of the essential parts of natural information processing and lacks the power of amplitude and frequency dynamics. Furthermore, delays used in TDNN, in many cases, reach outside the physiological range and display a rather non physiological adaptability.

Increasing knowledge on cell physiology and synaptic transmission in the 90'th led finally to **Networks of Leaky Integrate and Fire Neurons**. Leaving the purely continuous paradigm by introducing time discrete binary pulses for inter cellular communication, they mark another important step toward physiological compliant modeling. While the term 'leaky' refers to the dynamics of magnitude, emulating the time course of membrane potentials, 'integrate' describes the spatial and temporal integration of graded dynamic potentials with a resolution of μs, capable to interpret phase dynamics as well as frequency dynamics. Finally the term 'fire' describes the ability of IF neurons to generate single, time discrete pulses (spikes) at a time resolution of μs and therefore, to generate phase dynamics (temporal relationships to spikes from other cells) as well as frequency dynamics (temporal distance to preceding spikes of the same cell).

It is easy to observe that networks of IF cells require specific inter-link-elements to transform the emitted binary pulses into graded dynamic potentials. Those *synapses* are modeled as separate dynamic transducer, employing 2'nd order filters as already proposed by Hodgkin and Huxley in their initial work 1949 [HK49]. These filters respond to the dirac-like Spikes by a specific continuous time course, consisting of a rising phase (depolarization) to reach a certain magnitude and a decay phase (repolarization) to return to the resting state. Most notably, the dynamic parameters of these graded potentials (rise-time, magnitude and decay time) can code several dimensions of information stored at the local site of every synapse by properties of ion channels and transmitter substances. In contrast to single 'weight' values, employed by many statistical neural networks, synapses in IF networks account for a substantial part of the information processing and can change their parameters on different time scales (for details see section 3.1).

Throughout the history of dynamic modeling it has been a major question for engineers and biologists, to reveal the basic principles when and how synaptic transmission properties are to be changed. The first and still relevant models for synaptic adaptation reach back to 1949 when Donald Hebb [Heb49] published it's scheme for synaptic adaptation entirely based on local potential differences between the post-synaptic membrane and the receiving cell. Today well known as the *Hebb Rule*, this basic principle has been largely modified and detailed by numerous models of synaptic transmission. Recommended further readings on this topic include Henry Markram [MLFS97] and Lary Abbot [Abb98] and their models for *Long Term Potentiation (LTP)* and *Long Term Depression (LTD)*.

A last, but not least important feature of IF networks results from the fact that a single IF cell receives input (graded post-synaptic potentials - PSP) from more than one cell. Generated at the many branches of the natural dendrite, the postsynaptic potentials overlap within the dendrite, performing a spatial integration of information received from different spatially distributed cells. Since those PSP can be excitatory (EPSP - depolarizing the membrane) as well as inhibitory (IPSP - hyperpolarizing the membrane) their interference can result in total suppression as well as maximum amplification. Since the branches of the dendrite are of different length and morphology, PSP traveling along these branches will reach the cell body at different times. In some IF network models these delays are taken into account and mark another parameter for information coding and decoding as already used in the TDNN described above.

In general, model networks based on Integrate and Fire neurons mirror the ability of natural nerve cells to integrate information over space and time based upon graded dynamic potentials containing the information locally stored and modified in synapses, and communicate by uniform binary spikes, generated at specific points in time and space.

One of the latest and a very successful modeling approaches using IF neurons is the **Spike Response Model (SRM)** of Gerstner and van Hemmen [GV94],[GK02],[VGH$^+$90], [VDC02]. Here the synaptic transfer function is described by the spike response kernel:

$$\epsilon_n(t) = J_n \epsilon(t - \Delta_n) \tag{1.1}$$

Here, J_n describes the linear scalable amplitude of the dynamic potential in response to a binary spike arriving from the sending neuron n. The term (Δ_n) refers to the specific delay of this response with regard to the spike time t_n of the sending neuron n. Since ϵ marks a standard filter response kernel of the type $1 = \int_{-\infty}^{\infty} ds\epsilon(s)$, the synaptic transfer function consists of a dynamic potential with two parameters of freedom: 1. the linear scalable amplitude J and 2. the delay Δ adjustable in the range of μs. The combination of dynamic responses with different amplitudes and delays at the site of the synapses combines the power of TDNN and IF transmission models and marks a significant step toward physiological compliant modeling. However, the SRM transmission model lacks the ability to employ different and changing synaptic response kernels. It will be shown later that especially the variation of rise time and decay time constant between different synapses mark a powerful tool for information processing in the nervous system. Additionally, physiological synapses exhibit a dynamic response pattern, where each response to a uniform spike depends on the previous firing history and will decay in amplitude and duration as a result of preceding spikes. Although SRM synapses can alter their amplitude this is, like in statistical networks, rather seen as a long term macroscopic behavior, than as a usage of magnitude dynamics for information processing at the microscopic level.

At the site of the Neuron, the SRM also adds significant physiological parameters to the dynamic modeling approach. First of all, the superposition of synaptic potentials in time and space is taken into account when integrating synaptic potentials over both parameters. Second, a physiologically well proven refractory period after each emitted spike is introduced by the standard refractory kernel $\eta(t - t^f)$. The pure soma potential $\upsilon(t)$ of SRM is therefore described by:

$$\upsilon(t) = \sum_n J_n \sum_{t_n^f} \epsilon(t - \Delta_n - t_n^f) + \sum_{t^f} \eta(t - t^f) \tag{1.2}$$

The limitation here arises again from the usage of a standard kernel η for the refractory conditions of every cell. Individual physiological Neurons can, depending on their function and morphology, exhibit a wide range of absolute and relative refractory periods and therefore, exhibit significant differences in parameters like excitability, maximum firing rate and firing time.

Finally, the SRM determines the firing time of each cell as a firing probability instead of the deterministic threshold comparison. The reasoning behind is the seemingly stochastic firing behavior of spike generation. On the other hand, if inter-spike-intervals of $1 - 10\mu s$ are used for information processing during sound source localization (as is shown later) it seems unlikely that the exact timing of spikes can be sufficiently modeled by varying the intensity of a Poisson Process as originally done in SRM. For the purpose of auditory modeling, it seems therefore more helpful, to stay with a deterministic spike generation function.

In conclusion, Spike Response Models (SRM) take advantage of the time dynamics and can interpret information coded in phase dynamics as well as frequency dynamics. However the dynamic of amplitudes is limited to the macroscopic time frame and firing times make not full use of the microscopic time resolution seen in physiological networks.

For the purpose of this work, it seemed therefore helpful to develop a even more detailed IF modeling approach named **Spike Interaction Model (SIM)**. It consists of synaptic elements with variable kernel parameters and cell models with variable refractory kernels and dynamic response properties. These elements and their interaction are described in detail in section 3 "Neural Base Library". SIM can identify and code all three dimensions of time dynamics on the detailed time scale of μs and they exclusively use single spike events for information coding, transfer and interpretation.

So far, very few is known about spatial structures, connectivity and ion-potentials within cells and dendrites. Therefore, also SIM can only attempt to reproduce current knowledge on specific neural structures, but due to the developed simple to use and to modify construction kit (see Section 3 Neural Base Library) they should be a useful tool for biologists and engineers to interpret their findings and to postulate further principles of natural information processing.

1.2.2 Reasons to use Spike Interaction Models

As mentioned during the Introduction of this section, the computational expensive usage of Spike Interaction Models has 5 reasons, four of them are relevant to this work :

1. SIMs provide three additional dimensions for dynamic information coding and processing
2. SIMs are capable to resolve and process information at the time scale of μs
3. SIMs model parameters on a level of detail, accessible during physiological experiments
4. SIMs are technically implementable on a Software and Hardware Level
5. SIMs can perform un-supervised learning entirely based upon local conditions
   (not used in the proposed architecture)

Reason 1: Additional Dimensions for dynamic Information Coding and Processing

This feature of SIM's is probably the most relevant reason why nature has chosen a type of information representation fairly unusual and computational intensive to conventional computers as of today. Beside the well established 3 dimensions of space and the commonly used dimension of magnitude, SIM use *Time* as additional parameter of information representation.

Although this might seem to add only one coding parameter, anyone familiar with dynamic signal processing, will be aware that the timing of events has tree intrinsic parameters:

- the time-distance of events in relationship to other events - phase
- the time-frame of event repetition - frequency
- and the time-course of parameter values within a certain event - dynamic of magnitudes

In case of periodic invariant signals (i.e. a sinus wave) these parameters remain stable and are rather signal parameters than coding dimensions. But natural sensory signals, used for human communication, are usually a-periodic and time variant.
Therefore, a coding scheme preserving these signal parameters will be able to extract the dynamics of phase relationships (i.e. the changing time-distance between spikes in the left and right auditory nerve), the dynamics of frequency (i.e. the changing time-frame of spike repetition in a specific auditory fiber) and the dynamics of magnitude (i.e. the changing time-course of membrane potential within a auditory cell).

Since auditory sensors code all time parameters with high accuracy, they become available to information processing throughout the nervous system. Hence, changing phase relationships between the left and the right ear can be extracted even when amplitude and frequency of both signals are identical. Short term frequency components or gaps at the level of μs in acoustical signals are decoded as frequency dynamics and alter the auditory perception of clicks and speech significantly. And dynamic magnitude responses of auditory cells can clearly distinguish the onset of acoustical signals from their continuation in case of absolutely identical physical signals.

In distinction to the current spike response models, SIM extend the kernel transfer function of synapses and neurons by dynamic properties leading to non - uniform dynamic responses depending on the firing history of the element and its soundings. This way specific parts of the signal like onsets or intrinsic changes become enhanced and available for evaluation. A detailed description of the extended kernel function is provided in sections 3.1.2 and 3.3.3 of this thesis.

It becomes clear that the intrinsic parameters of time, as mentioned above, are used as independent dimensions for information processing by natural neural networks. Taking into account that they add to the 3D spatial dimension, represented by spatial neighboring and 3D local and global connectivity, it can be imagined that a small nucleus of neural cells exceeds the computing power of a modern PC which is limited to the sequential comparison of binary values, completely ignoring the dimensions of time and space for information processing. A new type of computing architecture, based on analog-digital models of nerve cells could take advantage of at least the time parameters and realize a truly parallel information processing, capable to process several thousands of today's processor instructions within some nanoseconds. Especially for perceptual tasks, this new paradigm of information processing and processor architecture could open a wide range of computing power not accessible by today's CPU's.

Spike Interaction Models, duplicating essential features of the time-dynamics of graduated membrane potentials, are capable to model this tremendous computing power at least in principle.

Reason 2: Capability to resolve and process information at the time scale of μs

One of the most commonly agreed principles of natural information processing is the *Coincidence Detection (CoD)* of spikes. It refers to the spatial integration of postsynaptic potentials at the site of the neuron and is illustrated in figure 1.1.

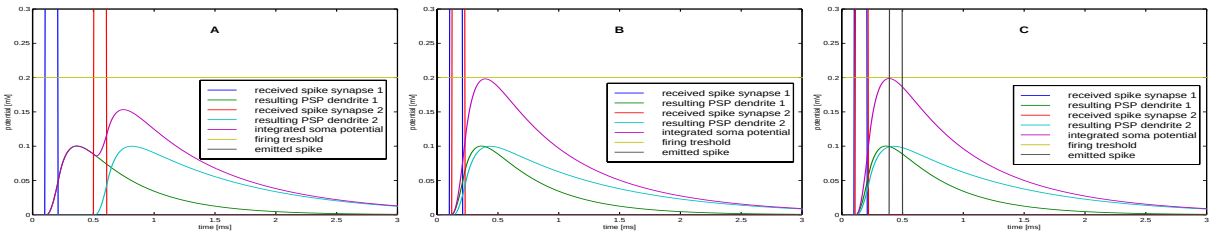

Figure 1.1: **Panel A** - *two pulses arrive with a time delay of 500μs at different synapses of the receiving neuron - they generate postsynaptic potentials of slightly different shape - these PSP overlap and shape the soma potential but do not reach the firing threshold - the receiving cell does not emit an own spike - no coincidence is detected,* **Panel B** - *the arrival delay has shortened to 20μs - still the resulting soma potential does not reach the threshold and still no coincidence is detected,* **Panel C** - *The two pulses coincide with only 10μs time difference and cause the receiving neuron to emit an own spike*

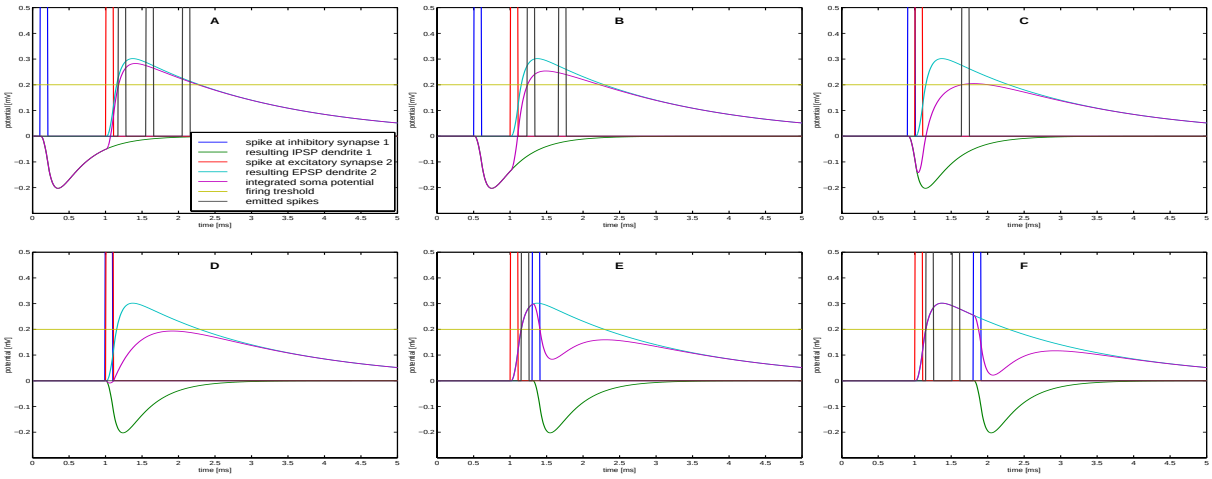

Figure 1.2: **Panel A** - *no interference of inhibitory (IPSP) and excitatory (EPSP) postsynaptic potentials due to an arrival delay larger than the IPSP duration,* **Panel B** - *interference of IPSP and EPSP results in a diminished soma potential and a decrease of emitted spikes,* **Panel C** - *arrival of an inhibitory pulse 100μs before the excitatory pulse causes delayed spiking and further decreases the number of spikes,* **Panel D** - *arrival of the inhibitory pulse coincident with the excitatory pulse erases the effect of excitation and prevents the neuron from firing,* **Panel E** - *arrival of inhibition shortly after excitation still diminishes the number of pulses emitted but does not cause time delay,* **Panel F** - *late arrival of inhibition with regard to excitation restores the reaction of the neuron to the excitatory input.*

As displayed in figure 1.1, although a single spike lasts more than 100μs, SIM networks are capable to detect time differences in the range of 10μs and less, as they are prominent in auditory signals i.e. between the left and the right auditory nerve. Also at many other places during sensory information processing, the exact timing of PSP arriving at the site of the soma

determines, if and when a specific cell responds. The emitted spikes again will interact at the next stage of processing and in this way, natural neural networks use phase differences of a few μs as information parameters employing the simple mechanism of coincidence detection.

As mentioned above, timely coincidence can not only lead to amplification of PSP among excitatory synapses but also cause interference effects, if inhibitory potentials are involved. As can been seen in figure 1.2, the timely exact coincidence of excitatory postsynaptic potentials (EPSP) and inhibitory postsynaptic potentials (IPSP) can significantly diminish, or even erase, the effect of arriving spikes. A typical result is a delayed firing as shown in Panel C of Figure 1.2 or a slowed down firing rate visible under the conditions of Panel B, C, E and F.

By these two examples it should become clear that the timing of single spikes matters and significantly changes the firing pattern of nerve cells. One might argue that if these time relationships between single spikes are stable within a given time frame, it should also be possible to model them as a delay between continuous signals (rate model). However, this is not the case. Especially acoustical communication signals change on a very short time scale, and as psychoacoustic experiments show, acoustical events in the range of microseconds alter our perception as well as the associated reactions to these signals.

Therefore, to duplicate natural information processing principles it is mandatory, to simulate them on the single spike level and assure a time resolution in the μs range. Since Spike Interaction Models are capable to do so, this has been another reason to employ them for the proposed neural architecture.

Reason 3: Accessibility of physiological parameters

One of the general goals in dynamic modeling of neural networks is, to bridge the gap between experimental physiological knowledge and technical applicable models. This subsection aims to show that SIM feature the ability to make modeling results directly comparable to experimental results and help biologists as well as engineers to understand the mechanisms of specific neural functionality.

Recording techniques in neurobiological experiments have been significantly improved during the last 30 years. By now the single neuron is no more just an anatomically observable object, it can be also studied under *"working conditions"* in vivo and in vitro (living slices). Besides improved imaging methods to study the morphology and their postnatal development, high resolution electrodes allow now access to single cells in the living brain.

Section 5.1 describes the neurophysiological experiments performed to study the echo suppression in living brain cells of the Mongolian Gerbil (Meriones Unigulatus) during this study. In order to validate the proposed model, identical stimuli have been presented to the animals as well as to the computational model and, as shown later, indeed very similar results could be obtained. However, in order to perform this type of validation the model has to give access to measurable and therefore comparable parameters.

Depending on the experimental setup, the most commonly accessible parameter is the number of spikes emitted within a given time frame after the onset of the stimulus. Hereby, the time resolution of today's experimental setups goes down to $5\mu s$. In order to compare the results it is therefore necessary, to choose a model which gives access to these parameters - i.e. models

single spikes at a time resolution at least close to the experiment. Since the employed SIM simulates all elements at a time resolution of $10\mu s$ it fulfills this requirement fairly well. Even total compliance could be reached by choosing a smaller simulation time step of the employed SIMULINK simulation system. In this model the $10\mu s$ level has been chosen, to keep the ability of the model to compute real world signals of $1-2s$ duration within a reasonable time.

Secondly, anyone monitoring the firing pattern of natural nerve cells will observe their dynamic behavior. A large portion of cells respond to changes within their input pattern with a dense burst of spikes and slows down to a periodic firing as the timely input relations remain stable. Since the initial burst in many cases preconditions the following spike pattern, it is necessary to include this general feature of nerve cells into the model. This was one of the arguments to further improve the Spike Response Model and work with the dynamic IF Cells of the Spike Interaction Model.

Finally, additional features of natural spikes like shape and duration of action potentials could be studied in nature, during the physiological experiments and have been transferred to the model.

Beside the accessible parameters in the specific experimental setup used in this study, the world of neurophysiology knows a lot more about nerve cells at the level of ion-channels, transmitter properties and morphological differences between specific cell types. The modeling approach of dynamic IF Neurons with variable internal properties could at least include some of them. For example time constants of cells in the MSO (Medial Superior Olive - a brainstem region seen to bee responsible for Interaural Time Delay evaluation) observed in the early work of Grothe [GS93] and of time constants of synapses within the DNLL (Dorsal Nucleus of the Lateral Lemniscus - another region in the lower brainstem hypothetically realizing the echo suppression during sound source localization) observed by Wu in 1998 [Wu98] have been included into the model.

Although, at this study, SIM do not model the specific properties of ion-channels and transmitters, their ability to define separate dynamic properties to each synapse and cell, opens a wide range of modeling parameters for static and dynamic details, observed in physiology. This way, macroscopically observed cell and network behavior can be studied by variations of microscopic model parameters and these models may help biologists, to understand the functional relevance of intrinsic cell parameters as well as their role within the network of specific nuclei and projections.
On the other hand, the modeling of microscopic, physiologically plausible dynamic parameters might help the engineer, to achieve an overall functionality of neural architectures similar to those, studied in psychoacoustic experiments and useful in real life.

Reason 4: Easy to implement on a Software and Hardware Level

The value of any model to a physiologist is often limited by the complexity of action, necessary to implement structural and parametric modifications in order to watch their implications in a simple and easy to understand manner. The accessibility and the handling of programming and modification is therefore a critical parameter, when modeling neural networks, aiming to help during the interpretation of physiological findings.
On the other hand, a valuable model for engineering purposes needs to be able, to perform in

a real world environment. To achieve this, computing has to be done on a real-time level or at least during very short computing times.

Both, the level of detail and user accessibility demanded by physiologists and the computing speed needed by engineers, usually compete with each other and are barely to be aligned in one modeling approach. The here presented architecture addresses this antagonism by taking advantage of today's advanced commercial simulation tools and silicon implementation techniques.
Since SIM entirely consist of single elements modeling nerve cells, synapses and connecting elements like axons and dendrites, a small library of those elements proves sufficient to create any type of neural network model. Additionally, their parameters can be modified from outside without complicated programming. But at the same time, the dynamic character of SIM requires the generation and solution of higher order differential equations - a rather complicated matter usually not in focus of physiological research.

A well suited base to combine the two requirements is provided by the dynamic simulation tool of the MATLAB family, named SIMULINK. This entirely graphical oriented software tool takes care of the generation and solution of the appropriate differential equations and opens the world of dynamic simulation to the inexperienced user. Additionally, it provides a number of standard libraries containing elements for signal generation, display and general processing from simple math up to higher order filtering.

During the course of this thesis, a specific Neural Base Library (NBL) has been created, containing several types of basic SIM elements as described in chapter 3. They are instantly available after installing MATLAB 5.2 with SIMULINK 2.2 or higher, at any laptop or desktop and loading a 250k file, free for distribution by the author. Using this library, anyone can create Spike Interaction Models by simply using the mouse pointer to move those elements out of the library into a graphical simulation panel, connecting them by visible lines to variable types of sources or display elements provided by the SIMULINK environment. This way, it will take only about two hours to gain the ability to built simple SIM.

Since all of the neural elements will be instances of the Neural Base Library, their general properties can be simply changed by graphical modifications of library elements. At the same time, individual parameters, like the dynamic shape of specific PSP or dynamic thresholds of neurons, are accessible by editing the mask of each element in the simulation workspace. For further detail see chapter 4 Neural Model Architecture. Additionally, the SIMULINK environment allows the user, to group portions of the model into subsystems (i.e. all neurons and synapses of a specific functional nucleus) creating a general view onto more complex models, where single elements are still easily accessible.

Furthermore, by the use of "scopes", a display element provided by SIMULINK, any parameter (for example the time course of the soma potential within a specific neuron) can be displayed during the simulation or afterwards. The elements of the NBL therefore contain specific outputs providing visual access to all of their intrinsic parameters during simulation, if needed. This feature allows the experimenter to "look behind the curtain" and might help to understand the specific reaction of individual elements (i.e. firing patterns) under defined dynamic conditions.

Finally, the simple principle of graphical programming provided by the SIMLULINK environment and the NBL, allows the physiologist to implement the proposed, or any other SIM architecture easily on any standard PC, and to modify its parameters depending on whatever hypothesis he is looking for. However, the question is how to make them working in real time for engineering purposes?

There are 3 principle ways to speed up simulation times of SIM generated in SIMULINK environments.

1. **Automated generation of C-code** running on conventional PC processors
2. **Automated generation of Assembler code for Digital Signal Processors (DSP)**
3. **Silicon implementation on specific processor chips** based on pre-designed analog-digital layouts for SIM elements

The first two of these options are commercially provided by the "Real-Time Interface to Simulink" of the dSPACE GmbH in Paderborn and require only little additional efforts for implementation. The general speed up by conventional C code depends on the specific architecture and ranges between 2 and 5 times faster computation for the proposed architecture. Automated generation of assembler code, executed on the specifically optimized DSP, shortens the simulation time approximately by a factor of 10 for the presented rather complex architecture and can reach real time performance in many cases of more simple SIM.
Finally, the silicon implementation will reach real time performance in any case and provides specific integrated Hardware/Software solutions suitable for a wide range of commercial applications. Corresponding hardware layouts for basic SIM elements have been developed and tested within the DFG Graduate Colleg GRK 164/1-96 "Automated design of analog and mixed analog-digital structures applied to neural networks" at the Technical University of Ilmenau. For further details of hardware implementability see section 6.3.4 Silicon Implementability.

Reason 5: Ability to perform unsupervised learning based upon local conditions

Beside the four reasons to use Spike Interaction Models for the proposed architecture, another important feature of the SIM modeling approach, concerning the adaptation to environmental signals shall be mentioned. Although no learning or adaptation principles have been implemented in the " Neural Architecture for Echo Suppression during Sound Source Localization", many other SIM depend on this general ability of natural neural networks to modify their synaptic properties in order to extract and store knowledge about the environment and behavioral relevant signal features.

It has been mentioned during the discussion of dynamic modeling approaches that natural synapses change their dynamic properties depending on the local electrical and chemical conditions resulting from specific firing patterns. Although not all secrets of this mechanism have been revealed by today, the general principle of Donald Hebb is widely accepted and has been applied in many models. Therefore a Hebbian like learning principle has been introduced in the neural elements "learning dynamic synapse" and "learning dynamic neuron" of the Neural Base Library

presented in section 3. Here, a mechanism for long term modification of synaptic properties has been introduced resulting in a modified Hebbian Learning Rule called *Timing Of Potentials - or TOP learning.*

The implemented model for TOP Learning realizes Long Term Potentiation (LTP) as increase of synaptic conductance as well as Long Term Depression (LTD), as decrease of synaptic conductance. The amount and direction of modification here depends on the local relationship between postsynaptic potentials, resulting from a presynaptic spikes of the sending neuron, on one side, and dendritic potentials, resulting from PSP of neighboring synapses as well as back-propagating postsynaptic spikes of the receiving neuron on the other side. In summary, the difference of both potentials will define the number of NMDA channels available for activation and change the conductance of the synapse by either diminishing them in case of depression (LTD) or increasing them in case of potentiation (LTP). Using the specific parameters of dynamic IF neurons, this adaptive mechanism modifies not only the amplitude of postsynaptic responses but also the decay time constants, and therefore the influence of the specific synapse onto the soma potential of the receiving neuron.

By using this NBL element, or a even more detailed adaptation mechanisms, SIM are capable to modify their intrinsic transfer relations between specific cells as well as neural nuclei, depending on the local interaction of presynaptic and postsynaptic spikes and/or potentials. This way, they can duplicate learning as well as developmental changes in network connectivity, helping the physiologist to explain and the engineer to access, adaptive properties of neural networks.

In conclusion, these 5 reasons to use Spike Interaction Models make them the appropriate modeling approach for architectures aiming to duplicate physiological findings and make them applicable to technical systems acting in the real world environment. They have been used during this study as well as for other models to guide auditory attention [ZITP97] and to model auditory pitch perception (unpublished).

1.3 Persistent Inhibition in the DNLL

As mentioned during the introduction, a third motivation for this thesis has been derived from the physiological problem, to provide a plausible and functional model, reproducing the measured results on *Persistent Inhibition* in DNLL neurons and their role during echo suppression.

As of today, the physiological circuitry realizing the complex task of echo suppression is not fully revealed, but current experimental results suggest a significant role of the Dorsal Nucleus of the Lateral Lemniscus (DNLL). This small nucleus of auditory neurons marks a significant transformation station within the auditory brain stem and is mainly populated by inhibitory neurons, with some of them showing Excitatory/Inbitory (EI) response properties. Hence, they receive excitatory input from the contralateral (opposite) hemisphere and inhibitory input from the ipsilateral (identical) hemisphere. The physiological and functional details of this nucleus are described in the sections 4.5"The Lateral Lemniscus" on page 110 and will therefore not be detailed here.

In the late 90'th, a specific feature of DNLL cells has been revealed by Pollak et al. [YP94c],[YP98] during experiments with the mustache bat. If the bat had perceived a preceding sound from a location in one hemisphere, any subsequent sound arriving from the opposite hemisphere did not cause the appropriate DNLL cells to fire immediately as expected. They fired with a significant delay of up to 30 ms. Hence, they seem to be inhibited even after the inhibitory signal had ended. This phenomena is called **Persistent Inhibition - (PI)** and is described in detail in the subsection "Peristent Inhibition" on page 45.

At this point it shall be mentioned that actual physiological experiments in Bats and Cats [Yin94][Pol97][KK00][BP01] show evidence that inhibitory influences from the DNLL shape the direction sensitivity of specific EI Neurons within the Inferior Colliculus (IC), as the next higher auditory center.

Under onset conditions (no preceding sound) these IC cells become contralaterally excited by lower auditory centers, possibly the Cochlear Nucleus (CN). Several studies [FBAR93] [LK92] [VHKG92] [PP94] could reveal that the same IC cells perceive complementary inhibition from the DNLL.

Since the firing pattern of DNLL cells, for reasons given later, will depend on the direction of the arriving sound signal, either one of the IC hemispheres will be inhibited by the referring DNLL and therefore the IC displays a direction sensitive firing behavior as well. This is the normal condition and has been supported by several experiments, lately again by Litovski et al. in the IC of cat [LSC01].

But if the DNLL doesn't fire during that period, the referring IC Cells will not be inhibited and will therefore continue to duplicate the excitatory pattern of the Cochlear Nucleus despite the fact that the direction of sound has changed. Hence, the directional sensitivity of these IC Cells

is suppressed during the period of persistent inhibition and they respond as if the sound still arrives from the direction of the first sound.

This finding have be verified by experiments presented in this thesis and seems to hold for the DNLL of the Gerbil as well (see section 5.1 Physiological Experiments). Therefore, it seems likely to be a common feature of the auditory processing in mammals.

Comparing these findings with well known results from psychoacoustic experiments, persistent inhibition seems to account as a possible mechanism for the historical *Law of the first wave front* [Hen49],[CM76][Bla74] and the similar *Precedence Effect* first studied and published by Wallach et al. in the middle of the last century [WNR49]. Both effects will be further elaborated in section 2.2 Acoustical Echoes and 2.3 Summing Localization and the Precedence Effect, describing the basics of physiological echo suppression during sound source localization.

If persistent inhibition in the DNLL is the basic underlying principle of those proven effects the question arises: Where does it come from?

In their early work Yang and Pollak [YP94c] showed that the blocking of GABAergic inhibition within the DNLL and of glycinergic inhibition towards the DNLL, abolished the persistent inhibition and restored the neurons ability to respond to subsequent sounds as if there was no preceding sound from another direction. At the same time they could prove the early work of Boudreau [BT68] that the general formation of EI properties (without persistent inhibition) takes place in a lower auditory center (the Lateral Superior Olive - LSO) and not in the DNLL. Both features suggest that persistent inhibition is evoked within the DNLL itself and depends on specific inhibitory effects within this auditory nucleus.

Another indicator arises from the physiological findings of Glendering [GBHM81] , Shneidermann [SOH88] and Markovitz [MP93]. They found that the DNLL perceives strong inhibitory projections from the ipsilateral LSO and the contralateral DNLL.

The general hypothesis of this study is that the push-pull effect of inhibition and excitation within the DNLL and its surrounding auditory nuclei, naturally results in a period of persistent inhibition without a need for specific inhibitory inter-neurons, extended latencies or additional circuits. The proposed architecture aims to provide a physiologically consistent model of the several auditory nuclei involved and simulates relevant physiological and psychoacoustic experiments.

As mentioned above, one major motivation of this thesis was to provide a plausible explanation for the effect of persistent inhibition and make it available to mobile technical systems for echo suppression during sound source localization.

The following chapter 2 will therefore explain the general principles of auditory sound source localization and physiological echo suppression.

Chapter 3 then introduces a multi purpose dynamic simulation system developed and employed to model and explain the principles of sound source localization including the persistent inhibition effects within the DNLL.

The following Chapter 4 will describe each of the modeled centers within the auditory brainstem in detail. It starts with the outer ear and moves step by step up to the inferior colliculus resulting in a complex spike interaction model, capable to perform echo suppression during source localization based on detailed models of auditory neurons, nuclei and their connections.

Chapter 5 describes the experimental results derived from physiological experiments with the Mongolian Gerbil as well as from psycho-acoustical and real world experiments, using designed and free field recorded signals. Here, the responses of living biological cells are directly compared to the model cell behavior under dynamic conditions.

Finally, Chapter 6 discusses these results in the context of other biologically motivated and technical models for echo suppression and briefly introduces the achieved status of silicon implementation.

At the end, chapter 7 will provide a short outlook towards further improvements and possible applications of the developed Spiking architecture for echo suppression during sound source localization.

Chapter 2

Sound Localization and Echo Suppression

The aim of this chapter is, to provide a short introduction to the natural principles of sound localization, the generation and impact of echoes and their auditory suppression by the Precedence Effect. It contains a short overview of relevant psychoacoustic and physiological experiments including current hypothesis on neurophysiological circuits.

2.1 Interaural Disparities - Neural Cues for Sound Localization

Sound is one of the most prominent communication signals in the animal kingdom and is used for interaction in all kinds of natural media like air, water and even the subterranean environment. In general the term *sound* refers to the oscillation of molecules in elastic media. It's intensity and dispersion properties therefore depend on the characteristics of the medium it travels in.
A *sonic event* is usually composed of a number of sound components with specific frequency, phase and amplitude characteristics changing over time. The resulting physical phenomena is a time series of pressure changes spreading spherically from the sound source into all directions of the medium. It is characterized by the physical parameters sound pressure, frequency and phase, whereas the sound pressure or intensity decreases proportional to the distance from the sound source. Since naturally sensed sound intensities spread over a wide range, they are usually described by a logarithmic measure the *Sound Pressure Level (SPL)*. It is defined by the relationship:

$$SPL = 20log\frac{P_x}{P_0}[dB] \tag{2.1}$$

and relates the real sound pressure P_x to a reference pressure P_0 marking the human perceptual boarder at 2×10^{-5} Pa. Under this definition, the observable sound pressure for humans reaches from 0 dB (perception threshold) to 120 dB (pain barrier). The frequency of sound is described by the usual unit of Hertz $[Hz] = \frac{1}{s}$ - describing the number of repetitions per second. It's phase is either characterized by the absolute measure of *Time Difference (TD)* in seconds $[s]$ or the relative measure *Phase Difference (PD)* referring to a single cycle and measured in radiant $[rad]$. The part of science dealing with the physical phenomena of sound is generally named *Acoustics*, whereas the perceptual side of sound is referred to as *Audition*.

Sound is sensed by the most fragile sensory organ - the ear. Almost all species use two instances of ears at the left and the right hemisphere of the body. These hemispheres are referred to as *aural fields* and from here, the term *Binaural Hearing* arises - simply describing the dual character of sound perception.
Since Binaural Hearing is common within nature, it is obvious that auditory systems use the differences of the same sonic event arriving at the two ears for information processing. These differences are referred to as *Interaural Disparities*. The two most common Interaural Disparities are the **Interaural Intensity Difference (IID)** - describing the SPL Difference between the two ears measured in $[dB]$ and the **Interaural Time Difference (ITD)** - describing the phase difference between the two ears measured in $[\mu s]$. Please note that in German literature IID is sometimes mentioned as IPD - Interaural Pegel Difference not to confuse with the IPD - referring to Interaural Phase Differences as the relative measure for ITD. To avoid confusion, throughout this thesis IID will be used for Interaural Intensity Differences and ITD for Interaural Phase Differences. Although there are also differences within the frequency spectra of the two ears, no common measure is yet defined for them and they wont be subject to this study.

The question now becomes: Where do these interaural disparities arise from and how are they employed for sound source localization? The general principle is shown in figure 2.1 and is due to the different relative positions of the two ears with regard to the location of the sound source.

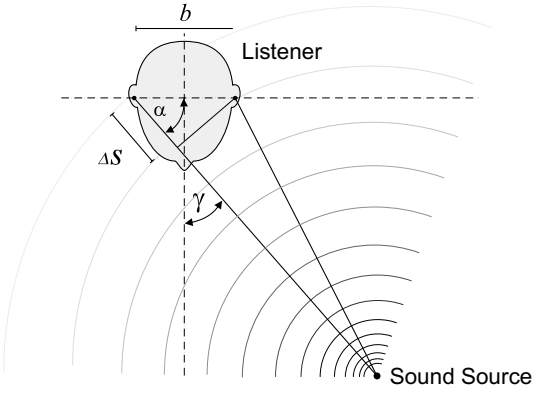

Figure 2.1: Interaural Disparities during sound perception

As the sound travels from the sound source toward the head of the listener, any wavefront first reaches the (ipsilateral) ear on the hemisphere of the sound source and only after traveling the distance ΔS it reaches the ear on the opposite (contralateral) side. If the traveling medium is air, with a sonic speed of $340ms^{-1}$, the Interaural Time Difference can be calculated to:

$$\Delta ITD = \frac{\Delta S}{340} = \frac{cos(\alpha) \cdot b}{340} \tag{2.2}$$

This simple relationship holds in principle for all frequencies as long as the deflection of the wave front remains small compared to the auditory base b. Hence, this relationship holds as long as the distance between sound source and listener is large enough (for humans the critical distance sits at about at 1 meter). However, since the wavelength of higher frequency sounds shortens to the the range of μs (i.e. wavelength of 5 kHz = 200 μs), the interaural traveling distance ΔS exceeds the wavelength and the Interaural Phase Difference will be identical for several different

sound source locations. This is one of the reasons why ITD's are mainly used to locate low frequency sounds up to 1 kHz.

Beside the principle of ITD generation, another feature becomes visible in figure 2.1. As the sound propagates through the medium it looses energy and therefore, the amplitude of each wave front decreases proportional to the distance from the source (see shading of the wave fronts in figure 2.1). As the sound reaches the ipsilateral ear, it contains more energy than at the time arriving at the contralateral side. This *Interaural Intensity Difference* is further enhanced by the shadowing effect of the head. Since the tissue of the head has a higher density than air, the sound looses additional energy during penetration. This energy absorption is a nonlinear function of frequency and rises with the frequency level, but is prominent throughout the entire spectrum. Hence, IID's result from distance differences with regard to the sound source location and energy absorption by head tissue. They mark a robust and widely used cue for sound source localization and will be the major cue employed by the proposed model of auditory processing.

The relationships mentioned above describe the main cues for sound source localization in the azimuthal plane. However, spectral cues and extended mechanisms for IID and ITD evaluation are employed to localize sound sources in the vertical plane and to distinguish between front and back locations. However, since the focus of this work is concentrated on the localization task within the horizontal plane and specially directed towards mechanisms for echo suppression, the introduction to interaural cues will stop at this point. For a deeper and more comprehensive introduction to interaural disparities one should refer to Blauert [Bla74], Zenner [Zen94], Cremer [CM76] or Zwicker and Fastl [ZF99].

As outlined in section 4 "Model Architecture and Physiological Correlates", interaural disparities become evaluated by the lower nuclei of auditory processing in the Superior Olivary Complex (SOC) and are the major information parameters the neural auditory system employs for localization of sound sources. Therefore, they also mark the critical information parameters coded and processed by the proposed architecture for echo suppression during sound source localization.

2.2 Acoustical Echoes

Acoustical Echoes are a common experience for everybody. They are usually associated with an extreme case, when short words are repetitively perceived for instance after calling into steep mountain walls. Here, the originally emitted sound is reflected by the hard mountain wall and returns to the callers ear after traveling the distance toward the wall and return. If this distance is long enough, the reflected sound returns with a delay longer than the time needed to end the call and is perceived as a repetition "by the mountains". For example a call of 200 ms duration is fully separately returned, if the distance to the wall is at least 34 m. This phenomena also occurs in any city with skyscrapers but due to the significant acoustical background it is not perceived as strong as in a quite mountain valley.

Another common, but annoying experience of echoes is known in large railway station halls or airports, when the announcements of loudspeakers seemingly overlap them self, making it a hard task to extract the speech content from the announcement. This has two reasons. The first is due to several loudspeakers emitting the same signal at the same time but at different distances to the listener. In this case, the sound of the far speaker reaches the ear significantly later then the sound of the near speaker and overlaps it with similar intensity, disrupting directional information as well as the content. By today, this problem has been widely managed by careful positioning of speakers and artificially adapted signal amplitudes and delays.
The second reason is also true for a single speaker in large reverberating rooms. Here, the perceived disturbance results from the reflection of sound by the walls of this room. If their distance to the listener is large enough, the reflected sounds will arrive at the listeners ear significantly after the sound arriving directly from the source. As long as the resulting delay is larger than about 30 ms and strong enough to exceed the background, it is perceived to interfere with the original and causes the same annoying "mixture".

The delay at which the listener perceives the reflected sound as a separate sonic event has been first investigated by Haas [Haa51] and is well described in the text book of Blauert [Bla74] pg.179 ff. It has been found to range between 2 and 30 ms and is referred to as **Echo-Threshold**.

However, the effect of overlapping reflected sound is also widely used to create a positively perceived amplification in large concert halls. By today, it has become an entire section of acoustical science and architecture to calculate and optimize the reflection properties of concert halls in order to diminish the delay of original and reflected sounds below the Echo-Threshold. In this case the human auditory system no longer perceives two separate sounds but groups them together and perceives a louder signal. This natural amplification has already been used by ancient speakers i.e. in the Roman arenas and during the Island "Thing" happening at the high mountain wall of the continental divide.

From these examples, the general nature of echoes as reflections of sound by hard surfaces becomes transparent. The amount of reflection depends on the density of the walls surface. As the sound travels trough a medium like air, it is reflected by any surface with a density higher than air. A simplified relationship is: The higher the density of the wall the higher the

amplitude and the broader the spectrum of the reflection. Therefore a concrete or tiled wall will reflect nearly all frequencies with much higher amplitude, than the same wall, covered by a soft curtain. Based on this, environments with many hard covered walls are referred to as *closed acoustical environments* or *reverberating environments.* On the opposite, specific rooms with walls covered by materials totally absorbing and not reflecting any sound are called *acoustically dead environments* or *unechoic chamber* and finally free field conditions with no significant reflection surfaces are called *acoustically open environments.*

A typical measure to describe an acoustically closed or reverberating environment is the so called *Reverberation-Radius.* Figure 2.2 shows a common example of a click sound recorded during this study in a reverberating environment with a Reverberation-Radius off 200ms.

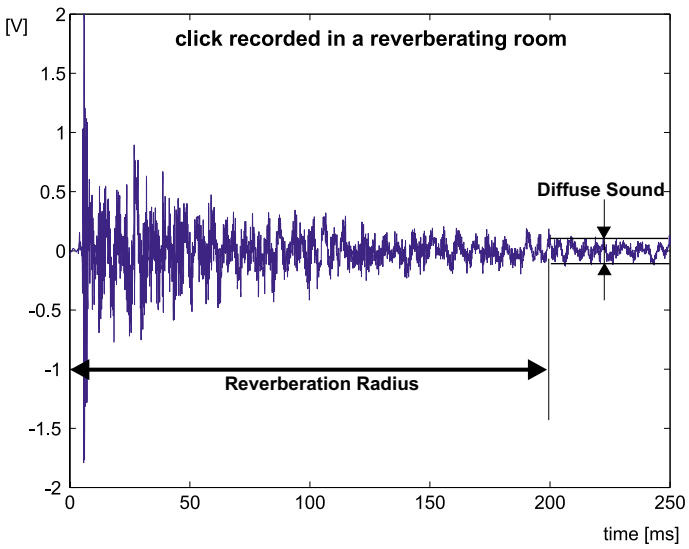

Figure 2.2: Reverberation radius of a click sound recorded in a highly reverberating environment

As shown, the Reverberation-Radius refers to that time after the sonic event, when the signal level of the primary sound just equals the level of the diffuse sound field. Within the Reverberation-Radius the primary sound is extended by overlapping sound reflections from the surrounding walls. Under the condition of figure 2.2 it is perceived as a prolongation of the sonic event and results in a less precise ability to localize the sound source. This is due to the fact that the Reverberation Radius exceeds the Echo-Treshold of 30 ms described above.
Auditory systems use this feature to determine the distance of a sound source, while evaluating the difference between the first sonic event and its reflected components during the prolonged primary sound. As can be seen, there is no sharp boarder line between the prolonged primary sound and the diffuse sound resulting from the background noise of the specific environment. As a general rule it is observable that: the higher the background level, the shorter the distance a sound source or echo is perceivable.
Since the determination of sound source distances will be not subject to this study, this effect will not be further elaborated at this point. For further details consult the early publications of Hornbostel [vEMH26]and Maxfield [Max33] or more the comprehensive description in the Blauert textbook pg.222 ff.[Bla74].

The reflection properties of sound also follows the general law of mirrored reflection known from light. Hence, the angle of arrival will be equal to the angle of reflection. As displayed in figure 2.3, in a closed and highly reverberating environment, the sound is reflected more than once and arrives at different points in time from a variety of directions, different from the source direction.

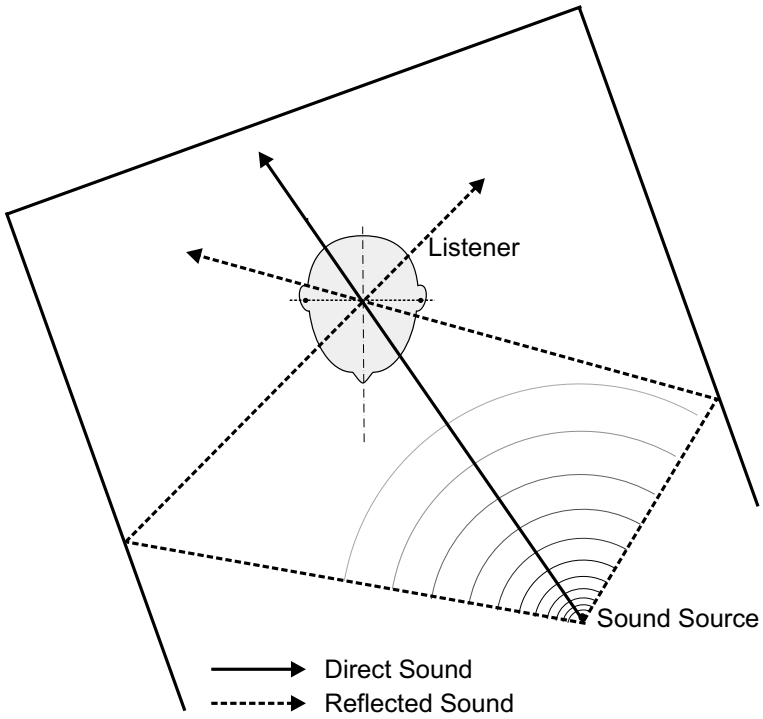

Figure 2.3: Echoes during sound perception

If the room is small as displayed in figure 2.3, the delays of the reflected sounds range below the Echo-Threshold and cause an amplification. The interesting effect is that in this case the amplified sound is still perceived as originating from the direction of the sound source, despite the physical fact of different components arriving as reflection from different walls surrounding the listener.

This effect of grouping sounds from any direction to the direction of the first sound is known as the **Law of the First Wave Front** and has been first investigated by Henry [Hen49] and later evaluated by Cremer [CM76].
Since the *Law of the first Wave Front* is the major effect realized by the human auditory system to suppress echoes during sound source localization it will be in focus of this study and shall be introduced in greater detail in the next section.

2.3 Summing Localization and the Precedence Effect

As mentioned in the last section, the *Law of the First Wave Front* applies to all sounds arriving at the human ear during a short period, after a sonic event. The upper boundary of that period is marked by the *Echo Threshold* at delays around 30 ms and the lower boundary sits at delays of 0.6 to 1 ms [Bla74]. It states that: *Any sound arriving after a first sonic event, with delays or intensities below the Echo Threshold is perceived to originate from the direction of the first sound* - see Blauert pg.178 [Bla74]. This effect is not based upon physical and propagation properties of sound, but is caused by the auditory processing within the nervous system. Therefore, it only occurs at a perceptual level and is so far not accessible for technical systems relying on microphone recordings of the physical phenomena.
During the last century, a large number of psychoacoustic experiments has been performed revealing most of the psychoacoustic features of this law. However, until now there is no clear evidence which physiological structures and mechanisms are responsible for the *Law of the first wavefront*. To introduce the current status, this section will provide a short overview on the main psychoacoustic experiments regarding echo suppression and current neurophysiological hypothesis on the neural circuits carrying it out.

2.3.1 Summing Localization

The Law of the first wavefront is known since long, and the earliest psychoacoustic experiments go back to the 30'th and 40'th of the last century.
In the 30'th, Blumlein [Blu31] was the first to realize the possibility, to generate spatial sound effects, when transmitting electroacoustical recordings of concerts. He tested which differences between two separately presented sound signals will cause the listener to perceive a sound location, different from that of the two speakers. Blumlein, DeBoer [DV39] and a number of acoustic engineers during the following years found that the listener perceives a virtual sound source location, if the sound arrives from two speakers with no, or very little, time or intensity difference at the two ears.

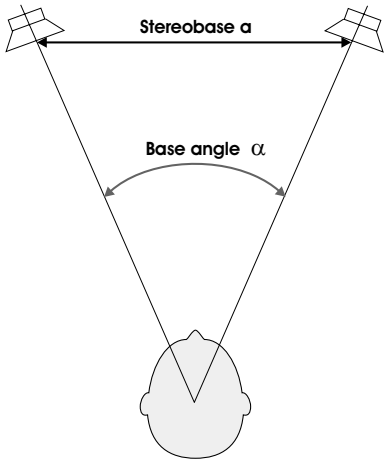

Figure 2.4: Traditional setting of stereo experiments

As displayed in Figure 2.4, the traditional stereo setting consists of two speakers, positioned with a symmetrical base angle α to the listener. If the two speakers emit absolute identical signals, the location of the sound is perceived to originate just from the midline. As the two signals become slightly delayed, the virtual sound source location moves toward the location of the earlier signal and reaches it at about 0.6 - 1 ms delay. Based on these insights, the recording of a classical concert was transmitted between Philadelphia and Washington first in 1933, using a two speaker setting to generate a stereo effect during perception.
Warncke [War41], concerned with the generation of stereo-phonic effects for the just born tone-movies, called this effect **Summing Localization** and employed it to generate spatial impressions for tone-films. A first systematic quantification of the Summing Localization effect has been done by Wendt [Wen63] in the 60'th and can be followed in detail at the Blauert Text Book [Bla74]. Most recently, Litovsky et.al [LSC01] comprehensively investigated the phenomena of binaural echo suppression including the Summing Localization. The revealed upper boundary of the Summing Localization effect around 1 ms remained stable throughout all kinds of experiments and marks the lower boundary for the Law of the precedence effect.

2.3.2 Precedence Effect

Another early experiment was published by Wallach in 1949 [WNR49] [WNR73]. He increased the time difference of the two speakers in stereo settings (see Fig.2.4) beyond the upper boundary of the Summing Localization and found the test persons, to perceive the location of the speaker carrying the preceding signal as the source location for both signals.
He called this effect **Precedence Effect** and obviously described the *Law of the First Wave Front* by this term. *The precedence effect therefore is nothing else, than the experimental observation of the Law of the First Wave Front.* From now on, it will be preferably employed to describe the effect of echo suppression during the specific period after a sonic event, covered by the Law of the First Wavefront.

A large number of psychoacoustic experiments in the 60'th and 70'th has been concerned with the precedence effect and studied it's boundaries as well as several parameters. However, a comprehensive presentation of them would clearly lead beyond the scope of this work. Therefore, only a rare selection of significant experiments will be presented here. For a detailed summary of experimental results one might refer to the original Blauert Textbook [Bla74]pg.177 ff, the actual version of it [Bla01] pg.409 ff. or the review of Litovsky et al.[LCYG99].

Echo Thresholds and their Parameters

The mayor parameter investigated to study the precedence effect has been the *echo treshold* and its dependency on Intensity Difference (ID) between signal and echo, Inter Signal Time Delay (ISTD), signal type and signal duration.

Already in the 50'th, Lochner and Burger [LB58] and Meyer and Schodder [MS52] independently investigated the Echo-Threshold, using the traditional stereo setting with a base angle $\alpha = 80°$

and speech signals of 5 to 15 syllables per second. By varying the Inter Signal Time Delay and Inter Signal Intensity Difference (ID) between the leading signal (S_0) and the trailing signal (S_T), they asked the test person to determine, when the echo was separately perceivable. The observed curves are displayed in Fig.2.5.

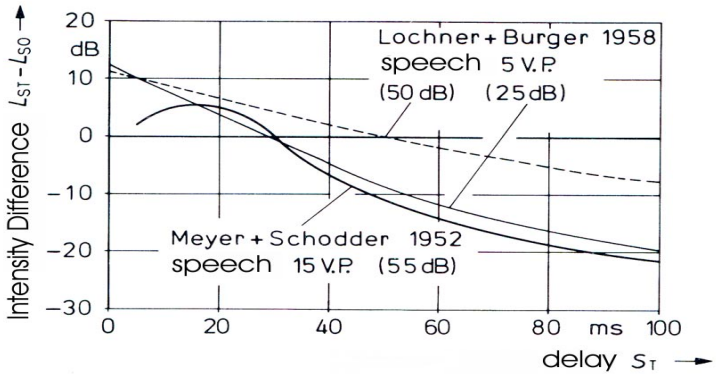

Figure 2.5: Echo-Thresholds determined using a standard stereo configuration with base angle of 80 degrees for continuing speech with an average speed of 5 syllables per second

Their results suggest that at very short delays, already very small intensity differences are sufficient to suppress the echo. Common to the two lower curves is the result that even when the trailing sound was absolutely louder than the leading sound (positive ID), it was not perceived as separate sonic event for delays up to $35ms$. Hence, the perception of the trailing echo is suppressed even if this is unnaturally loud.

Another important early experiment has been carried out by Damaske [Dam71]. He investigated the dependency between Echo-Threshold and *signal duration* by using broad-band noise signals of different length. Different from others, he did not employ the traditional stereo setting, but had the leading signal source (S_O) positioned in the midline of the listener and the trailing signal arriving from the left hemisphere (see Fig.2.6 upper right corner).

It can be observed in Fig. 2.6 that signal duration had only little influence on the Echo Threshold as upper limit of the precedence effect. Up to delays of $20ms$ and ID's of $10dB$ the three curves are nearly identical. However, a general impression looking at the whole experiment is that the precedence effect increases slightly as the signal becomes longer.
A second generally notable observation of Damaske is the fact that the precedence effect is also prominent if the signal originates from the midline and neither one of the two hemispheres.

The signal duration influence becomes more prominent when comparing the curves of equal intensity obtained at several different experiments. Figure 2.7 shows the curves of equal intensity between leading sound and echo obtained from David and Hanson [DH62] - curve (a and e), Lochner and Burger [LB58] - curve(b), Haas[Haa51] - curve (c) and Meyer and Schodder [MS52] - curve (d).

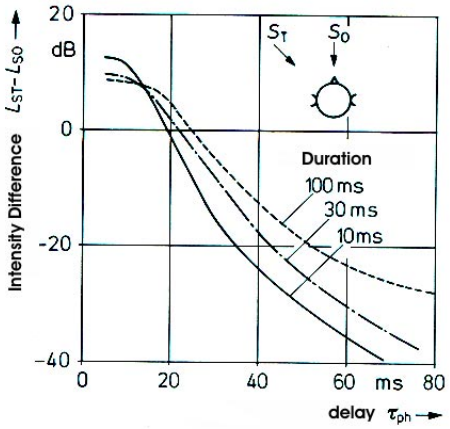

Figure 2.6: Echo-Thresholds for noise impulses of different durations under the condition of a frontal primary sound source S_O and a secondary source from $\alpha = 89°$

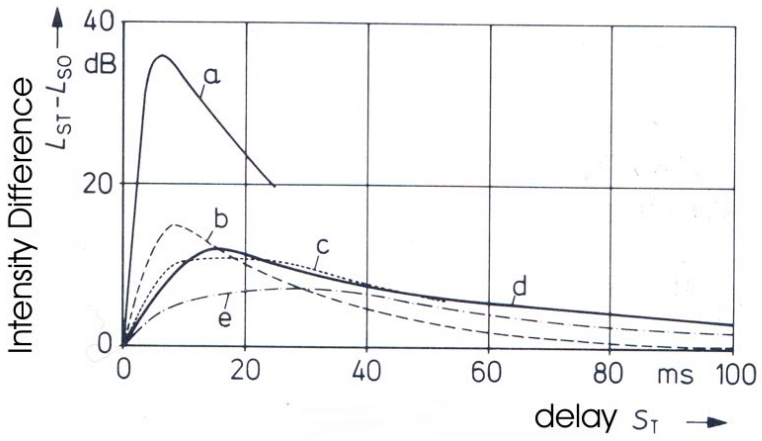

Figure 2.7: Echo thresholds for continuous speech of average speed (5 syllables per sencond); standard stereophonic loudspeaker arrangement, base angle $\alpha = 80°$ from Blauer [Bla96]pg.226

As displayed, for most speech signals (curves b,c and d) echoes, which are 10 dB louder, need to be delayed by 15-20 ms until the echo is perceived equally intense as the leading sound.
In case of long, slow raising pure tone signals, as employed by Lochner and Burger to obtain curve e, the echo is already perceived equally loud if it is 5 dB louder than the signal, but at much larger delays around 20-40 ms. However, the general course of curve e is very similar to those of the speech signals (b,c and d).

For a high pass click with steep signal raise, marked by curve a, 10dB louder echoes are perceived if they arrive only 2 ms after the leading signal. Hence, echoes of short signals and clicks are perceived at much shorter delays but need to be much louder (higher ID). But a general effect is visible at all 5 experiments. During a significant period after the leading signal (delays up to $60ms$) echoes become generally muted since they are perceived as equaly loud even if they have much higher Sound Pressure Level than the leading sound.

Based on these, and a number of other experiments, Blauert [Bla74]summarizes on pg. 184 f:

- The smallest echo-thresholds around $2ms$ are observable for short clicks of $1 - 5ms$
- For longer slow rising signals the echo-threshold is higher and relative constant - for speech it sits around $20ms$
- Steep onsets of signals as well as high intensities of the trailing sound shorten the Echo-Threshold. Accordingly, trailing sounds of smaller intensities need longer delays to be perceived as an echo [BS66].
- The precedence effect is also present, when the leading sound originates from the midline of the listener.
- The precedence effect is not due to free field sound propagation since it is also present during dichotic presentation - for example, separate presentation of the first sound to the left ear only and of the second sound to the right ear only by the use of head phones.

Based on this summary Blauert concludes that a contralateral inhibition process is very likely to be involved in the generation of the precedence effect under dichotic as well as free field conditions. This suggests that a contralateral inhibitory connection needs to be included in the neural model architecture in order to realize the precedence effect and a sufficient suppression of echo directions during sound source localization.

Discrimination Threshold

Meanwhile, a number of further experiments have been carried out and revealed an additional parameter of the precedence effect. As shown in Table 2.1, during Summing Localization lead and lag signal are perceived as a joined sonic event from a virtual source in between the speaker positions. The Echo-Threshold refers to the upper limit of the period, when echo suppression causes the listener to perceive lead and lag signals still as a joined sonic event but from the direction of the lead source. The new parameter **Discrimination Threshold** now - *describes the upper boundary of the discrimination suppression period, during which the echo is perceived as a separate sonic event but also originating from the direction of the lead source.* Above the Discrimination Threshold, lead and lag are perceived as separate sonic events originating from separate locations.

Echo Suppression and Discrimination Suppression have been studied among others by Yang and Grantham [YG97b] and Litovsky [LSC01].

Litovsky et al.[LSC01] found the Echo Threshold for 1 ms noise clicks in six normal-hearing listeners at 1-5 ms, but the Discrimination Suppression remained potent for delays of 10 ms or longer. At the longest delays tested, two distinct sounds were perceived, but they were not always heard at independent spatial locations.

The results of these experiments suggest that directional cues of the echo are not necessarily salient for all conditions, when the lag is subjectively not heard as a separate event. In other words, during discrimination suppression the echo information is perceived by the auditory system, but the localization mechanism remains disabled.

Overview on Echo Suppression Effects

Effect	Lead-Lag Intervall	Sonic Event	Location
Summing Localization	0 to 1 ms	joined	joined virtual
Echo Suppression	1 to 20 ms	joined	joined lead source
Discrimination Suppression	20 to 30 ms	separate	joined lead source
No Suppression	above 30 ms	separate	separate

Table 2.1: *Suppression effects during the perception of lead and lag stimulus pairs:*
joined - perception of only one sonic event or one sound source location
joined virtual - perception of only one sound source at a virtual location
joined lead source - perception of only one sound source at the location of the leading source
separate - perception as two separate sonic events or two sound source locations

This hypothesis is also supported by findings of Freyman, McCall and Clifton [FMC98], studying normal-hearing listeners' sensitivity to changes in the intensity of the lagging sound. They conclude that the precedence effect does not consist of a general suppression or attenuation of the lagging sound, but rather a suppression of directionality cues.

The Discrimination Threshold is a perceptual parameter describing the upper boundary of a period, where the echo is perceived but cannot be discriminated to arrive from a different direction. In most cases it is higher than the Echo Threshold and suggests that the auditory content of the echo is prominent in the auditory brain, while the directional information is still suppressed. In terms of the developed model architecture, discrimination suppression is observable at the level of the IC. Here echoes cause clearly distinguishable sonic events (spikes) but since they are nearly equal in intensity at both hemispheres their directional information is completely suppressed (see section 4.6 The Inferior Colliculus)

The Clifton Effect

Additionally to the investigation of perceptual thresholds, experiments of Clifton and Freyman [Cli87] [CF89] revealed the adaptive nature of the precedence mechanism. In his first experiment Clifton found that test persons, listening to clicks presented from a standard stereo setting with a few milliseconds delay, perceived both clicks for some seconds separately, when the location of the leading source was switched from one side to the other. During the second experiment when a click train with lead and lag stimulus of defined delay was presented several times, the subjects reported a "fade-out" of echo clicks after a number of clicks sets at each delay, regardless of rate. This result has been interpreted as a *buildup in inhibition* of echoes produced by the ongoing click train. This so called *Clifton effect* suggests that adaptation is prominent in the precedence circuit and might be carried out within higher auditory structures.

Extended results of the adaptive nature of echo suppression where observed by Freyman, Clifton and Litovsky [FCL91], when a train of conditioning clicks or noise bursts was presented in a stereo setting just before a click/noise pair containing an echo. In seven of nine listeners, perception of the lagging sound was strongly diminished by the presence of a train of "conditioning" clicks

presented just before the test click. Echo threshold increased, as the number of clicks in the train increased from 3 to 17. For a fixed number of clicks, the effect was essentially independent of click rate (from 1/s through 50/s) and duration of the train (from 0.5 through 8 s). But when only the lead sound was presented during the conditioning train, the perceptability of the lag sound appeared to be enhanced.

Another more complex experiment of Clifton [CFLM94]proposes that echoes provide information about room acoustics, which the listener picks up during the ongoing sound and uses to form expectations about what will be heard. When expectations are violated by changes in the echo, this disruption can be seen in a lowering of echo threshold, relative to the "built-up" threshold when expectations are fulfilled.

The Clifton effect describes the adaptive nature of the precedence effect, which can be modified by conditioning signals or changes in the spectral content. It suggests the involvement of higher auditory centers during the generation of the precedence effect.

Interaural Disparities and/or Spectral Disparities

Another question raised by experimenters since the 80'th is: *Whether the precedence effect relies on Interaural Disparities (ITD and IID as introduced before) or Spectral Disparities (resulting from spectral changes during the reflection of sound) or both of them?*

To answer this question, Yang and Grantham [YG97a] tested the *Spectral Overlap Hypothesis - SOH* of Blauert and Divenyi [BD88], and the *Localization Strength Hypothesis (LSH)* later proposed by Divenyi [Div92]. The Spectral Overlap Hypothesis states that: "The higher the coincidence between the spectra of the lead and the lag stimulus, the more suppression is exposed to the lag stimulus".

This hypothesis was supported by experimental results, when the test persons heard stimulus pairs with 5-ms, 1-octave, weighted noise bursts of 65 dB where lead and lag had been parametrically set to center frequencies of 0.5, 2.0, or 3.0 kHz. Discrimination thresholds were higher, when lead and lag center frequencies coincided than when they did not coincide.

The Localization Strength Hypothesis on the other hand states: "The greater the localization strength of the lead stimulus, the greater suppression it exerted on discriminability of the lag sound position" This was supported in another experiment, when lead and lag stimuli were 8-ms, 1.5-kHz weighted tone bursts of 65 dB, with lead and lag rise times parametrically set to 0, 2, or 4 ms. In this case, the amount of discrimination suppression increased as lead rise time became more abrupt or as lag rise time became more gradual. Since they could support both hypothesis, they found that spectral overlap and localization strength appear to be two relatively independent factors governing discrimination suppression.

The psychoacoustic experiments cited above, have been concerned with the precedence effect as a perceptual phenomena and investigated its dependencies on signal parameters and stimulation conditions. However, in order to find a consistent hypothesis to explain this phenomena, physiologists have studied the responses of living cells and nodes within the auditory system.

2.3.3 Neurophysiological Experiments and Hypotheses

Since the 70'th of the last century the question: *What physiological structures may be involved to generate the precedence effect?* concern many laboratories. Only some of the major experiments and derived hypothesis shall be mentioned here to justify the neural architecture chosen to model the precedence effect. However, the interested reader might refer to a comprehensive review published by Litovsky, Colburn, Yost, and Guzman in 1999 [LCYG99].

Involvement of the Auditory cortex

One of the first questions investigated was and still is: *To what extend is the auditory cortex involved with the echo suppression task of the precedence effect?.*

There has been a relevant number of experiments, suggesting a rather minor role of cortical structures regarding sound source localization. For example, Whitfiled experimented in the 70'th with unilateral ablation of the whole auditory cortex in cats. His publication in 1974 [Whi74] states that the precedence effect becomes disrupted after ablation, but later reports [WDCW78] put these results in question, since some of the animals still turned their head to the leading source after one hemisphere of the auditors cortex was ablated. This might be partly caused by the learning effect generally involved in cortex ablation studies and therefore only of limited significance.

Experiments with brain damaged humans have been carried out by Cornelisse [CK87]. Here, patients with lesion in the left hemisphere, in areas outside the temporo-parietal region, did not display deficits under any condition. Patients with discrete right temporo-parietal lobe lesion were able to localize single clicks, but frequently reversed the apparent perceptual locus of paired clicks in the hemifield contralateral to the side of lesion.

Induced by publications of Hafter et al. [eaH88], who reported the release of *post-onset adaptation* by additional short trigger signals with different spectral properties, Blauert [BCV89] tried to observe a similar "active release process", to overcome the precedence effect, but found this not to be true for a number of exploratory precedence-effect settings.

Moore et al.[MCR90] investigated the ability of human patients with unilateral temporal lobe lesion, multiple sclerosis, or dyslexia, previously shown to disrupt neural timing. Their results suggest that sound localization, using stimulus conditions known to elicit the precedence effect, places greater demands on neural timing and integration than conventional tests of localization, and may depend on a more sensitive index of neural function.

Finally in 2000, Litovsky, Hawley, Fligor and Zurek [LHFZ00] disproved a hypothesis of Saberi and Perrot [SP90] suggesting that listeners can *unlearn* the suppression of the lag's directional information after training with an adaptive psychophysical procedure. Since Listeners showed no sign of unlearning after 9-13 hours they conclude, hat directional information contained in the lagging source is not easily accessible, hence the suppression of this directional information is not realized in cortical structures, accessible to psychophysical learning mechanisms.

On the other hand, recent studies show evidence that echo suppression at least involves higher auditory centers. Selected examples are:

One experiment of Liebenthal, [LP97] exhibits a significant and specific reduction in binaural peak amplitude of cortical responses to auditory evoked potentials for the echo-evoked middle-latency component. A later experiment [LP99] was concerned with position judgment and auditory-evoked potentials (AEPs) in response to single- and pairs of binaural and monaural clicks, simulating a source and its echo. Here again, the binaural echo suppression depended upon echo lag, although less strongly than the psychoacoustic position judgment. With a seemingly wide step they Liebenthal concludes from these results a primary cortical involvement in echo-lateralization suppression.

Mickey and his colleagues [MM01] studied the responses of cortical neurons in areas A1 and A2 of anesthetized cats. Single broadband clicks were presented from various frontal locations. Additionally, paired clicks were presented with various delays and Intensity Differences from two loudspeakers located 50 degrees to the left and right of midline. Units typically responded to single clicks or paired clicks with a single burst of spikes. At delays of 1-4 ms, unit responses typically signaled locations near that of the leading source with substantial undershoot - in agreement with localization dominance. A superposed Intensity Difference shifted location estimates toward the more intense source, reaching an asymptote at 15-20 dB.

But taking in account the multi-lateral influences to the auditory cortex, the recordings of Liebenthal and Mickey rather confirm the existence of strong inhibitory influences somewhere in the auditory system than prove a active role of the auditory cortex during their generation, as suggested by Liebenthal only.

Although there is no clear message, to what extend the auditory cortex is involved in the generation of the precedence effect, most of the experiments suggest the involvement of a higher auditory circuit, but not necessarily the auditory cortex. From the current point of view, a likely candidate for this may bee found in the first major auditory center below the auditory cortex - the inferior colliculus (IC).

Involvement of the Inferior Colliculus

The inferior colliculus (IC) accounts for the first auditory center within the brainstem merging the seemingly separate pathways for IID and ITD evaluation (see section 4.1 The Auditory Brainstem). Since it is common to all vertebrates able to hear and known to be involved in binaural auditory processing, many investigations regarding the precedence effect focused on this central part of the auditory system.

For example Yin [Yin94] recorded from single neurons in the the Inferior Colliculus of anesthetized cat. Click stimuli were delivered under two different situations: over headphones in dichotic experiments and through two speakers in an anechoic room in free-field studies. He found the cell's responses in the cat's IC suppressed for short Inter Stimulus Time Delays (ISTD) and the time to reach 50% recovery, ranged from 1 to 100 msec with a median of 20 msec. With short ISTD's in the summing localization range (about +/- 2 msec) he found cells also to show responses consistant with the human psychophysical results that the sound source is localized to a phantom image between the two speakers and toward the leading one. Based on these consistent findings he concluded that the discharge of at least some cells in the IC is related to the perceived location of the sound source.

A psychophysical experiment involving brain damaged humans was published by Litovsky, Fligor and Traino [LFT02] involving a 48 year old man (RJC) with a small traumatic hemorrhage of the right dorsal midbrain, including the IC. The results suggest that: (1) localizing sounds within a given hemifield relies on the integrity of the contralateral IC, (2) unilateral IC lesion gives the illusion that sound sources in the 'bad' hemifield are displaced towards the 'good' hemifield, (3) the IC mediates aspects of echo suppression, and (4) lesion in the IC does not impede spatial release from masking in speech intelligibility, possibly due to that function being more heavily mediated by cortical or lower regions.

Among others, Keller and Takahashi,[KT96] studied the IC of the barn owl. They presented brief sounds simulating signal and echo under free field conditions and recorded extracellularly from individual space-specific neurons within the IC. Space-specific neurons responded strongly to the direct sound, but their response to a simulated echo was suppressed, typically, if the echo arrived within 5 ms or less of the direct sound. With inter-stimulus delays of less than 10 ms, the owl consistently turned its head toward the leading speaker. Longer delays elicited head turns to either speaker with approximately equal frequency and in some cases to both speakers sequentially.

Results of special interest were observed by Fitzpatrick [FKBT95] from single neurons in the inferior colliculus of the unanesthetized rabbit. Here, monaural and binaural click pairs (conditioner and probe) where delivered through earphones and a special type of cells (early high) was found, well responding to echoes (probes) even during conditions when they where NOT localizable. Hence, localization suppression does not necessarily depend upon excitation suppression it could also result from the suppression of inhibition and this way, make the echo recognizable but not localizable.

Further support to this hypothesis is given by experiments of Burger and Pollak [BP01] [PBP$^+$02]. They recorded from single cells in the IC of the free-tailed bat and found that EI neurons in the inferior colliculus (IC) that are excited by one ear and inhibited by the other can code interaural intensity disparities (IIDs). Although EI properties of many cells are formed in lower nuclei and imposed on some IC cells via an excitatory projection, many other EI neurons may be formed de novo in the IC. By reversibly inactivating the Dorsal Nucleus of the Lateral Lemniscus (DNLL) they showed that the EI properties of many IC cells are formed de novo via an inhibitory projection from the DNLL on the opposite side.

They also found that signals excitatory to the IC evoke an inhibition in the opposite DNLL that persists for tens of milliseconds after the signal has ended. During this period, strongly suppressed EI cells in the IC are deprived of inhibition from the DNLL, thereby allowing the IC cell to respond to trailing binaural signals to which it previously responded poorly or not at all. By relieving inhibition at the IC, it becomes obvious that an initial binaural signal essentially reconfigures the circuit and thereby allows IC cells to respond to trailing binaural signals that were inhibited when presented alone.
Thus, DNLL innervation creates a change in responsiveness of some IC cells to binaural signals, depending on the reception of an earlier sound. The experimental results of Burger and Pollack suggest that the circuitry linking the DNLL with the opposite central nucleus of the IC is important for the processing of IIDs that change over time, such as the IIDs generated by moving stimuli or echoes that emanate from different regions of space.

Somewhat similar to Fitzpatrick and Burger and Pollak, actual experiments of Litovsky and Delgutte [LD02] studying the inferior colliculus of cats conclude as well that the inhibitory inputs causing suppression in the cats IC may originate in part from subcollicular auditory nuclei. However, the key message of a comparative experiment between humans and neurons in the IC of cats, published by Litovsky, Rakerd, Yin and Hartmann in 1997 [LRYH97] is that the precedence effect operates also in the median sagittal plane, where binaural differences are virtually absent. From here it is concluded that precedence is mediated by binaural and spectral localization cues and a hypothesis is established that models attributing the precedence effect entirely to binaural differences are no longer viable. From the existence of suppression within the midline they conclude, that interaural disparity cues are not necessary for neural correlates of the precedence effect to be manifested.

While the experimental results of Litovsky on working echo suppression in the median sagittal plane are doubtlessly true, their conclusion seems to go too far. As could be demonstrated by the proposed model architecture, a precedence like effect naturally evolves even in the median sagittal plane without involvement of spectral cues, if the IID sensitivity functions of the left and right LSO overlap sufficiently (see section 4.4.3 LSO Model).

As a common result of all those experiments on the involvement of the IC during echo suppression it seems clear that the response properties of specific neurons in the inferior colliculus are mediated during a period of at least 2-20 ms after the lead stimulus, resulting in a specifically unsuppressed reaction to the trailing echoes. The interesting question is now, whether these mediations arrive from lower auditory nuclei in the superior olivary complex (SOC) or the dorsal ducleus of the dateral demniscus (DNLL) or they are intrinsically generated within the different substructures of the IC.

The experiment of Burger and Pollack [BP01] already suggested a significant role of the DNLL within this circuit with regard to IID's. Therefore some additional experiments concerned with the functional role of the DNLL will be described in the next paragraph.

Involvement of the DNLL

As described more consistently in section 4.5 Lateral Lemniscus, the DNLL is one of the major nuclei within the lateral lemniscus (LL) carrying information from the medial superior olive (MSO), and the lateral superior olive (LSO) towards the central inferior colliculus. During the last 10 years there has been a significant amount of work concerned to reveal the physiology and functionality of the DNLL.

Leading to the hypothesis of Burger and Pollak [BP01] [PBP$^+$02] cited above, an extensive amount of work has been done by Pollak at al. to study and prove the GABAergic inhibition of IC Cells by the DNLL [PPWL92],[YPR92], [PP93c], [PP94], [KPP95],[WLP95]. These projections are described in further detail in section 4.6 The Inferior Colliculus and have proven sufficiency to significantly modify and sometimes fully suppress the response of specific cell groups within the central IC. Of special interest to many of these experiments was the change in timely response of newly formed IC-EI cells to lag or echo stimuli. Here it has been shown that GABAergic inhibition

arising from the DNLL generates some EI properties in the contralateral IC but weakens or even abolishes responses during a period of up to 20 ms after the lead stimulus. The major role of DNLL cells to generate this property has been further supported by pharmacological and lesion experiments of Park [PP93a],[PP93b], Yang [YP97] and Kelly [KLv96],[KL97],[IvK96].

Persistent Inhibition in the DNLL

Another prominent feature of DNLL Cells, called **Persistent Inhibition - (PI)**, has been revealed during follow up experiments by the groups of Pollak in the mustache bat and Kelly in the rat [YP94a],[YLP96],[Pol97],[YP98],[KK96],[BVdRR98],[KBP99], [BKP00]. During these independent experiments with varying conditions they found that a substantial portion of DNLL cells exhibits extended EI properties. Hence, they receive excitatory inputs from the contralateral LSO and inhibitory projections from the ipsilateral counterpart [KBK98]. However, different from EI Cells within the LSO, they stay persistent inhibited after a first stimulus for as long as 20 ms. This *Persistent Inhibition* is typically level-dependent and increases in duration with increasing sound level. According to Bauer [BKP00], the Persistent Inhibition is also sensitive to the duration of the stimulus, with short (5 ms) tones being less effective than longer ($>$ 20 ms) tones in generating persistent inhibition. By contrast, Yang [YP98] found that stimulus duration had only little influence on the duration of the persistent inhibition. Having in mind the early psychoacoustic experiments of David and Hanson [DH62] both can be correct since they found the echo threshold to be short for very short signals and clicks but to stay around 20 ms for longer signals as well as speech (see figure 2.7).

The existence of persistent inhibition in DNLL-EI cells has been also confirmed by the physiological experiments during this study. Here we recorded from the DNLL of the gerbil and where able to reproduce the general results of some experiments by Pollak and Yang. Methods and results will be presented in detail in chapter 5 “Experimental Results“.

What remains unclear until now, is the underlying mechanism generating persistent inhibition of these DNLL cells. Therefore, it is the aim of this study, to provide a valid mechanism for the generation of Persistent Inhibition within the DNLL and prove its efficiency to suppress echoes at the level of the IC. So far, three major hypothesis regarding the mechanisms to generate persistent inhibition have been formulated:

- Kelly and Kidd [KK00] suggest that NMDA receptors of DNLL cells account for the persistent component of inhibition, while AMPA receptors seem to bee involved with the “normal“ direct inhibition by the ipsilateral SOC. One hypothesis derived from here is that *NMDA receptors might cause extended time constants and therefore generate long lasting inhibitory potentials.*

- Another hypothesis relies on *efferent projections from the IC or the auditory cortex.* Although such efferent influences possibly exist, and may cause the adaptive behavior revealed by the clifton effect, it seems unlikely to have them generating such a timely precise and specific behavior, clearly depending on the spatio-temporal structure of the sensory input.

- A third hypothesis finally involves *strong contralateral inhibition between the two DNLLs on the left and the right hemisphere.* Shown by vanAdel et al. [vAKK99] a disection of the commissure of probst and therefore a cut of the inhibitory influences between the two DNLL resulted in a much less effective reduction of response amplitude after ipsilateral acoustic stimulation. Responses to both short (+/-1.0 ms) and long(1.0- 30.0 ms) intervals were affected. Strong GABAergic projections between left and right DNLL were also observed by Chen, Kelly and Wu [CKW99] in slices of rat brains and by Zhang [ZLKW98] in living rat brains. Furthermore, such projections are seen to be part of the Persistent Inhibition circuit by Yang and Pollack [YP98] [PBP+02].

Based upon the observation of dynamic cell potentials within the model architecture an *extended hypothesis on the generation of Persistent Inhibition within the DLLL* has been established during this study. It states that: *Persistent Inhibition within the DNLL results from the Push-Pull effect of either relatively weak contralateral excitation by the opposite LSO or strong twofold inhibition by the ipsilateral LSO and the contralateral DNLL.*

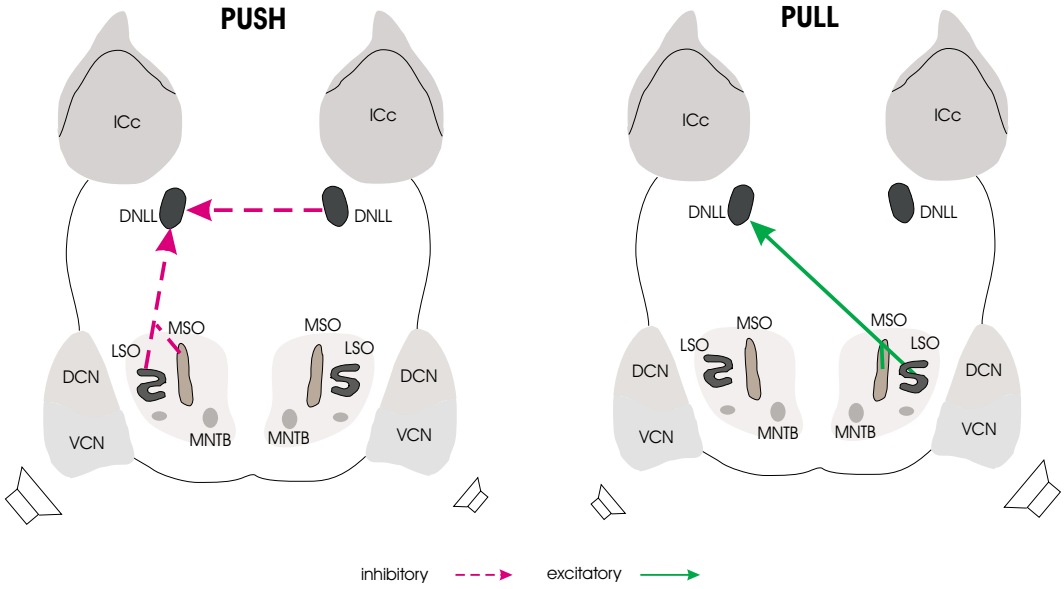

Figure 2.8: Push-Pull effect on DNLL - EI cells generating Persistent Inhibition
left panel: *Inhibitory Push of the left DNLL cells toward maximum hyperpolarization by inhibitory inputs from the right DNLL and the left LSO/MSO in response to a lead stimulus from the left hemisphere - note that no excitation is present*
right panel: *Excitatory Pull of the left DNLL cells toward the firing threshold by the right LSO during a lag stimulus from the right hemisphere*

As shown in the left panel of Figure 2.8, in case of an ipsilateral leading stimulus, the soma potential of left DNLL cells with EI properties is strongly hyperpolarized and pushed toward low soma potentials due to GABAergic inhibition from the opposite DNLL and glycinergic inhibition from the ipsilateral LSO. Different from the conditions in the LSO and the IC, at this time no competing excitatory input to the DNLL is active. The time needed to repolarize back to the resting potential, therefore only depends on the intrinsic membrane and channel properties of the DNLL cells.

If a trailing sound arrives within that time frame from a different direction, as shown in the right panel of figure 2.8, the soma potential will be pulled toward the firing threshold. But assuming a relative weak excitation compared to the achieved level of hyperpolarization this will just shorten the time needed to repolarize the cell but will not cause immediate firing as expected without a inhibitory lead signal. Simulations led to persistent inhibition between 10-20 ms depending on the achieved hyperpolarization during the lead stimulus and the intensity of the lag stimulus.

Since the level of hyperpolarization depends on the number of inhibitory impulses (i.e. IPSP's) received during the lead stimulus, hyperpolarization will be stronger if the lead was loud, had a steep onset, or lasted longer. However, physiological hyperpolarization has an upper limit around -75 mV. If this is reached, the soma potential will stay there and not decrease further. Therefore, persistent inhibition has an upper limit and will not further increase in case of an ongoing or very intense lead stimulus. Depending on the relation between maximum hyperpolarization and the strength and number of excitatory impulses (EPSP's) during the lag stimulus, typical periods of persistent inhibition have been found to range between 10 and 20 ms. As discussed in section 5.2 "Psychoacoustic Experiments", the properties of this mechanism correlate well with many of the psychoacoustic experiments cited above with regard to inter stimulus delay (ISD) as well as inter stimulus intensity difference (ISID).

Since this hypothesis somewhat extends the first and third hypothesis mentioned above, it incorporates current experimental knowledge, while providing a plausible and simulative supported explanation for the observed Persistent Inhibition of DNLL - EI cells. Overall it supports a major tendency in the literature to assign a substantial part of the circuits causing the precedence effect to the inhibitory projections of the DNLL toward the central IC.

Involvement of Lower Auditory Nuclei

Of course, there have been also experiments concluding an involvement of lower auditory nuclei like the cochlear nucleus (CN) and the superior olivary complex (SOC) during echo suppression. By example, Kaltenbach [KMF+93] found a specific forward masking of cells within the dorsal cochlear nucleus (DCN) and discussed a possible role of this effect during echo suppression and Grothe [GN00] mentioned a possible role of inhibitory effects in the medial superior olive (MSO) during echo suppression.

However, the point of view taken in this study is that those nuclei are largely concerned with the generation of time precise and specific spatially arranged excitation pattern, necessary to enable the DNLL push-pull effect introduced above as well as a correct spatio temporal processing within the IC and the auditory cortex. These are specifically:

1. The decoding of source locations within the left and right hemisphere into spatially separated firing pattern of left and right nuclei by evaluating IID's and ITD's within both the LSO and the MSO.

2. The encoding of absolute intensities (SPL) by the firing rate and the number of frequency channels involved.

3. The encoding of onset steepness by dynamic firing pattern, exhibiting intense spike bursts of many frequency channels in case of steep onsets and phase locked or chopper responses for slow raising tones.

4. The encoding and enhancement of spectral and modulation cues including their delivery to higher auditory centers.

The specific nature of cell responses within the auditory nerve and the nuclei of the CN and the SOC will be subject to the next chapters and shall not be elaborated beyond the general expectations listed above at this point.

Finally, only a last general observation of Fitzpatrick [FKK+99] shall be mentioned. His group recorded excitation pattern to paired clicks with varying interstimulus intervals, from several structures of the ascending auditory system in unanesthetized cats including the auditory nerve, the antero-ventral cochlear nucleus, the superior olivary complex, the inferior colliculus and the primary auditory cortex. Their main finding was a progressive increase in the duration of the suppressive effect of the leading sound on the response to the lagging sound. In neurons of the auditory nerve, the cochlear nucleus, and the superior olivary complex, 50% recovery of the response to the echo sound occurred, on average, for intervals of approximately 2 ms. In the inferior colliculus, 50% recovery occurred at an average separation of approximately 10 ms, and in the auditory cortex at approximately 20 ms. However, these are very top level observation not taking in account the specific nature and complex arrangement in most of these structures and can only be seen as a general rule that significant suppression of echoes occurs only at higher centers of auditory processing. i.e. all of the echo information is preserved during the sensation and first decoding of auditory scenes.

Summary

Intensive studies within the last 60 year have revealed that,

- Interaural time differences (ITD's) and interaural intensity differences (IID's) are the major cues for auditory sound localization in the azimuthal plane.
- Echoes are natural phenomena in closed acoustical environments, caused by sound reflections of dense surfaces. They arrive shortly after the original sound with a variety of delays and from different directions.
- The perception of echoes is totally suppressed during the echo suppression period and their directional information is further suppressed during the period of discrimination suppression.
- Echo suppression is stronger for loud signals with steep onsets and longer duration. Echoes louder than the signal can shorten the suppression period and overcome suppression in rare cases.
- Early echoes are stronger suppressed than later ones.
- Echoes are also suppressed when the sound originates from neither one of the listeners auditory hemispheres and no interaural disparities are present.

- Echo suppression can be built up by conditioning stimuli and disrupted by sudden modifications of the frequency spectrum and/or the interaural cues - clifton effect.
- Information on echoes is present even if their direction is not perceived
- A substantial part of echo suppression might result from disabling the directional sensitivity of specific EI cells in the central inferior colliculus normally inhibited by the DNLL.
- The timely exact dis-inhibition of specific IC Cells can result from the persistent inhibition of DNLL cells following a lead sound.
- Persistent inhibition in the DNLL might be caused by a specific push-pull effect of strong doubled inhibitory and weaker excitatory projections from the LSO on both hemispheres and the contralateral DNLL.

In accordance with most of the psychoacoustic and physiological experiments cited above the proposed model realizes echo suppression at the level of the LSO - DNLL - IC projection. It essentially exhibits the following features:

1. Echo suppression under dichotic as well as real world free field conditions
2. Intensity dependency - stronger echo suppression for lead signals with high intensities. Weaker suppression of very intense echoes.
3. ISD dependency - less effective echo suppression as the inter stimulus delay (ISD) between signal and echo increases.
4. Duration dependency - longer echo suppression for longer durations of the lead signal with upper limits around 20 ms
5. Effective echo suppression if the sound originates from the midline of the listener.
6. Preservation of echo information during suppression of directional information.

Chapter 3

The Neural Base Library

As mentioned in section 1.2.2, a major advantage of the Spike Interaction Model (SIM) proposed in this study lies within its modular structure based upon the basic neural elements of a Neural Base Library (NBL) and ready for graphical programming as well as silicon implementation. The base elements contained in this library are shown in figure 3.1 and will be introduced step by step during this chapter. At the end of the chapter, some general considerations and simulation conventions will be introduced in order to ease the understanding of simulation results and parameters, shown during the subsequent introduction of the model architecture in chapter 4.

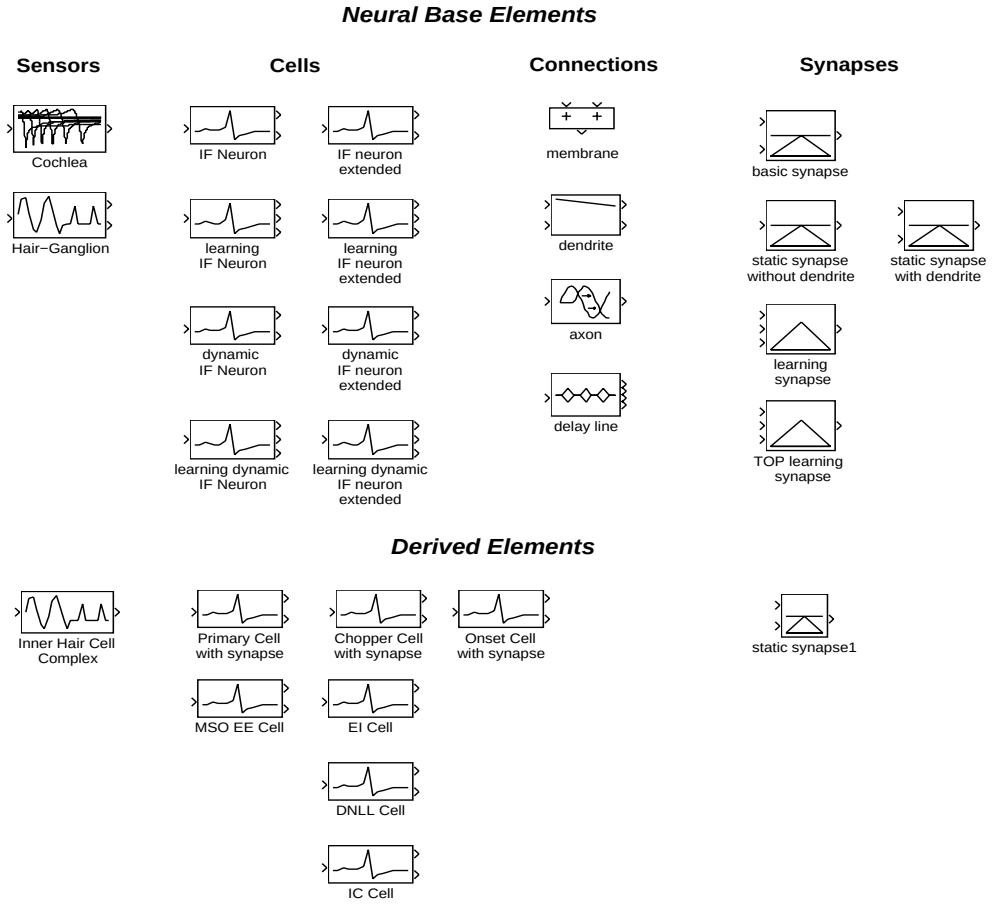

Figure 3.1: Neural Base Library - Overview

The upper panel of the Neural Base Library, shown in figure 3.1, displays the neural base elements themself, while the lower panel contains the different cell types derived from those base elements and applied in the proposed model architecture.

The neural base elements are grouped into four types: Sensors, Synapses, Cells and Connections. Each library element is represented by a single instance, visible as a graphical symbol, and contains two layers - a structural and a parametric layer. The structural layer defines the internal structure of each element common to all instances. Hence, whenever the structure of a library element is modified these changes become effective in all instances and derived elements within the model. This object oriented approach is very helpful in case of large models and enables the experimenter to modify for example the behavior of all dynamic IF neurons or of all DNLL cells by only a single modification of the library element.

On the contrary, the parametric layer is specific to each instance of the library element and enables the experimenter to set specific parameters to each cell or synapse or to a group of them contained in a subsystem as described later. Additionally, the parametric layer of all cell models contains a number of internal parameters possible to display through the separate "Internals" port. They can be chosen individually for each cell in order to access them during and after simulation. Figure 3.2 displays the structural and parametric layer of the library element "IF Neuron" for example.

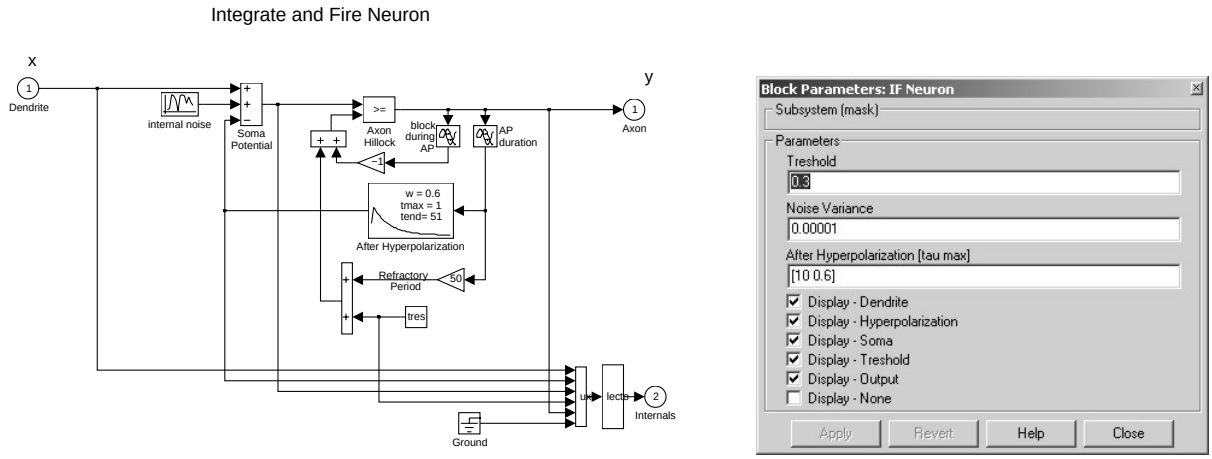

Figure 3.2: **left panel** - *structural layer of the Neual Base Library element IF Neuron*
right panel - *parametric layer of the IF Neuron Base Element.*

During the next subsections, the neural base elements, relevant to this model, will be introduced in detail. However, learning base elements have not been employed here and the reader might refer to other publications of the author like [ZITP97].

For the matter of understanding, we will start with synaptic elements and connections before introducing the basic cell models and the specific sensory elements.

3.1 Synapses

Within Spike Interaction Models and any other networks of Integrate and Fire neurons, the major task of synapses is, to generate graded and time continuous postsynaptic potentials (PSP) in response to presynaptic spikes arriving from the sending neuron. The specific shape of these PSP-functions $g(t)$ is essential to the performance of the whole network and differs with regard to 3 parameters:

- Rise time t_D- time needed to depolarize up to the maximum potential
- Maximum Potential w - maximal depolarization reached during one PSP - often referred to as synaptic weight
- Decay Time t_R - time needed to repolarize back to the resting potential

Furthermore, all three parameters are modified by dynamic processes in case of "dynamic synapses" and/or adaptive processes in case of "learning synapses".

3.1.1 Static Synapse

Here we will start with the simple static synapse, generating identical PSP in response to each spike of the sending neuron. The structural and parametric layer of the NBL element "Static Synapse" are shown in figure 3.3.

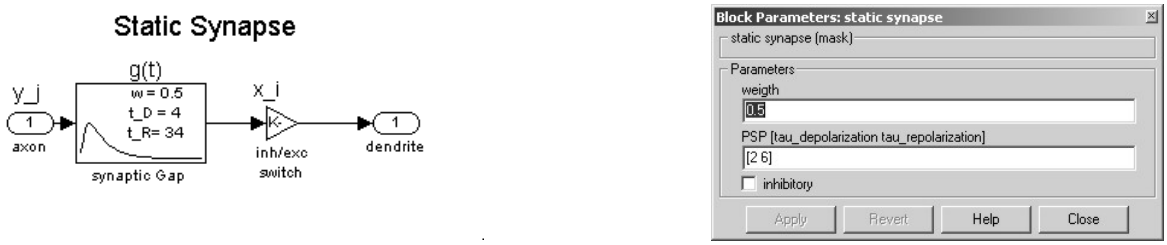

Figure 3.3: **left Panel** - *structural layer of the neural base library element static synapse*
right Panel - *parametric layer of the static synapse base element.*

The structural layer obviously contains only two elements, the synaptic transfer function $g(t)$, modeling the processes within the synaptic gap, and a switch to invert the synaptic response in case of inhibitory synapses. This results in the fairly simple transfer function:

$$x_i(t) = y_j(t) * g(t) \times inh \tag{3.1}$$

with $inh = 1$ for excitatory synapses, and $inh = -1$ for inhibitory synapses and $*$ representing the convolution between the arriving presynaptic spike $y_j(t)$ and the synaptic gap transfer function $g(t)$.
On the parametric level, the element contains tree parameters responsible to define the shape of the postsynaptic potential. Since this shape is very critical to the function of the entire network considerable effort has been taken, to design a model, flexible to incorporate the dynamic properties of the underlying complex electrical and chemical processes.

If a spike of the sending neuron reaches the presynaptic terminal, a specific transmitter substance is released into the synaptic gap, which in turn activates a number of voltage and chemical gated ion channels at the postsynaptic terminal. The resulting ion transfer, especially the Influx of Na+ ions, causes the postsynaptic membrane potential to raise. This process shall be named *Depolarization*. On the other hand, as soon as the membrane potential leaves it's resting value, an active ion transport is enabled, seeking to reestablish the resting state - this process shall be referred to as *Repolarization*. Since both processes counteract each other, their specific time constants and efficiencies will define the parameters of the resulting PSP at each synapse. However, this view onto Depolarization and Repolarization is still a summary on many separate processes on the level of Na+, K+, and Cl- ions and many different types of transmitters and channels. A fairly realistic synaptic model therefore needs to make all three PSP parameters independent and separately controllable.

A common technical model of the postsynaptic membrane potential is a capacitor element. The exponential charge function of such a capacitor with capacitance C is known as:

$$V(t) = V_0 + \frac{1}{C}\int I dt \tag{3.2}$$

and in the Laplace Domain this refers to a $pT1$ Element marked by the equation:

$$G(s) = \frac{K}{1+sT} \tag{3.3}$$

However, the usage of capacitors or pT1 Elements as synaptic models has three major drawbacks observable in figure 3.4.

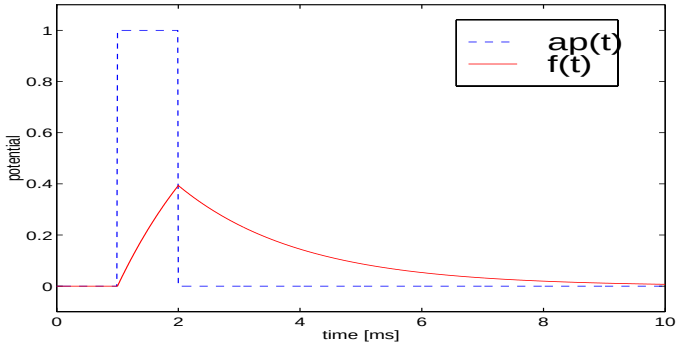

Figure 3.4: Response of a pT1 element with $\tau = 2ms$ and $K = 1$ to a presynaptic spike of 1 ms duration

1. Depolarization and Repolarization share the same time constant. Therefore rise time, maximum potential and decay time are not independently controllable.

2. The Depolarization process ends at the end of the presynaptic spike and here the maximum potential is reached. Hence, the physiological significant synaptic delay of 1-2 ms cannot be modeled and the maximum potential depends on the spike duration.

3. The response function is discontinuous causing discontinuities in case of spatio-temporal integration.

To overcome these limitations, Depolarization and Repolarization have been developed into separate models in the Laplace domain and rejoined to derive a more flexible synaptic transfer function $g(t)$.

Depolarization

Using the equivalent circuit diagram of figure 3.5, the membrane capacitance C_m is accompanied by a depolarization conductance g_d resulting in an charge current I_d trough both elements. Here the capacitance C_m accompanied by the conductance g_d can be best described in the Laplace domain by an Integrative Delay Element of the IT_n type. By the way, already Hodgkin and Huxley [HH52] suggested the usage of T_n Elements to model the membrane conductance. So the the IT_1 process:

$$G_D(s) = \frac{K_D}{s(1+sT_D)} \tag{3.4}$$

has been employed to model the Depolarization (or charge) process. In the time domain the impulse response of this process this would refer to:

$$g(t) = \mathcal{L}^{-1}\left\{1 * \frac{K_D}{s(1+sT_D)}\right\} = K_D * (1 - e^{-t/\tau_D}) \qquad with \quad \tau_D = C_m/g_d \tag{3.5}$$

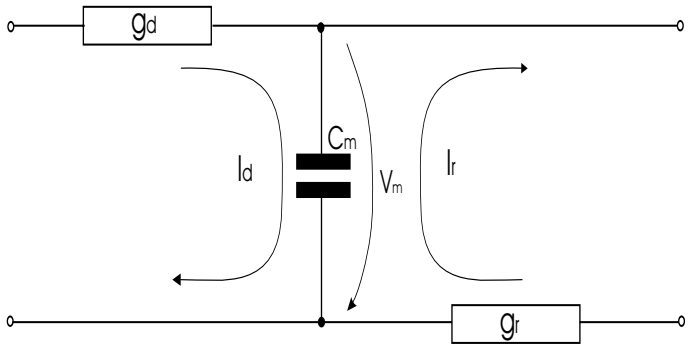

Figure 3.5: Equivalent circuit diagram for the postsynaptic membrane

Repolarization

To describe the outward flow of positive ions it is assumed that a discharge flow through the same membrane capacitance is modulated by the number of active ion pumps which in turn depends on the potential difference between actual membrane potential and resting potential. By usage of the same Equivalent Circuit Diagram (figure 3.5), the repolarization conductance g_r regulates the repolarization current I_r trough the same membrane conductance C_m, leading to a similar Laplace Element described by:

$$G_R(s) = \frac{K_R}{s(1+sT_R)} \tag{3.6}$$

The impulse response in the time domain will as well be defined by the repolarization time constant τ_R and the efficiency constant K:

$$g(t) = \mathcal{L}^{-1}\left\{1 * \frac{K_R}{s(1+sT_R)}\right\} = K_R * (1 - e^{-t/\tau_R}) \qquad with \quad \tau_R = C_m/g_r \tag{3.7}$$

Combination

Since the postsynaptic membrane potential after a single excitatory presynaptic spike is always depolarized and returns to the resting state after a limited amount of time, it can bee assumed that:

- The Efficiencies of the depolarization and the repolarization process are equal, leading to a complete abolishment of both effects after a certain amount of time. ($K_R = K_D$).
- The time constant of the repolarization process is alway larger than that of the depolarization process ($\tau_R > \tau_D$)

In the Laplace domain the combination of the two processes under these assumptions becomes a simple subtraction resulting in a third order transfer element of the $PIDT_2$ type as shown in equation 3.8

$$\begin{aligned} G(s) &= G_D(s) - G_R(s) = \frac{K}{s(1+sT_D)} - \frac{K}{s(1+sT_R)} \\ &= \frac{K(1+sT_R) - K(1+sT_D)}{s(1+sT_D)(1+sT_R)} = \frac{Ks(T_R - T_D)}{s(1+sT_D)(1+sT_R)} \end{aligned} \tag{3.8}$$

$$G(s) = \frac{Ks(T_R - T_D)}{T_D T_R s^3 + (T_D + T_R)s^2 + s} \tag{3.9}$$

After reformatting the denominator, equation 3.9 is derived as the Laplace domain equivalent of a third order differential equation. The spike response of such an element is shown in figure 3.6.

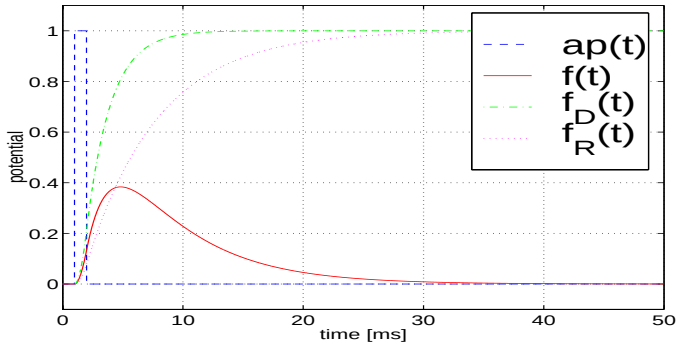

Figure 3.6: Spike response of a $PIDT_2$ Element as difference of two separate processes with the time constants $\tau_D = 2ms, \tau_R = 6ms$ and a joined efficiency of $K = 1$

To further simplify the synaptic simulation, canceling of s leads to a fairly simple PT_2 element with identical response behavior but only a second order differential equation to be solved during simulation.

$$\begin{aligned} G(s) &= \frac{K(T_R - T_D)}{(1+sT_D)(1+sT_R)} \\ G(s) &= \frac{K(T_R - T_D)}{T_D T_R s^2 + (T_D + T_R)s + 1} \end{aligned} \tag{3.10}$$

As can be seen in figure 3.6, the PSP function reaches it's maximum $2ms$ AFTER the end of the presynaptic spike and returns to the resting potential $5 \times \tau_R = 30ms$ after the presynaptic spike.

This way, synaptic delay as well as physiological causality is reproduced and the discontinuity of the time course is removed. Furthermore, depolarization is much faster than repolarization indicating the possibility to control both processes seperately.

The question is now, if the maximum potential can be controlled independently in this model. To achieve a control parameter for the maximum potential, we go back to the time domain and subtract the depolarization and repolarization process to:

$$\begin{aligned} g(t) &= K(1-e^{-t/\tau_D}) - K(1-e^{-t/\tau_R}) \\ &= K(e^{-t/\tau_R} - e^{-t/\tau_D}) \end{aligned} \tag{3.11}$$

The resulting synaptic transfer function is commonly known as α-function and frequently used in advanced integrate and fire networks. To find the maximum value of the α-function it's first derivative is set to zero resulting in equation 3.12 to define the point in time, when $g(t)$ reaches its maximum.

$$\begin{aligned} g'(t) &= K(-\frac{1}{\tau_R}e^{-t/\tau_R} + \frac{1}{\tau_D}e^{-t/\tau_D}) \\ 0 &= K(-\frac{1}{\tau_R}e^{-t/\tau_R} + \frac{1}{\tau_D}e^{-t/\tau_D}) \\ 0 &= -\frac{1}{\tau_R}e^{-\frac{t}{\tau_R}+\frac{t}{\tau_D}} + \frac{1}{\tau_D} \\ 0 &= -\frac{1}{\tau_R}e^{t\frac{\tau_R-\tau_D}{\tau_R\tau_D}} + \frac{1}{\tau_D} \\ \frac{\tau_R}{\tau_D} &= e^{t\frac{\tau_R-\tau_D}{\tau_R\tau_D}} \\ ln(\frac{\tau_R}{\tau_D}) &= t(\frac{\tau_R-\tau_D}{\tau_R\tau_D}) \\ t_{max} &= \frac{\tau_R\tau_D}{\tau_R-\tau_D} \times ln(\frac{\tau_R}{\tau_D}) \end{aligned} \tag{3.12}$$

Using this equation, the library element tells the experimenter the absolute rise time t_D for any parameter set chosen (see figure 3.7). Since the absolute value of $g(t)$ at $t = t_{max}$ shall be independently controllable accounting for the synaptic efficiency w, we will use equation 3.12 within equation 3.11 and derive the following relationship to calculate K with regard to a specific target value of w.

$$\begin{aligned} w = g(t_{max}) &= K(e^{-\frac{\tau_D\tau_R}{\tau_R(\tau_R-\tau_D)}\times ln(\frac{\tau_R}{\tau_D})} - e^{-\frac{\tau_D\tau_R}{\tau_D(\tau_R-\tau_D)}\times ln(\frac{\tau_R}{\tau_D})}) \\ \frac{w}{K} &= \left(e^{ln(\frac{\tau_R}{\tau_D})}\right)^{-\frac{\tau_D}{\tau_R-\tau_D}} - \left(e^{ln(\frac{\tau_R}{\tau_D})}\right)^{-\frac{\tau_R}{\tau_R-\tau_D}} \\ \frac{w}{K} &= \left(\frac{\tau_R}{\tau_D}\right)^{-\frac{\tau_D}{\tau_R-\tau_D}} - \left(\frac{\tau_R}{\tau_D}\right)^{-\frac{\tau_R}{\tau_R-\tau_D}} \\ K &= \frac{w}{\left(\frac{\tau_R}{\tau_D}\right)^{-\frac{\tau_D}{\tau_R-\tau_D}} - \left(\frac{\tau_R}{\tau_D}\right)^{-\frac{\tau_R}{\tau_R-\tau_D}}} \end{aligned} \tag{3.13}$$

Using this relationship to calculate the simulation parameter K in the background during the Initialization Phase of the Model, the Experimenter simply defines w and derives the exact maximum potential regardless of the time constants chosen.

Figure 3.7 displays the independence of the three parameters. In panel A, maximum value has been kept constant and only τ_D has been varied between 1 and 5 ms. As can be observed, all tree response functions reach the same maximum and disappear after 50 ms but their rise time can be controlled independently. Similar in Panel B the three response functions reach their maximum at the same time and vanish at the same point in time but exhibit significantly different maximum potentials w. Finally in Panel C, τ_R is the only parameter changed, while the other two are kept constant. Here the same Maximum is reached after the same time but decay time varies significantly.

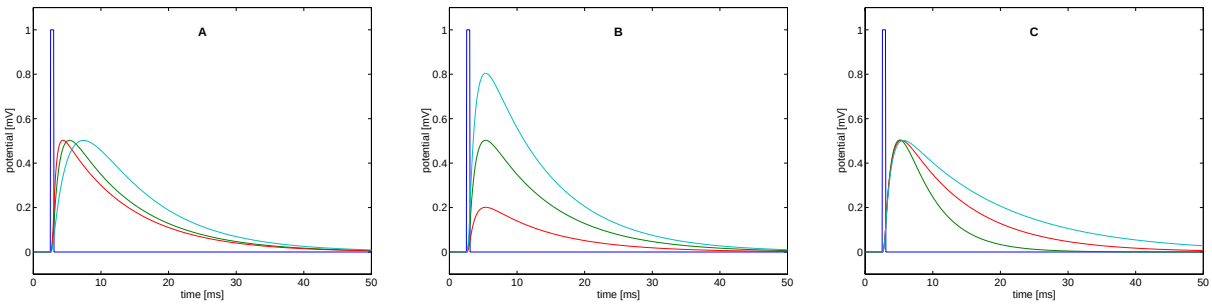

Figure 3.7:
panel A - *independent variation of Rise Time*
panel B - *independent variation of Maximum Potential*
panel C - *independent variation of Decay Time*

Since the PT_2 transfer function obviously fulfills the requirements formulated above, it has been employed to realize the synaptic gap model in the static synapse element. Out of the different options to simulate transfer functions within the SIMULINK environment, the standard block *Transfer Fcn* in the Laplace domain has been chosen and configured as shown in figure 3.8.

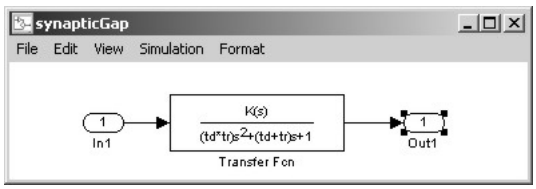

Figure 3.8: Transfer function in the Laplace domain employed to model the synaptic gap

During initialization, a specific mask calculates the parameters (K, td and tr) and displays the general shape as well as the key values within the icon as shown in figure 3.3. Since this procedure happens only once during initialization, it will not influence simulation time but help the experimenter to visually evaluate the parameter sets chosen.

So far, the synaptic response has been considered exclusively under the rare case of a singular spike. But natural cells and neural networks nearly always emit a number of spikes at specific places and specific points in time. A very important question during the modeling of synapses is therefore the temporal integration in case of several spikes arriving at the presynaptic terminal. In figure 3.9, the simple case is shown, when the synapse is excited by a spike chain of equal

distances. Here it can be seen that the membrane potential reaches a dynamic equilibrium after a number of spike responses adding to each other. In case of the static synapse the value of the equilibrium directly depends upon the spike rate (250 Hz in Panel A-C and 125 Hz in Panel D-F) and the three parameters of the PSP shape (varied within the vertical referring panels).

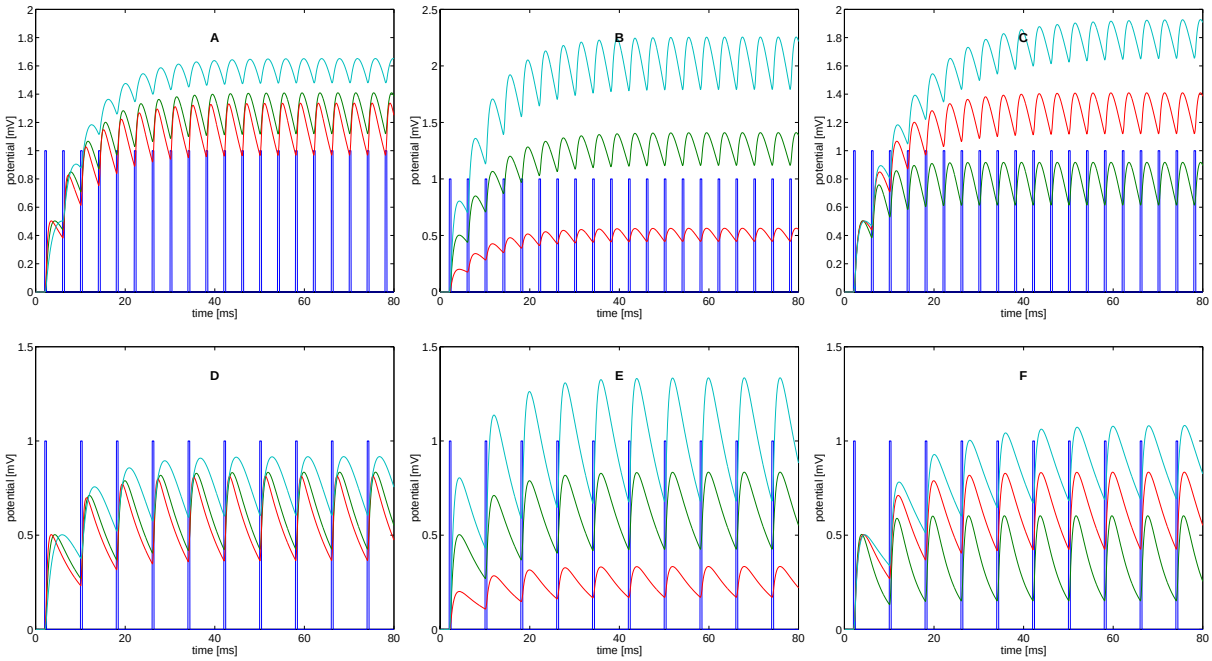

Figure 3.9:
Synaptic response to a spike frequency of 250 Hz with:
panel A - τ_D *varying between 1 ms (red) and 5 ms (cyan)*
panel B - w *varying between 0.2 ms (red) and 0.8 ms (cyan)*
panel C - τ_R *varying between 5 ms (green) and 15 ms (cyan)*
Synaptic response to a spike frequency of 125 Hz with:
panel D - τ_D *varying between 1 ms (red) and 5 ms (cyan),*
panel E - w *varying between 0.2 ms (red) and 0.8 ms (cyan),*
panel F - τ_R *varying between 5 ms (green) and 15 ms (cyan).*

Despite the different dependencies displayed in Figure 3.9, the major feature, also visible there, is the integrative nature of the PT_2 model. As observable, a subsequent PSP adds exactly to the level, which the sum of the previous PSP has reached at the time of the new spike. Therefore much higher membrane potentials are reached than a single PSP could produce. However, the possible membrane depolarization (or hyperpolarization in case of inhibitory synapses) is not unlimited. To include this natural limitation the library element “Dynamic Synapse“ has been developed and will be introduced in the next subsection.

3.1.2 Dynamic Synapse

As can be seen at the structural level of the NBL element "Dynamic Synapse", shown in figure 3.10, the basic transfer function $g(t)$ here is identical to the static model. But additional considerations have led to a subsequent channel function c(t) modeling the percentage of blocked ion channels after each spike.

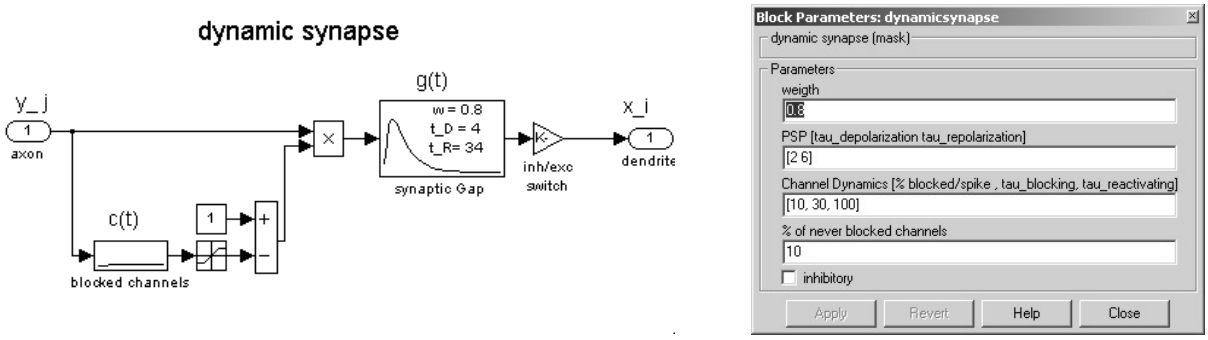

Figure 3.10:
left panel - *structural layer of the neural base library element dynamic synapse*
right panel - *parametric layer of the dynamic synapse base element.*

The assumption is, that each time a transmitter is released into the synaptic gap, a certain percentage of the available ion channels is activated, whereas their absolute number refers directly to the maximum postsynaptic potential generated. Directly after the spike, those channels are blocked and cannot be activated during a certain amount of time. Therefore, the percentage of blocked channels per spike as well as the time constants of channel blocking and channel reactivation are the new *channel dynamic parameters* added at the parametric level shown in the right panel of figure 3.10.
The functional result is that after the first spike, all channels are ready for activation and the PSP will take on the shape defined by $g(t)$. But after a number of spikes arriving within a short time frame, a high percentage of channels is already blocked and only few can still be activated resulting in a much lower maximum value of the PSP. In order to prevent this mechanism to fully block all channels, a limiter has been introduced to model. Du to the fact that the number of active channels will never become zero or negative, the experimenter is enabled to define a percentage of never blocked channels. At the parametric level this parameter is definable at the bottom of the mask.

To model the new function $c(t)$ of the dynamic synapses, the same PT_2 model as used for the synaptic gap has been employed once more. The argumentation behind is that the blocking and reactivation processes of the ion channels are very similar to the depolarization and repolarization of the membrane, since:

- both consist of competing processes canceling each other after a certain amount of time,
- both depend on the gaps between single presynaptic spikes - i.e. the spike frequency,
- both have different time constants for their sub-processes,

Considering the displayed structure of the dynamic synapse element show above, the transfer function of the element can be written as:

$$x_i(t) = \{[1 - y_j(t) * c(t)] \times y_j(t)\} * g(t) \quad \times inh \tag{3.14}$$

As can be imagined, the differential equation describing this function is already quite complex, but the object based approach of the NBL enables the experimenter to modify functions like this with very simple graphical changes. In case of the dynamic synapse, this would be surely possible by introducing separate models for ion and channel types. However, the dynamic properties shown below, proved sufficient for the model developed during this study and since the synaptic model usually has very many instances any additional computational effort influences to overall simulation time.

In order to display the dynamic response properties of dynamic synapses figure 3.11 displays their response under several significant conditions. Here, the parameters of the model element have been set to the following values: $w = 0.8$, $\tau_D = 2ms$, $\tau_R = 6ms$, % of blocked channels/spike = 5% , $\tau_{blocking} = 10ms$, $\tau_{reactivating} = 70ms$ and % of never blocked channels = 10%.

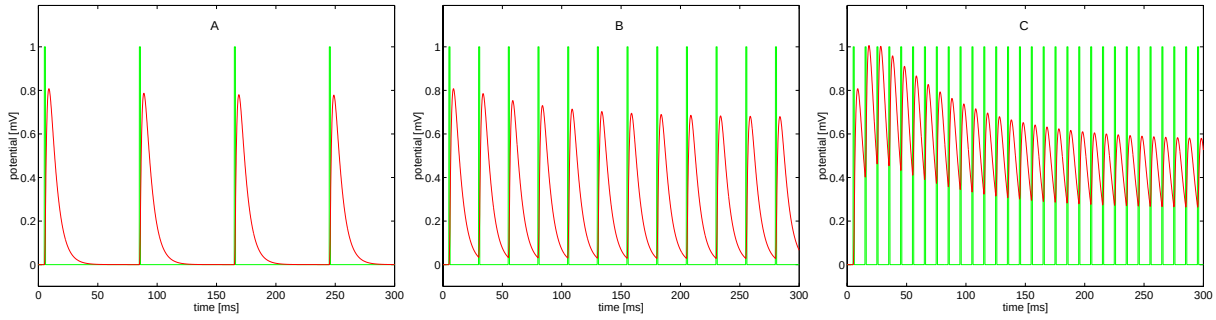

Figure 3.11:
panel A - *dynamic response to a spike frequency of 12.5 Hz*
panel B - *synaptic response to a spike frequency of 50 Hz*
panel C - *dynamic response to a spike frequency of 100 Hz*

In *panel A* the special case is shown, when the distance of the arriving spikes exceeds the reactivation time constant, hence all ion channels are again available and the PSP response to each spike is uniform.
If the distance between subsequent spikes decreases, as shown *panel B*, the response to the second spike is already diminished due to the limited channels available. Since the Spike distance of $20ms$ still exceeds the repolarization time constant τ_R there is no temporal integration and the membrane potential does not rise above the maximum value of a single PSP.
However, in *panel C*, both time constants are under-run and the natural behavior of dynamic synapses becomes prominent. During the first three spikes, the membrane potential is temporal integrated and rises above the maximum value of $g(t)$, but soon thereafter, much of the channels are blocked and the integrated individual responses to each spike are significant smaller, resulting in a dynamic equilibrium far below the one observed in case of the static synapse. If one now imagines that this synapse projects its membrane potential trough the dendritic tree toward the receiving neuron, a basic feature of cell responses throughout the nervous systems

can be explained. This means that most neurons respond to the onset of new signals or changing signal parameters with a burst of spikes (phasic response) and return to a much lower firing rate if the signal remains stable (tonic response). The dynamic character of synaptic transmission might be one of the major reasons for this and should not be neglected by dynamic neural models.

In figure 3.11 the dynamic response to a absolutely invariant signal leads at least under the conditions of panel C to a membrane potential changing significantly over time. In this case, only the spike frequency has been modified to display this feature but the parameters of channel dynamic enable the experimenter also to define these properties. The influence of each parameter is shown in figure 3.12. For clarity of the display, the presynaptic spikes are not shown but they were kept at a constant frequency of 200 Hz.

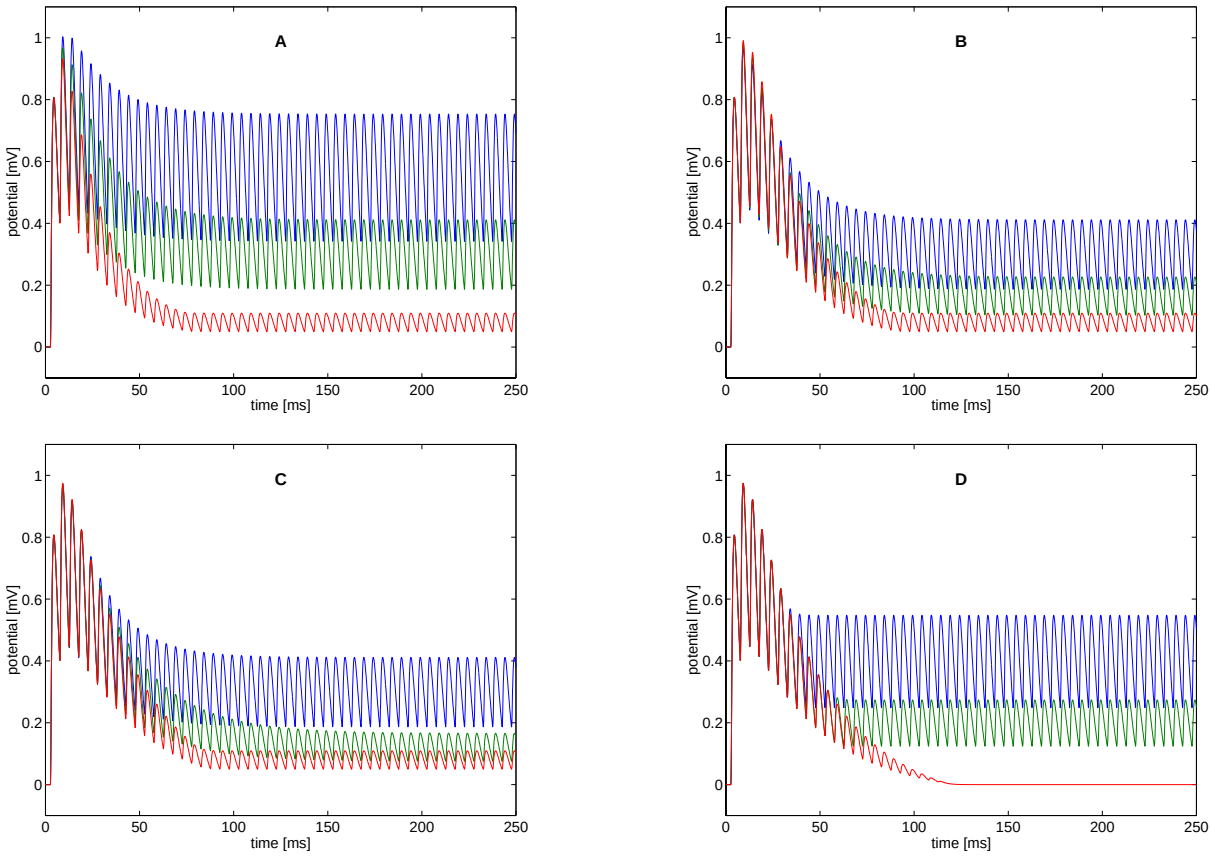

Figure 3.12:
Influence of the different parameters of dynamic synapses on their response pattern to a pulse frequency of 100 Hz **panel A** - *parameter % of blocked channels - values of 5% (blue), 10%(green) and 15% (red),*
panel B - *parameter* $\tau_{blocking}$ *- values of 10 ms (blue), 20 ms (green) and 30 ms (red),*
panel C - *parameter* $\tau_{reactivation}$ *- values of 40 ms (blue), 60 (green) and 80 ms (red),*
panel D - *parameter % of never blocked channels - values of 50% (blue), 25% (green) and 0% (red)*

As can be seen in panel A, the more channels are blocked after each spike, the lower is the value of the dynamic equilibrium. The influence of the two time constants, varied in panel B and C, is less significant but strongly influences the upper limit of the inter-spike-distance when the first

decay in the dynamic response happens (as shown in panel A of figure 3.11). A strong influence is observable when varying the percentage of never blocked channels. Here in the extreme case of 0% after roughly 100 ms the synapse doesn't react anymore, despite continuing presynaptic spikes, and will only restart after a significant gap.

Dynamic synapses are very helpful elements to duplicate the inherent dynamics of synaptic transmission but should be used with care. Due to their dynamic response properties they will alter the spike time as well as the firing rate of the receiving neuron compared to the sending neuron. Since at many stages of the auditory brainstem a precise spike timing and the preservation of phase locking is essential to decode interaural disparities, dynamic synapses have been used in the proposed architecture exclusively at the level of the inferior colliculus.

3.2 Interneural Connections

This second group of base elements is concerned to model the connections between the synapses and the cell body of the receiving neural cell - i.e. the dendritic tree - and to model the connections between the soma of a sending cell and the synapses it projects to - i.e. the axon.
As commonly known, the major difference between both structures is that the dendritic propagation uses graded potentials overlapping in time and space and containing a huge structural variety. By contrast axonal structures have less branches and propagate actively reproduced Action Potentials (AP) traveling along the axon with nearly identical shape.

3.2.1 Dendritic Model

In order to model dendritic processing two elements have been added to the Neural Base Library. First, and most importantly, the membrane model and secondly the dendritic leakage model. Both elements are shown in a typical setup in figure 3.13.

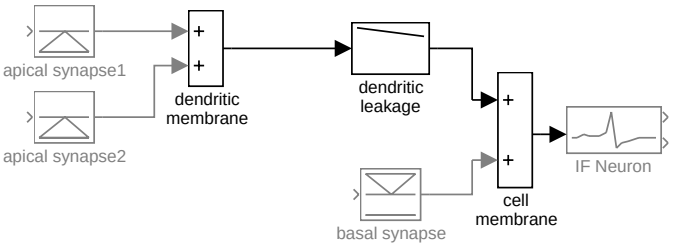

Figure 3.13: Simple model of a dendritic tree containing the library elements membrane and dendritic leakage

Although the membrane element is a very simple element, with no parameters and a simple summing structure, it's function is very important, since it realizes the spatial integration within the dendritic tree. The reason behind is that any neuron contains many synapses, receiving postsynaptic potentials from many different cells and nuclei. While every synapse performs a temporal

integration on its own, the resulting soma potential strongly depends on the interaction between the PSPs generated by the different synapses. Here, the membrane element integrates the EPSP and IPSP of all those synapses within the branches of the dendritic tree (dendritic membrane) as well as at the soma (cell membrane). Mathematically it is a simple linear summing element.

The dendritic leakage element is a bit more advanced and models the leakage current flowing through the dendritic membrane, diminishing the effect of apical synapses (synapses relatively far from the soma) compared to basal synapses (synapses close to the cell body). It's structural and parameter level are shown in figure 3.14.

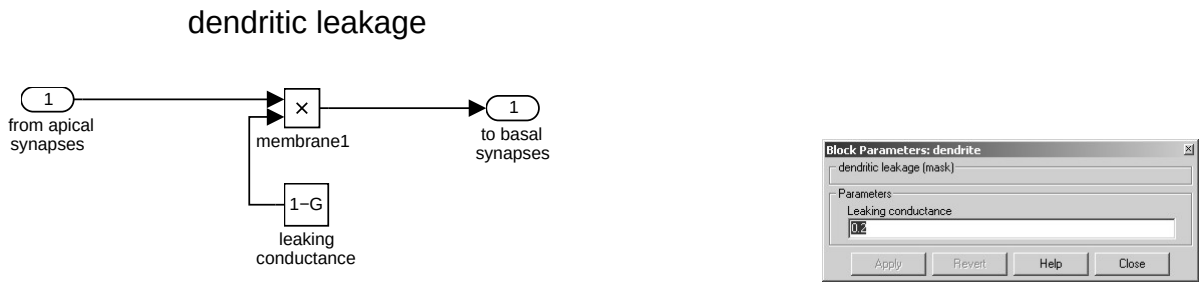

Figure 3.14:
left panel - *structural layer of the neural base library element dendritic leakage*
right panel - *parametric layer of the dendritic leakage base element.*

The basic structural element of this model is a leaking conductance $(1 - G)$ accessible from the outside and defining the percentage by which the postsynaptic potentials of apical synapses are diminished by the time they arrive at the site of the cell body.

Based upon these two dendritic elements, a huge variety of spatio-temporal dynamics within the dendrite can be generated. To give at least a rough impression, figure 3.15 provides a simple example, employing the dendritic structure shown in figure 3.13.
Here, panel A displays the spatial integration of two temporally integrated postsynaptic potentials generated by different synapses receiving input from different cells. Panel B shows the effect of the dendritic leakage element - here set to 20% leakage current. And panel C finally shows the resulting complex dynamic potential after integrating the IPSP of the basal inhibitory synapse.

3.2.2 Axonal Model

It has been mentioned before that axonal transmission differs from dendritic transmission by the use of active Action Potentials (AP). At every node of Ranvier, a complex mechanism of ion channels and pumps generates an AP with nearly identical shape. Since the axon is covered by Myelin in between the node of Ranvier, the AP seems to jump from one node to the next along the axon. This very rapid and nearly loss-free transmission process is used by the nervous system to cross even large distances between different brain areals but also between sensors, actors and nerve cells throughout the body.

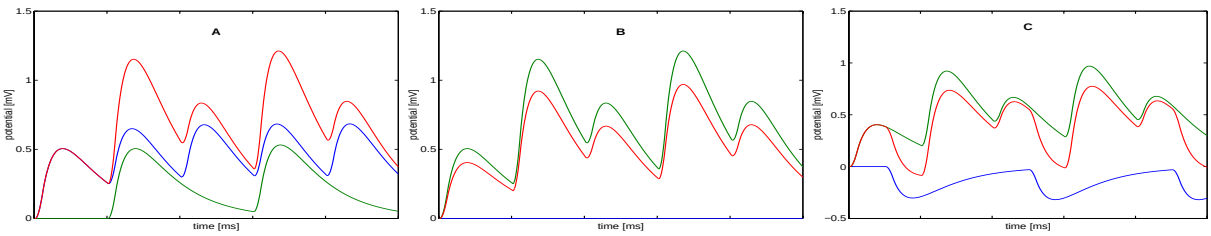

Figure 3.15:
panel A - *spatial integration of two PSP (blue and green) at the dendritic membrane, red = resulting potential*
panel B - *dendritic leakage between apical dendrite (green) and basal dendrite (red) = potential at soma*
panel C - *spatial integration (red) between apical EPSP's (green) and basal IPSP (blue)*

In order to model axonal transmission, AP-pulses generated by the cell models propagate along the lines of the model and reach one ore more synapses without loss. However, there is a timing issue. Even the natural rapid transmission will need a certain amount of time to cross larger distances, for example between the superior olivary complex and the inferior colliculus. The base element "Axonal Delay" therefore introduces a specific time delay to model the time a AP needs to reach it's target. This parameter is accessible from the parametric layer and can be set to any value between $10\mu s$ and several ms. Typical axonal delays in the auditory brainstem range from 10 to 200 μs.

Under natural conditions, a single axon makes many contacts with dendritic synapses of different cells and at different distances from its site of origin. To simplify the model structure, the base element "Delay Line" has been created by combining three axonal delays into one simulation block. The resulting internal structure is shown in Figure 3.16. The only parameter of both axonal elements is the unit delay, identically applying to each axonal delay block.

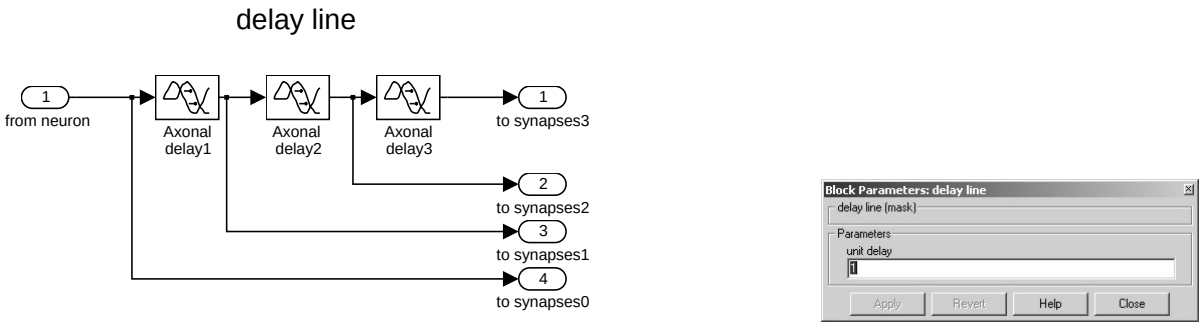

Figure 3.16:
left panel - *structural layer of the neural base library element delayline*
right panel - *parametric layer of the delayline and axonal dely base elements.*

The principal effect of axonal delays is displayed in figure 3.17. Here it becomes observable that a single action potential emitted from the sending neuron reaches its targets at different points in time - in this case after 2, 4 and 6 ms.

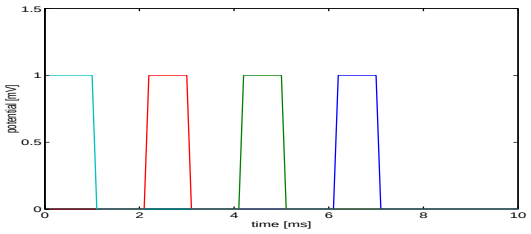

Figure 3.17: delayed arrival of an Action Potential after passing a delay line of three units

Finally, there shall be a word with regard to dendritic delays. In principle, dendritic delays are as well possible as axonal ones and the axonal delay block can also be used to introduce delays into the graded dendritic propagation. However, the effect will be the same, if the AP (at the presynaptic side of the synapse) is postponed or if the PSP (on the postsynaptic side of the synapse) is delayed. By convention, in this model architecture delays have always been applied to the axonal (or presynaptic) terminal. But this is only by convention and does not limit the experimenter to use the axonal delay block to model specific dendritic delays. It is capable not only to delay pulses but also graded potentials.

3.3 Cell Models

Out of the several cell models contained in the Neural Base Library only the non-learning *IF Neuron* and *dynamic IF Neuron* including their extended versions will be introduced here. For the learning models one might again refer to other publications like [ZITP97],[ZIT97].

3.3.1 IF Neuron

The general principle of Integrate and Fire (IF)-Neurons has been introduced already during chapter 1 of this thesis. It's basic character is the spatio-temporal integration of graded potentials resulting in the emittance of a uniform pulse in case of exceeding a defined firing threshold. This simple principle with some additional features is realized within the model of the "IF Neuron" displayed in figure 3.18.

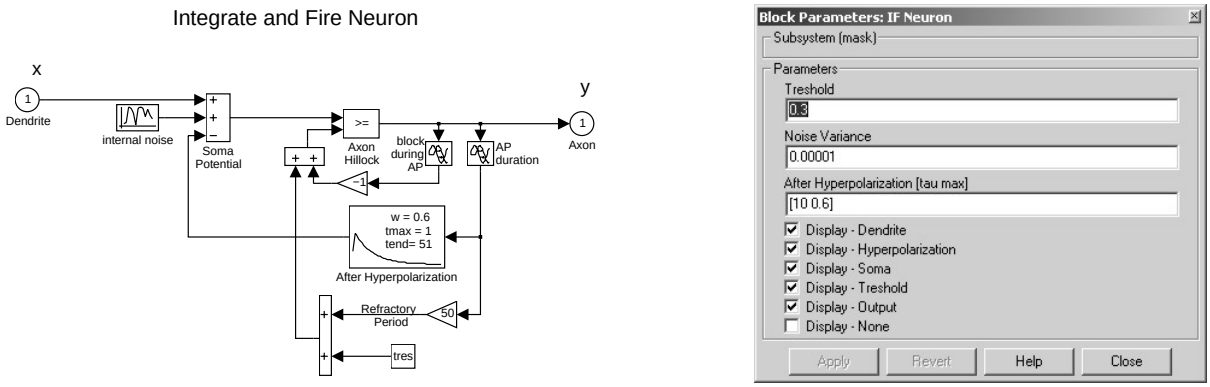

Figure 3.18: **left Panel** - *structural layer of the neural base library element IF Neuron*
right Panel - *parametric layer of the IF Neuron base element.*

As can bee seen in the upper left corner of the IF Neuron structure, the summation of the different dendritic PSP $x_i(t)$, directly feeds into the first major element of the model structure - the soma potential. It represents the membrane potential of the neuron's cell body. Besides the spatio-temporal summation of the external dendritic potentials, two internal potentials add to the value of the soma potential resulting in the following equation:

$$s(t) = \sum_{i=1}^{n} x_i(t) + \varepsilon(t) - ahp(t - t_f) \tag{3.15}$$

First, there is a internal noise term $\varepsilon(t)$, representing the fact that cell potentials undergo statistical changes and are not fully deterministic. The noise term represents this natural phenomena and can be set at the parametric level from outside. Internal noise is especially critical for networks with adapting (learning) behavior. For most cells of the proposed architecture, noise variance has been set to very small values like 0.00001 in order to achieve a fairly deterministic cell response. However, statistical effects can be used when increasing this variable.

The second internal potential $ahp(t)$ adding with negative sign to the soma potential is the so called *After Hyperpolarization Potential (AHP)*. The AHP, already known from the Spike Response Models of Gerstner [GV94] [GK02] and van Hemmen [VGH+90] [VDC02], models the membrane dynamic directly after a spike has been emitted by the cell. Here, the term $(t - t_f)$ describes the time passed since the last firing time (t_f) of the neuron. It is the very nature of the ionic currents involved in the generation of Action Potentials that after each AP the membrane of the entire cell is shortly hyperpolarized below the resting value at $-55mV$ and recovers with time constants between $100\mu s$ and $10ms$. To model the AHP, again the PT_2 element, known from the static synapse, has been employed but only the maximum value w and the repolarization time constant τ_R have been made accessible in the parametric layer. The rise time of the AHP Element is always set to $100\mu s$. The functional effect of the AHP is that during the AHP period the cell potential is internally diminished causing a decreased firing probability. In other words, more synaptic current is necessary to make the cell fire again shortly after it has generated an AP. The effects of all three parts on the soma potential are displayed in principle in figure 3.19, where the time course of PSP's is overlapped by some internal noise and the effect of an AHP resulting from a previous spike.

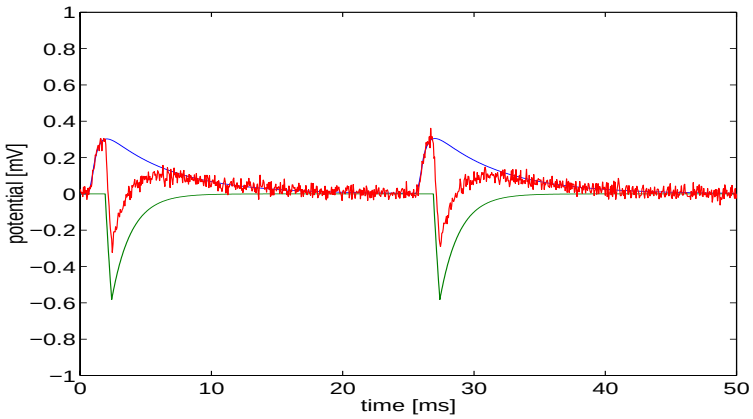

Figure 3.19: Soma potential (red) as linear summation of PSP's (blue), internal noise (varaince = 0.0002) and AHP effects (green).

The second major element of the IF Neuron structure is the threshold function modeling the processes at the axon hillock of the nerve cell. As can be observed in figure 3.20, the soma potential is here compared to a threshold value. The reason behind is that the active process of AP generation is triggered if the soma potential within the cell exceeds a certain value, usually referred to as firing threshold. If this value is exceeded only for a very short time, a self inducing process will generate an Action Potential of uniform shape in all cases. To ensure this deterministic nature of AP generation, beside the comparative element described as *axon hillock*, another element named *block during AP* is added to the model. Accompanied by a delay element (set to $10\mu s$) to prevent an algebraic loop, this AP block ensures that the threshold is always exceeded during the duration of one AP and the AP pulse is not cut of, even when the calculated noisy soma potential falls below the threshold.

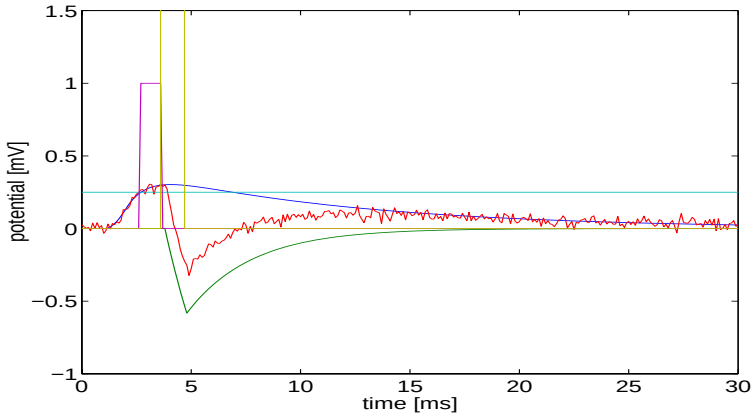

Figure 3.20: As result of a presynaptic potential (blue) the IF Neuron generates a uniform spike (magenta) if the soma potential (red) exceeds the threshold (cyan) and is followed by an absolute (yellow) and relative (green) refractory period.

Due to the nature of AP generation, every pulse needs to have the same duration as well as the same height. To ensure identical height, a uniform pulse height of 1 ($= 100mV$) has been chosen and is realized by the hillock comparator, as it returns boolean values of 0 or 1. To ensure a uniform pulse width, the AP blocking element prevents the pulse from cutting of early and the *AP duration* element ensures a cut of after exactly $100\mu s$ (1 model time step). The realization is again a simple delay element set to $100\mu s$ accompanied by a high gain value named *Refractory Period*. Each time the soma potential exceeds threshold, the AP duration element will hold the output high for $100\mu s$ and will afterwards assure that the threshold is lifted to a very high value, causing the hillock element to immediately return to 0 and to end the AP pulse. By this feed back loop, a uniform width of every AP is assured. However, the refractory period element has also a physiological correlate. Directly after the emittance of an AP, every cell is fully blocked and cannot fire for about as long as the AP lasted. This period is often referred to as *absolute refractory period*, while the AHP period in this case marks the *relative refractory period*. Since these terms are often mentioned but have been hardly assessed in a quantitative way, a value of $100\mu s$ for the absolute refractory period has been chosen. Depending on the AHP parameters, time constants of up to $10ms$ for the relative refractory period, have been used for the cell models of the architecture proposed in this study.

Finally, there is a fourth element adding to the threshold value, used for comparison at the axon hillock. It is simply named *threshold* and marks the internal threshold potential at which the AP generation process is triggered. It is freely definable at the parametric layer and is a simple constant in case of the non-dynamic IF neuron.

Having described the specific elements of the IF Neuron model, it's overall transfer function can be written as:

$$y(t) = \eta(s(t), tres(t)); \quad \text{with} \quad \begin{aligned} s(t) &= \sum_{i=1}^{n} x_i(t) + \varepsilon(t) - ahp(t - t_f) \\ tres(t) &= tres + ref(t) - block(t) \end{aligned} \tag{3.16}$$

Here $\eta(s, tres)$ describes the discontinuous threshold function returning 1 for $s >= tres$ and 0 for any other case, $s(t)$ represents the soma potential and has been already described as summation of dendritic PSP's - $x(t)$, the internal noise - $\varepsilon(t)$ and the after-hyperpolarization - $ahp(t)$. $tres(t)$ describes the comparative threshold value, containing the threshold constant - $tres$, the refractory period $ref(t)$ (which is 50 for $t_f =< t =< t_f + 100\mu s$ and 0 elsewhere) and the blocking function $block(t)$ (with $block(t) = -1$ for $t_f =< t < t_f + 10\mu s$).

As shown in figure 3.2, the principle parameters like threshold constant ($tres$), noise variance (ε), AHP maximum and decay time constant (ahp) are definable for each cell from the parametric level. Beside these parameters, the parametric level allows the experimenter to access the time course of different internal cell parameters via the "internals" port of the IF neuron base element. They can be selected by specific check boxes on the parametric level and are than propagated and selected via the internal structural elements shown at the bottom of the block structure.

If all boxes are checked, one can watch them during simulation on a scope-display element provided by the SIMULINK standard library. A typical example is shown in figure 3.21.

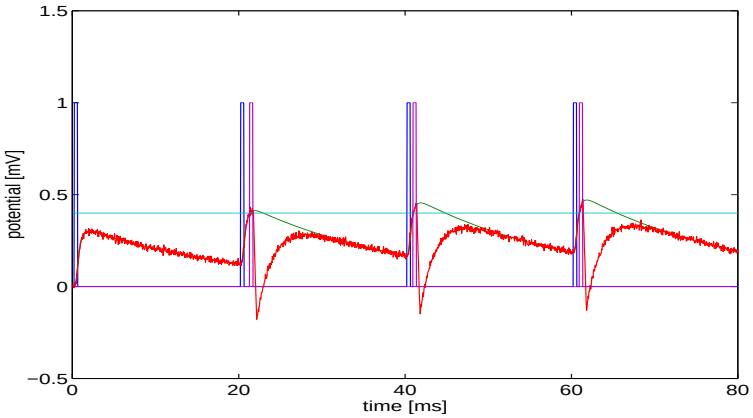

Figure 3.21: internal cell parameters of an IF neuron receiving EPSP from a single synapse: **blue**-*presynaptic spikes,* **green**- *resulting dendritic potential,* **red**- *soma potential including noise and after hyperpolarization,* **cyan**- *internal firing threshold,* **magenta** - *emitted action potentials*

Here, the presynaptic spike is additionally displayed (blue) to show the causality of the dendritic potential (green) resulting in this case from a single synapse. As can be seen, the soma potential

after the first spike, does not reach the internal firing threshold and no AP is emitted. Hence, the soma potential follows the dendritic potential with some noise. After the second presynaptic spike, the temporal integration of PSP's reaches the firing threshold and an AP is generated about $2ms$ after the presynaptic spike. This AP, in turn generates an After-Hyperpolarization, diminishing the soma potential for a certain period. In this case, the third presynaptic spike arrives after the AHP is vanished and the cell fires again. Since the soma potential, at this time, was already closer to threshold due to further temporal PSP integration, the temporal delay between presynaptic spike and the AP spike emitted shortens to $1ms$. However, the AP spike will always result from the PSP and therefore follow the presynaptic spike with some delay, even if the threshold would be lowered further. I.e. the causality and internal delay of cellular spike transmission is guaranteed under any condition.

3.3.2 Extended IF Neuron

A special version of the IF neuron is represented by the Base Element *Extended IF Neuron*. As shown in figure 3.22, it duplicates the IF neuron but includes a single synaptic transmission model at the dendritic entrance.

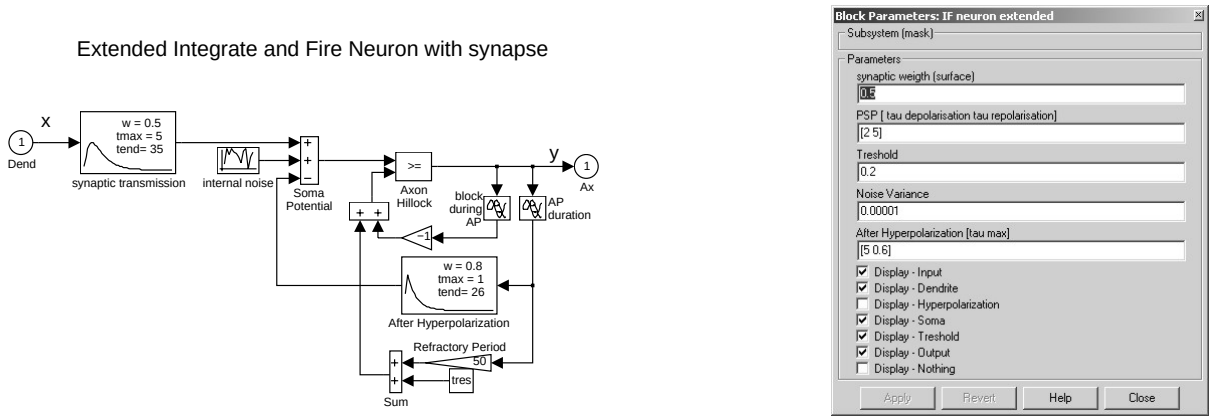

Figure 3.22: **left panel** - *structural layer of the neural base library element Extended IF Neuron* **right panel** - *parametric layer of the Extended IF Neuron base element.*

This element is particularly helpful if all synapses, projecting toward a specific neuron, can be assumed to act identical or if a one to one connection is likely. In this case, the number of necessary simulation elements can be significantly reduced if the synaptic transfer function is included in the cell model and the presynaptic spikes are feeded directly into the cell model. Although somewhat un-physiological, its effect will be identical to several separate but identical synapses projecting their PSP toward this cell. Logically, the synaptic transmission parameters *synaptic weight* w, τ_D and τ_R are now part of the cells parametric layer and can be set from here identical for all virtually assumed feeding synapses. However, special attention should be drawn to the fact that the extended IF Neuron element does not allow the construction of compartmental dendritic models (like apical or basal synapses) and prevents the experimenter from using different PSP shapes and/or dynamics. However, at some stages of the proposed model direct one-to-one connections have been established using this element.

3.3.3 Dynamic IF Neuron

For the simple case of a single synaptic input, the IF Neuron model described above, has proven to be fairly sufficient. However, this is rarely the case. In real biological systems, most cells receive quite different synaptic inputs. If many PSP's are integrated in the soma potential, the cell reaches it's maximum firing rate very fast, limited only by the absolute and relative refractory periods (see figure 3.24 Panel A).
Since this is not the physiological response of a nerve cell, another cell model has been developed and named "Dynamic IF Neuron". Under similar assumptions as in the case of dynamic synapses, an additional filter element has been added to this structure, modeling the limited resources and firing capabilities of a natural neuron.

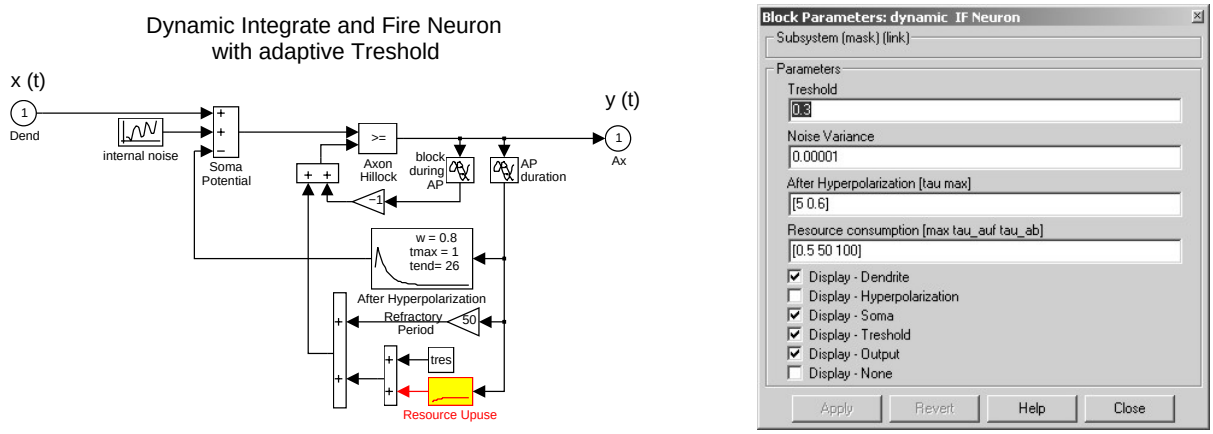

Figure 3.23: **left panel** - *structural layer of the neural base library element Dynamic IF Neuron* **right panel** - *parametric layer of the Dynamic IF Neuron base element.*

Enhanced in figure 3.23 this additional element is positioned within the feedback loop of the threshold function and is triggered by every AP emitted. This results in an adaptive threshold represented by the term $res(t - t_f)$ in the threshold function of equation 3.17 now written as:

$$(3.17)\quad y(t) = \eta(s(t), tres(t)); \quad \text{with} \quad \begin{aligned} s(t) &= \sum_{i=1}^{n} x_i(t) + \varepsilon(t) - ahp(t - t_f) \\ tres(t) &= tres + ref(t) - block(t) + res(t - t_f) \end{aligned}$$

The parameters of this PT_2 element are controllable from the parametric level and contain again the maximum value max the resource-blocking -time-constant τ_{auf} and the resource-decay-time-constant τ_{ab}. Through these parameters, the experimenter can now control which maximum firing rate is reached after which time. By modification of the resource parameters, the cell's response to an identical activation pattern can be modified between a strictly onset response (see figure 3.24 - panel D), where only a single burst is emitted at the onset of the signal, and from there continuously across *Pauser* response (panel C) and *Primary* response (panel B) toward the *Chopper* response, shown in panel E. These response types will be introduced in more detail in the next chapter and shall only be mentioned by example at this point.

The dynamic model behind the resource element, is fairly simple and goes back to that channel

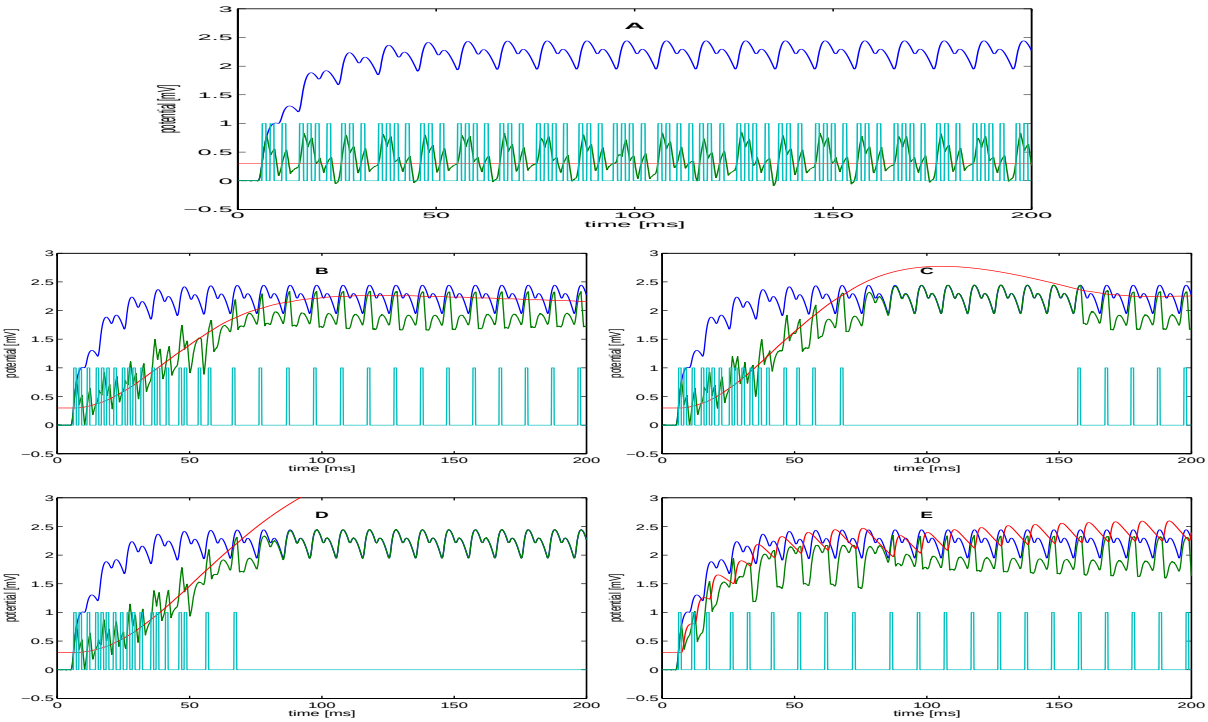

Figure 3.24:
panel A - *Response of the static IF Neuron to high multi-synaptic excitation for reference, blue - dendritic potential; green - soma potential; red - internal firing threshold; cyan - emitted AP-spikes*
panel B - *Primary-like response of a dynamic IF neuron to the same input with dynamic parameters set to* $max = 0.1; tau - auf = 30; tau - ab = 100$
panel C - *Pauser response of the dynamic IF neuron to the same input with dynamic parameters set to* $max = 0.15; tau - auf = 50; tau - ab = 100$
panel D - *Onset response of the dynamic IF neuron to the same input with dynamic parameters set to* $max = 0.2; tau - auf = 100; tau - ab = 200$
panel E - *Chopper response of the dynamic IF neuron to the same input with dynamic parameters set to* $max = 0.5; tau - auf = 1; tau - ab = 40$

resource theory. The more often a specific cell has already fired within a given time frame, the more ion channels are blocked and cannot be activated until after their recovery period is over. The resulting effect is that after every AP, the threshold to be exceeded in order to generate another AP is lifted by a small amount. During the first *ms* this will not influence the firing behavior much, but as the emitted AP's become more dense, the threshold value will significantly increase and slow down the persistent firing rate. As can be seen in panel D of figure 3.24, it reaches a high level equilibrium, where the firing rate depends mainly on the resource-reactivation-time-constant τ_{ab}.

The dynamic IF neuron has been used in the proposed architecture, to model the specific cells at the lateral superior olive (LSO) as well as within the dorsal nucleus of the lateral lemniscus (DNLL) and the inferior colliculus (IC). The referring cell types in the *Derived Elements* section of the Neural Base Library have been derived from this base element with their dynamic parameters set to specific values as displayed in the next chapter.

3.3.4 Extended Dynamic IF Neuron

Last but not least, the fourth non-learning cell model shall be introduced here. As one might imagine it combines the extended IF neuron with the dynamic one and is therefore called "Extended Dynamic IF neuron". It's structure and internal equations are identical to the dynamic IF Neuron except that a synaptic transmission model is again added at the dendritic entrance. The reasons are mainly of computational nature and similar to those, mentioned in case of the extended IF neuron. The internal structure and parametric level shown in figure 3.25, might therefore not surprise.

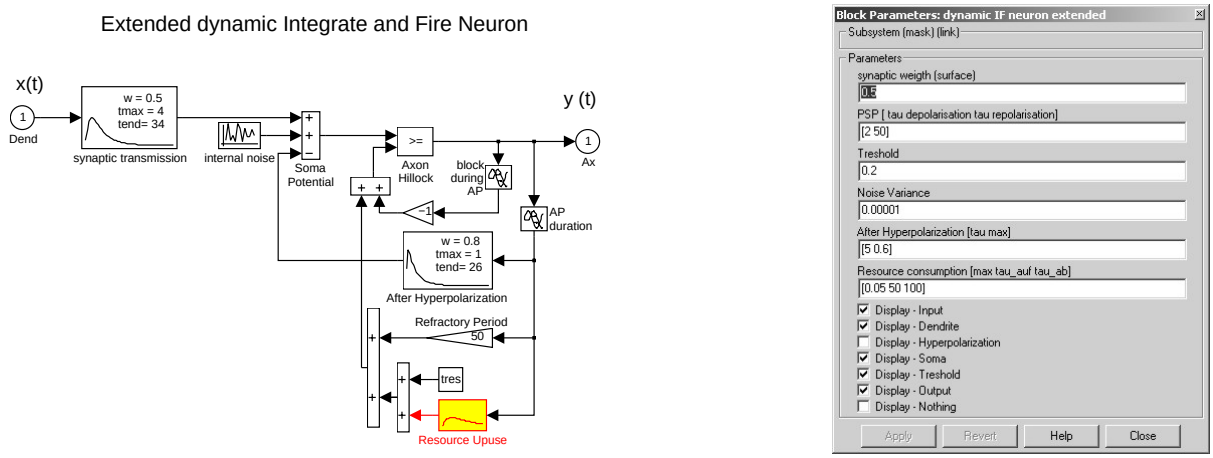

Figure 3.25:
left panel - *structural layer of the neural base library element Extended Dynamic IF Neuron*
right panel - *parametric layer of the Extended Dynamic IF Neuron base element.*

This base element was especially used to derive the three cochlear nucleus (CN) cell types of the derived elements section. *Primary-, Copper- as well as Onset-Cells* are necessarily of dynamic nature and since the model assumes a direct one-to-one feeding of the CN cells by the auditory nerve fibers, the single synapses have been included into the cell model to clarify the model structure and decrease the computational effort.

3.4 Sensors

The fourth and last group of neural base elements shall be mentioned only briefly at this point, since the motivation behind them as well as their detailed internal structure will be discussed during the introduction of the inner ear model at the next chapter. However, the structural and parametric levels of the two auditory sensory elements used in this architecture shall be introduced here for reasons of completeness.

3.4.1 Cochlea

The "Cochlea" base element realizes a specific filter cascade of 2'nd order filters designed to model 16 different locations along the basilar membrane. Since the cochlea as a whole as well as the elongation of the basilar membrane is a purely mechanical mechanism, there is no neural correlate to this element. However, a detailed and timely exact model of this mechanical structure is essential to decode the frequency parameters as well as interaural disparities of the acoustical signal arriving at the two ears.
To realize this filter cascade the Cochlea base element makes use of a specific type of MATLAB function optimized to use with MATLAB/SIMULINK called *s-function*. Following a predefined I/O syntax, s-functions can contain any type of mathematical functions written in MATLAB code. The s-function used here is called *fltcasc* and can bee seen at the parametric level displayed in figure 3.26. The second parameter of the s-function is a variable named *flt16441*, which contains the 5 filter coefficients for each of the 16 frequency filters used in this model.

Figure 3.26: **left panel** - *structural layer of the neural base library element Cochlea*
right panel - *parametric layer of the Cochlea base element.*

The *fltcasc* s-function in principal could load any filter file and therefore realize any number of filters. In this model 16 filters for each hemisphere have been chosen, to achieve a sufficient frequency resolution, while not running into computational problems due to the increasing number of cells and synapses, in case of more frequency channels.

Due to some hidden calculation and loading processes happening during the model initialization phase in the background, the experimenter will not need to care for the s-function nor the filter file. The only prerequisite is that both files reside in the NBL folder included in the MATLAB path, as mentioned above. Noticeable at this point is the fact that one should name the s-functions differently if the Cochlea element is used more than once. Unfortunately, the object oriented library approach does not hold for s-functions so far and therefore, the separate instances have to be created by hand. This is fairly simple by saving them with another file-name and changing the function-name in the first line of the code.

3.4.2 Hair Ganglion

The last base library element to be introduced here is the "Hair Ganglion" model. Its task is, to model the first stage of auditory coding, when the mechanical deflection of the inner hair cells, generate oscillatory cell potentials for each frequency channel. Their amplitude, phase and frequency are coded by the AP-spikes of the ganglion cells, generated by this model element.
To model this coding procedure, a library element, very similar to the static IF neuron, is employed. It represents mainly the ganglion cell, emitting the first spikes of the auditory system and feeding them into the auditory nerve fibers. Structure and parameters of the hair-ganglion model are again shown in figure 3.27

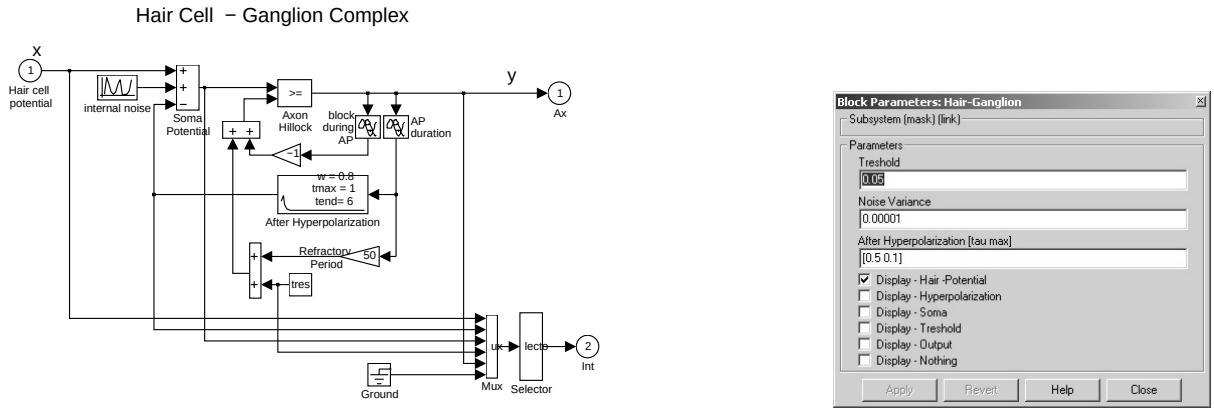

Figure 3.27:
left panel - *structural layer of the neural base library element Hair Ganglion Complex*
right panel - *parametric layer of the Hair-Ganglion base element.*

The essential difference to a normal IF neuron is here that there is no dendritic input to that ganglion cell. Instead it receives sinusoidal hair cell potentials, moving the ganglion soma potential between de- and hyperpolarization back and forth. Each time the cell is depolarized, a number of spikes, depending on the amplitude of hair elongation and the frequency of this channel is generated by this model. Since the typically low firing threshold is reached at exactly the same cycle time of each sine wave, this coding mechanism, realizes phase locking up to $2kHz$ as well as the encoding of interaural time differences (ITD's). The upper limit for phase locking at $2kHz$ is given by the positive cycle time of $250\mu s$ for that frequency, which is just sufficient to generate a single spike per cycle.

To achieve more than one spike per cycle, the hair-ganglion Model has typically short AHP time constants, resulting in a high firing rate during the positive cycle. However, the number of spikes per cycle and therefore the amplitude coding depth achievable with one hair-ganglion Model is limited to 8-10 spikes in case of the $500Hz$ channel (cycle time of 2 ms) and decreases with increasing channel frequency. To reduce this limitation and reflect the one to many connectivity of hair cells, 3 ganglion cells with different firing thresholds have been assigned to each frequency channel and combined within the derived Base Library element *Inner Hair Cell Complex* shown in figure 3.28.

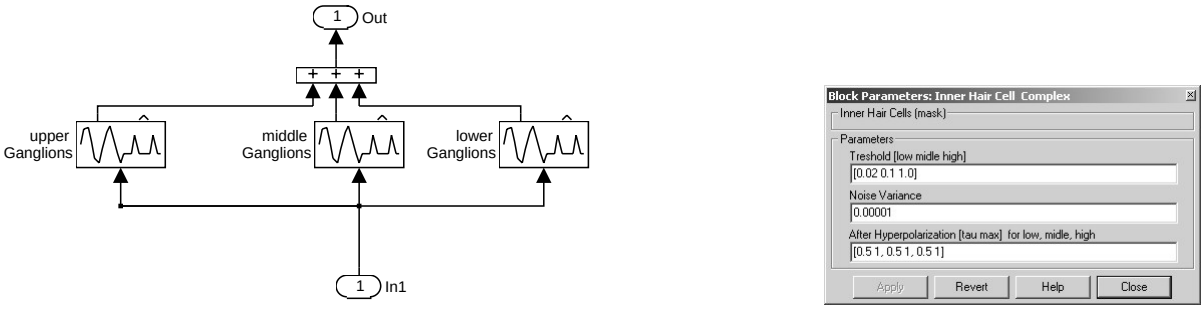

Figure 3.28:
left Panel - *structural layer of the neural base library element Inner Hair Cell Complex*
right Panel - *parametric layer of the Inner Hair Cell Complex.*

As can bee seen in figure 3.28 right panel, the parameters are here to be set for 3 Hair - Ganglion models in one mask, whereas the firing thresholds vary between 0.1 (10 mV) and 1 (100mV). A more detailed evaluation of the Hair Cell Ganglion coding principle and the effect of the tripple cell approach will be provided in subsection 4.2.3“Hair Ganglion Model“.

Having now introduced the essential elements of the Neural Base Library, the next section provides some general simulation parameters and conventions used for this specific model.

3.5 Simulation Conventions

The MATLAB /SIMULINK environment used for the simulation of the proposed model architecture allows the user to simulate dynamic models under a variety of different conditions. These conditions are accessible via the *Simulation / Parameters* menu at the top of the simulation environment. After opening this menu a three rider panel, shown in figure 3.29 becomes visible.

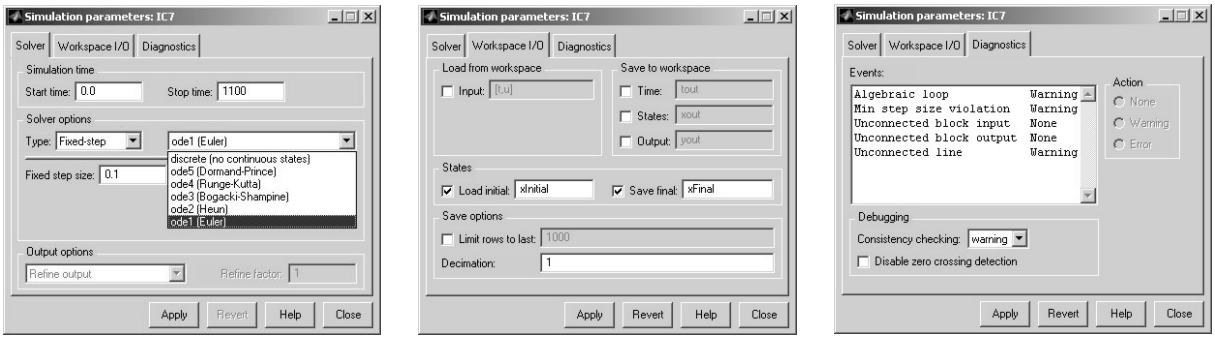

Figure 3.29: **left panel** - *parameters of the Solver rider*
middle panel - *parameters of the Workspace I/O rider*
right panel - *parameters of the Diagnostics rider*

The most important of these riders is the one named *Solver*. At it's top, the experimenter can choose the start and stop time for simulation. These numbers refer to the **Simulation-Time-Step (STS)** of the SIMULINK System, which is identical with the model clock and is used to display the time course within the display elements. In order to interpret the simulation results, *the experimenter must set a convention for the time represented by 1 Simulation Time Step*. Throughout this model **1 STS = 100** μs. Using this convention, a single spike of $100\mu s$ duration for example, lasts exactly one simulation time step and a 1 second Signal is simulated after 10000 STS. It is important to notice that all time constants need to be set using this time measure. For instance to realize a time constant of $5ms$ the referring parameter needs to be set to $50 = 50 \times 100\mu s = 5ms$.

The *Start Time* parameter is usually set to zero, whereas the *Stop Time* has to be selected carefully to cover the essential parts of the signal provided. A special batch routine named "SIMULATE.M" has been designed, to determine the signal length up front and set the Stop Time parameter automatically.

The second part of the Solver-Rider contains the *Solver Options*. Here, the experimenter can determine, which method is used, to solve the differential equations during simulation. The principle solver type can be chosen between *variable step size* and *fixed step size*. In general, the variable step size will lead to more accurate results since the step width is always optimized with regard to the variance of the simulated time courses. However, in case of large models, with many variables simulated at the same time, the optimal step size becomes always very small and the overall simulation time increases significantly.
Therefore, a fixed **Calculation Time Step of 0.1 STS** has been chosen for this model. This means that the value of all internal states and variables are calculated every 0.1 STS, i.e. every 10 μs. If, for instance, the firing threshold of a cell model is now exceeded and has not been above

threshold $10\mu s$ before, a spike will be emitted immediately, although the next full STS is not yet reached. This way, the proposed model architecture realizes a time resolution of constantly $10\mu s$ at all levels.

Finally the experimenter has to chose a *Solver Method*. The list of options is shown in figure 3.29-left panel and contains all typical solver methods including a discrete one in the beginning. In general, the Neural Base Library elements can be used with any continuous or discrete solver method but it is suggested, to use a continuous method, since the many continuous values will otherwise be discretized and this will cause different timings and changes in the dynamic properties. Out of the continuous solver methods, the Euler Method has been chosen for this model, since it could prove sufficient accuracy while realizing a quite fast and effective mechanism. The *Output Options* at the bottom of this rider can be neglected for this model.

The second rider named *Workspace I/O* enables the user, to save and load results at the beginning and the end of the simulation. If one checks the *Input-* check box, input parameters are loaded directly from the MATLAB workspace and must be contained in the [t,u] variable pair. This option should not be used if possible. To work around, the "load from file" block in the model architecture allows to load input variables of any type from *.mat files* containing the time in the first row and any number of input sequences in the rows $2 \times n$.

Somewhat similar to understand are the output variables *Time, States and Output*. Although they can be renamed after checking the referring box, they should not be used, if a standard *Scope* display element already displays the parameter to be stored. Every Scope element contains a "Settings" rider, where a specific workspace variable can be defined the systems saves its simulation results into. The MATLAB command "SAVE" without any syntax will then store all workspace variables into a file, where they remain accessible for later evaluation. To automate this procedure the batch routines "SIMLUATE" and "EXPERIMENT" have been designed and included in the MATLAB path.

A sometimes useful tool of the second rider is contained in the *States* section. Since SIMULINK internally uses a number of internal states x_n to represent all simulated parameters, these states can be saved at the end of the simulation and loaded at another point in time. As long as the model structure remains absolutely identical, the loading of former saved states causes the model to continue the simulation exactly at the same point it stopped before. This is especially helpful, if simulations need to be interrupted, in order not to lose the achieved simulation results. The name of the state variables can be chosen with no significant limitation.

The third rider *Diagnostics*, shall be mentioned here only shortly. If the consistency check is on, MATLAB will provide a list of warnings on the MATLAB Workspace, reminding the user on open outputs, inputs and algebraic loops (i.e. the feed back definition of a variable by itself without time delay). However, model errors like algebraic loops will be always displayed in the workspace and prevent the model from running. In order to prevent the user from a long list of warnings at the start of every simulation due to not connected "Internals" ports of currently not watched cells, the consistency check has been disabled for this model by default.

To summarize the main conventions and parameters of this model, these are:

- 1 Simulation Time Step (STS) refers to $100\mu s$ is used to display the models output and shall be used to define all time constants,
- 1 Calculation Time Step (CTS) is 10 times shorter at $10\mu s$ and is used to define the time resolution of the model calculations,
- this model should be simulated using the Euler Method and a fixed step size of 0.1 defining the relationship between STS and CTS,
- simulations can be interrupted if the state variables are saved after simulation end and loaded to the identical model later on,
- simulation results should be saved using the storage options of all display elements and/or the special batch routines “SIMULATE“ and “EXPERIMENT“.

3.6 Neural Simulation Systems

Before entering the model architecture itself, which is exclusively composed of NBL elements, a short overview on other existing neural network simulation systems will try to answer the question why we had to develop our own environment for simulation of neurons instead using one of the numerous tools in reach of today's neural computation engineers.

By today the number of biologists and bio-engineers striving to validate and evaluate their experimental by realistic models of neurons has significantly increased, and rapidly developing computing power of conventional computers has enabled far more detailed simulation methods than 10 or 20 years ago.

However, the full complexity of dynamic multidimensional information processing in natural neural networks is still not accessible and will remain so due to the general limitations of the conventional computing paradigm discussed in section 1.2.2 and 6.3 of this thesis. Therefore, one of the major reasons to develop our own simulation system was, to find a simulation system which combines three seemingly contradictory features:

1. Including the necessary level of detail to duplicate dynamic properties of neural systems at the level of synapses and cells - enabling the direct comparison of simulation results to recordings from physiological experiments.
2. Using the available power of today's conventional computers and professional software systems to ensure fast performance, graphical interaction and their usage without detailed programming abilities.
3. Ensuring scalability to model entire neural systems (like the auditory brain stem) by enabling a fully parallel implementation as special analog-digital hardware.

To the knowledge of the author, none of the numerous neural simulation systems available in the late 90'th of the 20'th century was able to combine these features. To support this hypothesis some major neural simulation systems will be shortly introduced at this point. For a more comprehensive overview on neural simulation systems, one might refer to *http://www.hirnforschung.net/cneuro/*.

The several simulation systems listed there can be generally divided into two types: First, the *biological driven simulation systems* and second the *functional driven simulation systems*.

A typical example of **biological driven simulation systems** is the **SNNAP** system provided by the department of Neurobiology and Anatomy at the University of Texas Health Science Center. It exhibits a very high accuracy of cell modeling including Hodgkin-Huxley models, ion channel properties and synaptic transmission details. However, a general draw back of biological simulation systems is that this detailed modeling results in a large number of parameters, most of time not accessible even to the biologist, facing the user with many parameters of great impact to the overall function which have to be guessed or extracted from not always comparable sources in

literature. Additionally, their ability to model detailed compartmental neural connections is paid for by a limited network size. Designed for small networks of 3-10 cells SNNAP can now simulate up to 10000 neurons in the batch mode, but in this case it is far from real time performance and no more of value to technical applications. This holds as well for other advanced biological simulation systems like **NEURON** provided by Michael Hines and John W. Moore at the Department of Neurobiology at the Duke University, **MCELL** provided by T.M. Bartol from the Salk Institute and J.R. Stiles from Pittsburg Supercomputing Centre, **CATACOMB** provided by Mike Hasselmo from Boston University or the **GEPASI** biochemical kinetic simulator from the Virginia Bioinformatics Institute.

All of those biological simulation systems lack the ability to be transferred into electronic circuitry, opening the door to truly parallel computing and technical applicability.

The probably most advanced **functional simulation system** based on Integrate and Fire Neurons is the **GENESIS** tool developed at the California Institute of Technology. The earliest GENESIS simulations were biologically realistic large-scale simulations of entire cortical networks (Wilson and Bower, 1992). On the other Hand, the De Schutter and Bower cerebellar Purkinje cell model in 1994 is a typical example of a large detailed single-cell model, with 4550 compartments and 8021 ionic conductances. GENESIS is now often being used for large systems-level models of cerebellar pathways (Stricanne, Morissette and Bower, 1998), and at the other extreme, is increasingly used to relate cellular and network properties to biochemical signaling pathways (Bhalla and Iyengar, 1999).

Originally, GENESIS war purely based on a UNIX scripting language, demanding quite some training of the user, to set up and simulate neural networks. Although, the graphical Interface XODUS has been added in between, its graphical programming capabilities are still limited from the perspective of a biological researcher. Another reason, not to use this system for the purpose of this thesis was the fact that the broad area of application has added a significant amount of complexity to the system increasing the amount of computational overhead significantly. This has somewhat changed since libraries of entire cell models and synapses (similar to the NBL) have been added during the last years but at the time, this thesis was started, the amount of configuration work necessary to set up small networks and display intrinsic parameters has been quite extensive. Finally, also GENSIS is not concerned with parallel computing and entirely relies on dynamic simulations on conventional computing architectures. However, to the believe of the author, only truly parallel and distributed computing in specialized silicon will enable the technical usage of detailed neural cell models.

Other even more functional oriented simulation systems include the **NEOSIM** simulator from the University of Edinburgh, the statistical Neural Network Simulators **ADALINE, BPN, HOPEFIELD, BAM, BOLTZMANN, CPN, SOM and ART1** provided by K. Kutza at *http://www.geocities.com/CapeCanaveral/1624/* and the **NEURON** simulator developed by M. Hines and T. Carnevale at Yale University together with J.W.Moore at Duke University. Those simulators are rather purely concerned to simulate functional aspects of neural cell systems. Therefore they lack the necessary detail to compare their results to physiological recordings and explain neuro-biological experiments.

Last but not least, the development of the Neural Base Library, introduced in this chapter, was driven by a long track of positive experience with the MATLAB simulation environment, providing a huge and simple to access variety of algorithmic signal processing functionalities including the advanced dynamic simulation system (SIMULINK) with many predefined library and display elements and highly developed graphical programming cababilities. The MATLAB Inc. support furthermore ensures continuing portability to all kinds of operating systems and hardware configurations as well as increasing portability onto specialized DSP hardware (i.e. by dSPACE Converters).

In conclusion, non of the available neural simulation systems available during the late 90'th could combine the three aspects mentioned above and the author decided to develop an own library based on the well established MATLAB system.

Based on the elements of this Neural Base Library, the next chapter will describe the developed model architecture of the auditory brainstem, which contains models of several auditory nodes and their connectivity. It is exclusively composed of NBL elements.

Chapter 4

Model Architecture and Physiological Correlates

This chapter will now introduce the neural model architecture in parallel with some essential aspects of the referring physiological structures. It starts with a overview of the major nuclei and their connectivity within the auditory brainstem. Afterwards, the five model stages ascending from the inner ear toward the inferior colliculus and their physiological counterparts will be described in detail including the internal model structure and the simulation parameters used.

4.1 The Auditory Brainstem - Overview and Connectivity

Based on much more accurate and detailed recording and modeling methods during the last five decades, neuroscience has gathered a significant increase of knowledge on the microscopic structure, the connectivity and the functionality of auditory nuclei within the brainstem and higher auditory centers. Although the principles of hearing are still not fully understood, it becomes evident that sonic events, in the way they are visible on oscilloscopes after microphone recording, are never projected directly to the auditory cortex. What is processed there, is a variety of auditory features extracted from the physical phenomena of sound within the auditory brainstem. Human hearing experience, i.e. the impression of a continuous acoustic environment consisting of tones, sounds, speech, music, noise or silence, finally relies on the combination of those features extracted, coded and pre-processed within the auditory centers of the brainstem.

Different from the visual system, the auditory path (displayed in principle in figure 4.1) branches directly after entering the brainstem into separate pathways crossing the two hemispheres several times and reconnecting first time in the inferior colliculus. In between, numerous excitatory and inhibitory interconnections as well as a huge variety of dynamic cell response properties have been detected. Quite well established is the view onto two separate pathways for the decoding of interaural time differences - referred to as the *Timing Pathway* and the decoding of interaural intensity differences - called *Intensity Pathway*.

The **Timing Pathway** mainly emerges bilateral from the ventral cochlear nucleus (VCN) and includes the medial superior olive (MSO) as well as inhibitory projections of the MNTB MNTB - Medial Nucleus of the Trapezoid Body - a small structure in the superior olive transforming excitatory inputs from the contralateral hemisphere into inhibitory signals to the ipsilateral LSO

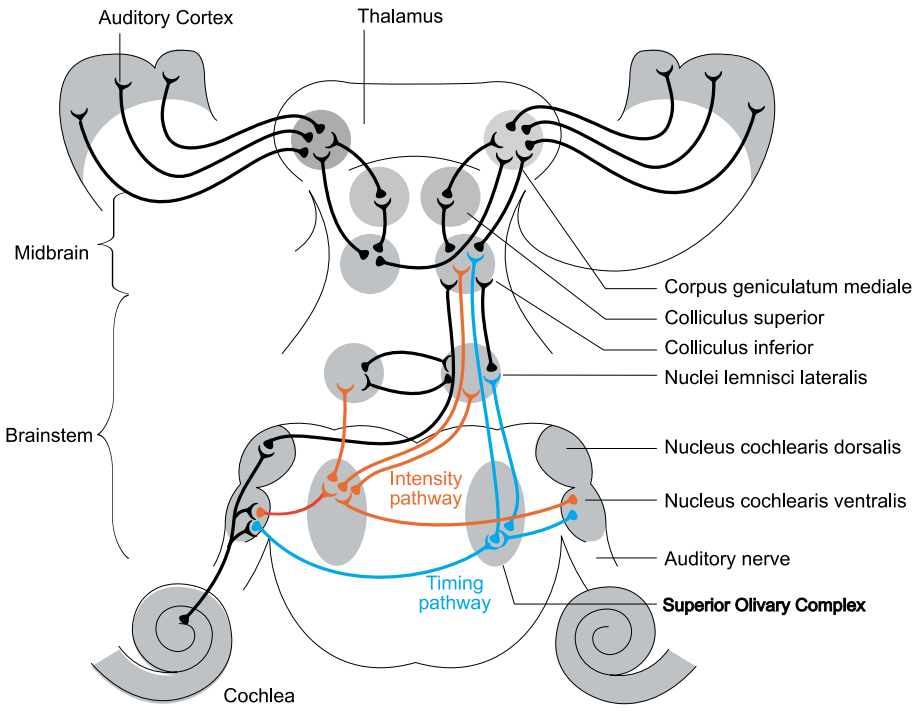

Figure 4.1: Ascending auditory pathways

und MSO and LNTB within the superior olivar complex (SOC). While most of the ascending fibers of the Timing Pathway approach the IC, at least some branch into the DNLL generating very similar responses of MSO Cells and specific DNLL cells of the EE type.

The **Intensity Pathway** originates also in the Ventral Part of the Cochlear Nucleus (VCN) and projects ipsilateral to the LSO as another specific sub-nucleus of the SOC. The same LSO receives inhibitory inputs from the contralateral VCN via the ipsilateral MNTB. From the LSO, afferent projections cross again and excite the contralateral IC as well as the contralateral DNLL. Additional inhibitory projections have been found to approach the ipsilateral DNLL.

As mentioned before, there is convincing evidence for efficient mechanisms to suppress echoes based on the interaction within the Intensity Pathway including additional bilateral inhibition at the site of the DNLL. Therefore, the specific projections of the Intensity Pathway towards and away from the DNLL will be modeled here in detail. At this point it must be mentioned that many auditory perceptual and associative tasks are assigned to the auditory midbrain as well as to thalamic and cortical structures displayed in principle in Figure 4.1. However, those will not be subject to this model.

The quite complex connectivity of auditory centers within the brainstem is displayed with greater detail in figure 4.2. Throughout this chapter this seemingly confusing network will be revealed and step by step evaluated resulting in a fairly simple principle to localize sounds and suppress their echoes. Since most of the excitatory as well as inhibitory projections are known to cross between the two hemispheres, information contained within binaural disparities becomes decoded mainly in the different nuclei of the SOC and interact already at the level of the DNLL.

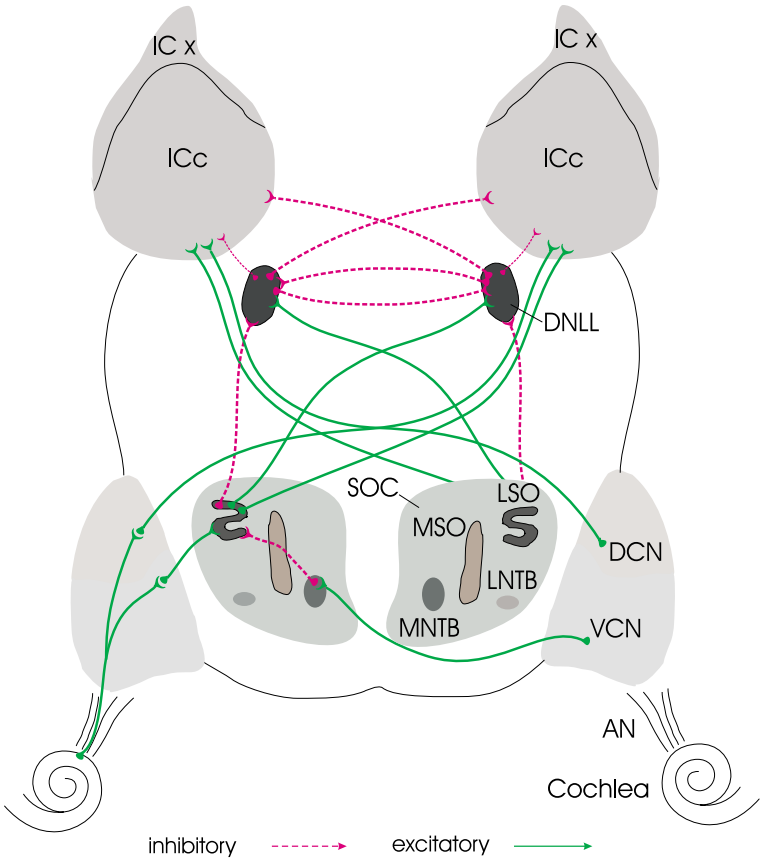

Figure 4.2: Overview of the main nuclei of the auditory brainstem and their connectivity **AN**- *Auditory Nerve,* **VCN**- *Ventral Cochlear Nucleus,* **DCN**- *Dorsal Cochlear Nucleus,* **SOC**- *Superior Olivary Complex,* **MSO**- *Medial Superior Olive,* **LSO**- *Lateral Superior Olive,* **MNTB**- *Medial Nucleus of the Trapezoid Body,* **LNTB**- *Lateral Nucleus of the Trapezoid Body,* **DNLL**- *Dorsal Nucleus of the Lateral Lemniscus,* **ICc**- *central Inferior Colliculus,* **ICx**- *external Inferior Colliculus.*

As can be observed, most of the ascending projections within the auditory brainstem are of excitatory, glutamatic nature. However, recent experiments have revealed that timely exact and efficient inhibitory glycinergic and GABAergic projections within the brainstem, account for a substantial part of the pre-processing carried out in the SOC, the DNLL and the IC. Although the connectivity pattern of figure 4.2 seems quite complex, it does not claim to cover all of the existing projections. In fact, it is a rather limited limited view onto those nuclei and connections, relevant to sound source localization and echo suppression based upon interaural intensity differences.

The general architecture of the model developed to simulate the intrinsic dynamics and functionality of the different nuclei shown above consists of five major stages visualized in figure 4.3. The first stage is concerned with the realistic frequency decomposition and neural coding in the inner ear and the projections from the cochlear hair cells to the ganglion cells of the auditory nerve. The second stage models the onset detection and generation of dynamic spike pattern at the level of the cochlear nucleus (AVCN). The following third stage models the azimuthal detection of sound source locations based on the evaluation of interaural intensity differences

within the LSO nuclei. They project to a fourth stage modeling the echo suppression based on persistent inhibition caused by interactions between the left and right DNLL. Finally, the fifth stage accounts for the correct detection of sound source locations even under echoic conditions and contains an artificial directional sensor despite the models of the left and right IC cells.

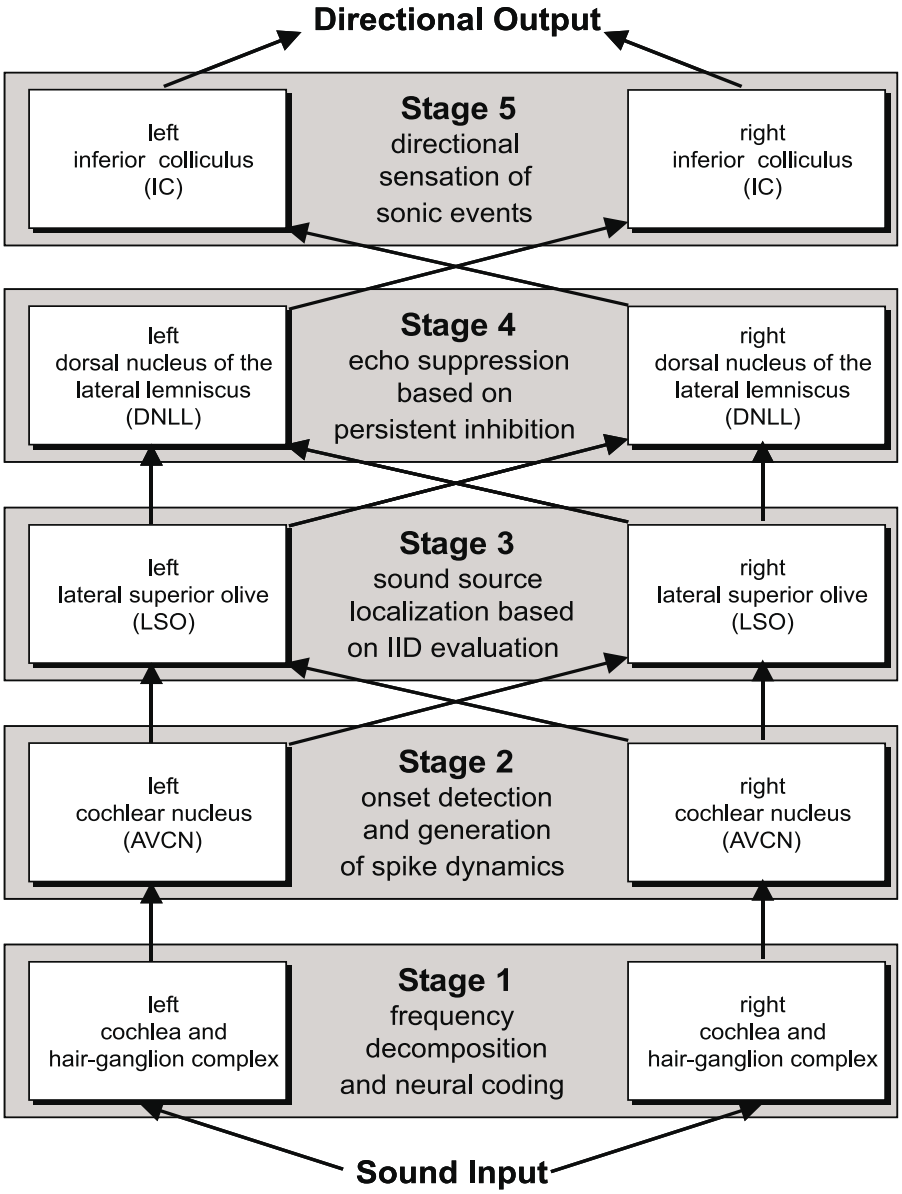

Figure 4.3: Five stage principle architecture of the auditory brain stem model

In order to reveal the modus operandi under different dynamic conditions as well as to achieve a functional model, all of the connections shown in figure 4.2 have been included in the functional SIMULINK model shown figure 4.4.

As visible, the model architecture contains separate structures for all stages of processing and each of the two hemispheres. The architecture starts with a bilateral model of the inner ear including mechanical transduction in the cochlea (light blue) and neural coding by inner hair cell - ganglion complex (cyan).

It is followed by the first center of auditory information processing - the cochlear nucleus (CN) represented in green color only by its antero ventral part (AVCN). From here, the green connections transfer the excitation pattern toward the contralateral central inferior colliculus (at the top of the figure), and toward the ipsilateral LSO (connections at the bottom). Furthermore, inhibitory projections toward the LSO are realized by the light green structures representing the left and right MNTB.

Feeded from both AVCN and MNTB, the blue nuclei refer to the LSO and their connections. They send contralateral excitatory projections toward the DNLL (rising triangles in red color), whereas ipsilateral projections toward the DNLL are of inhibitory nature (falling red triangles). Additionally, excitatory projections toward the contralateral IC are shown by the blue lines approaching the top of the model.

The next stage is marked by the red structures of the DNLL model. As can be seen, beside the contralateral inhibitory projections toward the contralateral IC, there are weak (small triangles) inhibitory influences of the ipsilateral IC and strong inhibitory projections toward the contralateral DNLL.

Finally, at the top of the model, the orange structures contain the cell models of the specific IC cells relevant to this study, receiving input from the contralateral CN, the contralateral LSO and bilaterally from both hemispheres of the DNLL.

At the top of the model, an artificial structure is shown in black, to generate a directional signal, able to guide a technical system like the motor of a robot head. At the bottom, blue filled elements are used to feed the model either with artificial sinusoidal signals, generated by the binaural sinus generator, or recorded files from free field or real world experiments.

The following sections will now describe the internal structure, as well as the dynamic parameters and physiological correlates for each of the nuclei, shown in figure 4.4. Since all of these nuclei are based upon the elements of the neural base library (NBL) (see chapter 3) the description will be limited to the functional level, taking in account that all employed elements have been introduced on the structural, parametric and dynamic level during the last chapter. Last but not least, it shall be mentioned that the architecture shown above, is a one to one view onto the computational model, as it is visible to the experimenter and can be run by MATLAB/SIMULINK.

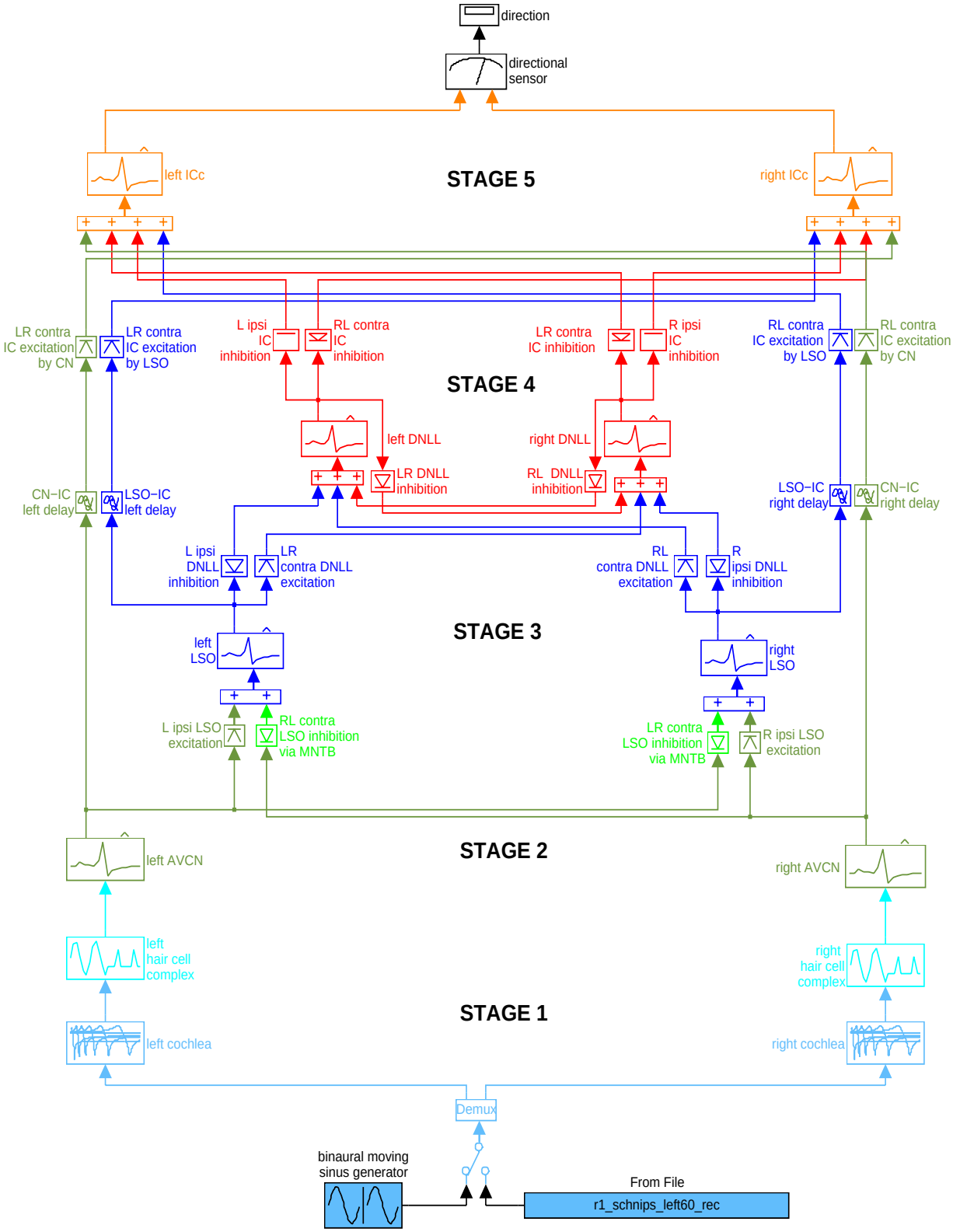

Figure 4.4: Overview of the neural model architecture

4.2 Stage 1 - The Human Ear

4.2.1 Outer, Middle and Inner Ear - Physiological Aspects

Different from the visual system, acoustical signals undergo a number of physical modulations and transformations before they become coded into auditory action potentials. Since the primary sensory range of the auditory system is limited to two 1-dimensional time series at the left and the right ear, the reconstruction of a 3-dimensional auditory space depends on the intrinsic parameters of the time series, like frequency, amplitude, phase, AM modulation and FM modulation. The task of the auditory sensors at the inner ear is, to transform these physical parameters as exactly as possible into specific firing pattern of the acoustical nerve and to separate them step by step from a purely temporal code into a spatio-temporal representation with distributed firing pattern.

Outer and Middle Ear

Hearing doesn't start in the inner ear. It is the shape of our body, our head and of the outer ear which generate specific modulations, important to discriminate sounds in the vertical plane as well as to distinguish between front and rear sounds. Furthermore, the middle ear, containing the tymphanic membrane and the three bones, malleus, incus and stapes, enables the tremendous dynamic hearing range by selective attenuation and the transformation of air pressure waves into micro-mechanic fluid waves within the cochlea. As shown in figure 4.5, beside the bony spiral of the cochlea, the inner ear contains the semicircular ducts of the balance system and the VIII. or Acoustical Nerve.

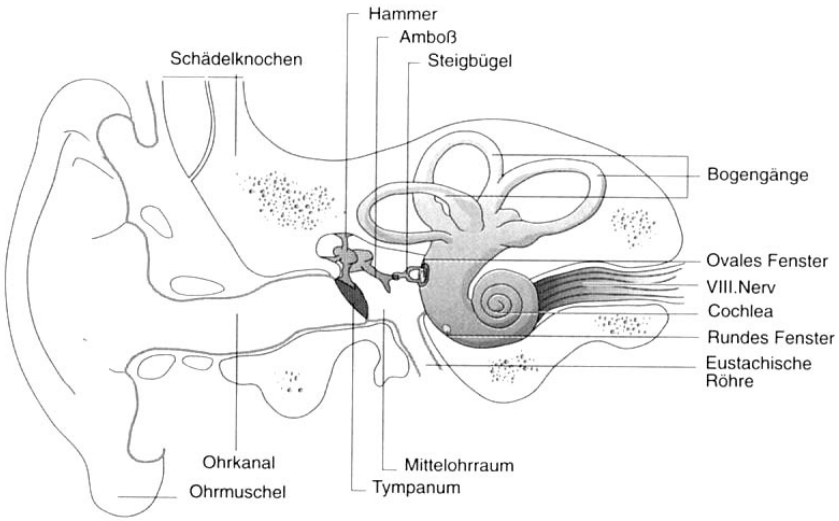

Figure 4.5: Schematic drawing of the essential parts of the human ear according to [DMS01]S.360

All structures of the outer and middle ear are interesting objects of auditory science by them self, but have been excluded from that model, since its aim is currently limited to sound localization

within the horizontal plane under normal condition with no extreme dynamic sound properties. However, further models might add at least some of the modulation properties assigned to the outer and middle ear and extend the application range of this architecture.

Inner Ear

The entire primary coding mechanism is to be found within the spiral shaped bony cochlea of the inner ear. Here mother nature has combined micro-mechanical as well as electrical principles in an amazing manner to achieve a very efficient coding system. The general principle of this system is shown in figure 4.6.

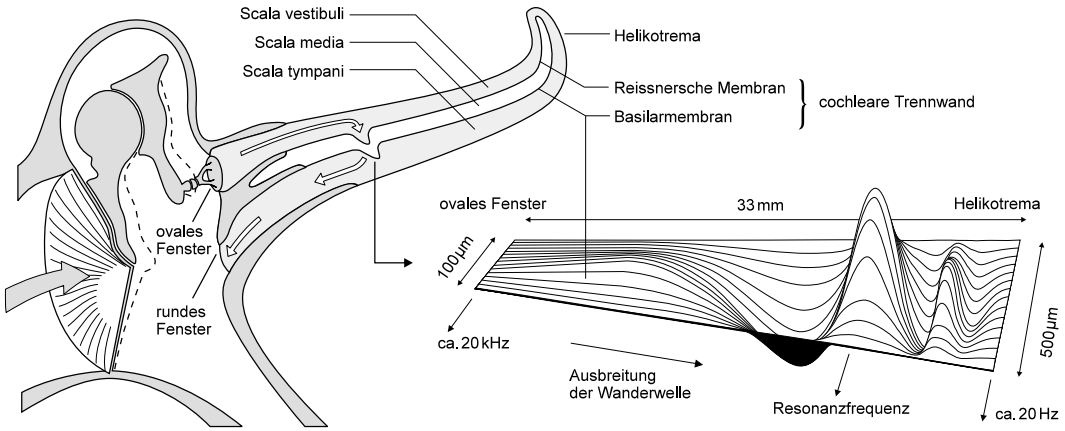

Figure 4.6: Simplified (straightened) display of the cochlea and its basilar membrane in motion

As shown above, the mechanical oscillations of the stapes are transmitted into fluid oscillations at the oval window at the basal end of the fluid filled Cochlea. From here, a traveling wave propagates along the cochlea toward the apex, or the Helikotrema at the inner end of the Cochlea. The necessary pressure equalization happens at the round window at the opposite side of the basal Cochlea end. This traveling wave also moves the cochlear divider including the *Organ of Corti*, shown in figure 4.7, as the organ of auditory coding.

The cross section of the cochlea, shown in figure 4.7 panel A, clarifies the position of the organ of corti within the scala media bordering the scala tymphani. Panel B has magnified the organ of corti and shows it's essential parts. The basilar membrane including 3 rows of outer hair cells (OHC) and 1 row of inner hair cells (IHC), and the tectorial membrane lying above with the ends of the hairs attached to it.
As the traveling wave now moves along the cochlea, it realizes the first stage of decoding and creates one of the major decoding principles of the auditory system - the **tonotopy**. Since the basilar membrane is narrow and stiff near the oval window and becomes more broader and more flexible toward the apex, its resonance frequencies changes from high frequencies, near the oval window, toward low frequencies, near the apex. Based on these mechanical properties, a sound of a specific frequency will cause the basilar membrane to resonate at a specific location along the cochlea as shown in figure 4.6.

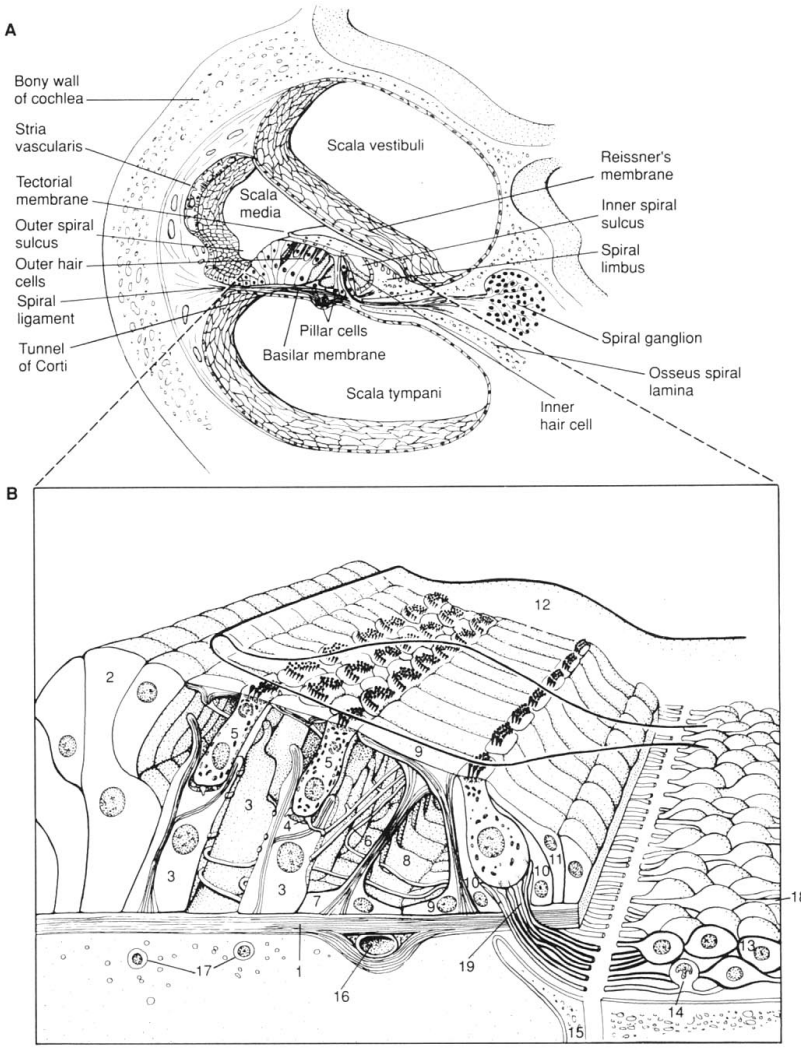

Figure 4.7: Structure and essential parts of the organ of corti by permission of [KSJ91]
1-11 cell types of the basilar membrane with outer hair cells (5), inner hair cells (10), tectorial membrane (12) and ganglion cells (13/14)

The spectral components of a specific sound are this way spatially decoded covering a range of 50 Hz (near the helikotrema) up to 20 kHz (near the oval window). This spatially ordered representation of frequency components is called *Tonotopy* and holds for many auditory regions up to the primary auditory cortex.
However, the louder a sound, the more energy will be transferred into the cochlea and the movement of the basilar membrane will not just increase in amplitude but also in the overall portion being set in motion, i.e. a very loud sound will move nearly all parts of the cochlea beginning from the base and reaching the helikotrema after some *cochlear time-lag*. On the other hand, a gentle sound will cause the cochlea to move only at a very specific location.

If one now takes a look at a specific location along the cochlea, the mechano-electrical transduction principle becomes visible as displayed in figure 4.8.

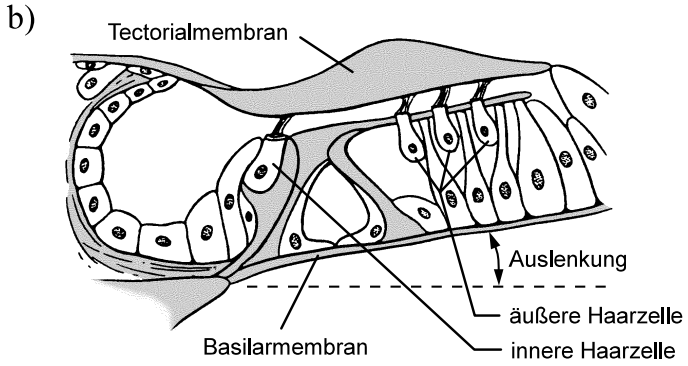

Figure 4.8: Mechano-electrical transduction by deflection of hair cell Sterocilia as a result of the oscillation of the basilar membrane

As the basilar membrane including their hair cells is moved by the traveling wave, the hairs (stereocilia) become periodically bent from one side to the other, since their ends are fixated within the tectorial membrane. This deflection of hairs, exactly follows the the time course of the membrane oscillations at this specific point, i.e. the time course of the specific resonance frequency component called *best matching frequency (BMF) or characteristic frequency (CF)* of that location. Due to the different length of the inner hair cell - stereocilia, their bending results in an opening and closing of ion channels as shown in principle in figure 4.9 panel A.

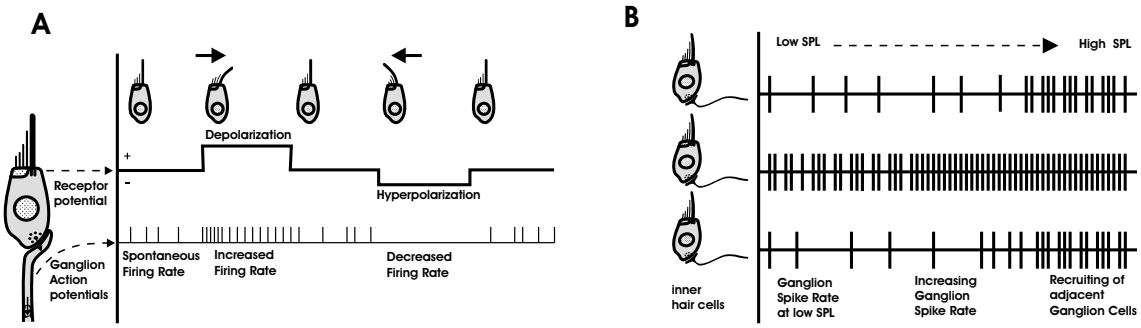

Figure 4.9:
panel A - *periodic depolarization and hyperpolarization of hair cells by the deflection of stereocilia.*
panel B - *recruiting of adjacent cells to code higher SPL's.*

If a inner hair cell becomes depolarized by an influx of positive ions during the bending of hairs to the right, transmitter is released at the numerous somatic terminals of *spiral ganglion cells*. This way the oscillating inner hair cell potential is directly transferred to about 30 different ganglion cells. However, every ganglion cell is only excited by one inner hair cell. This excitation leads to a change of the ganglion's firing rate, depending on the current cycle of movement at a specific cochlea position. In fact, the ganglion cell will only fire during the positive (depolarizing) part of the oscillation cycle. Panel B of figure 4.9 shows another important coding scheme. As the SPL (and therefore the deflection of stereocilia) increases, a single ganglion cell might reach its maximum firing rate. In this case, the rate code is accompanied by a place code. Adjacent ganglion cells, which receive their input from the same hair cell, become recruited and start firing at higher depolarization values (i.e. firing thresholds).

Finally, the 30 to 50 thousand axons of the spiral ganglion cells make up the **acoustical nerve (AN)**. It leaves the inner ear toward the cochlear nucleus and carries all of the acoustical information now coded as auditory action potentials. In principle, the *frequency components* of each sound are coded by place (tonotopy) in the different fibers of the acoustical nerve, the *phase* of each component is coded in the timing of spikes within that specific fiber and the *amplitude* of that component is coded in the number of spikes per cycle as well as the number of AN fibers (ganglion cells) being recruited. The coding of AM (amplitude modulation) and FM (frequency modulation) will be demonstrated in the next section using the developed model of the inner ear.

While the inner hair cells nearly exclusively excite the afferent fibers of the acoustical nerve, the circumstances for the outer hair cells are quite opposite. Only about 10% of the sensoric fibers originate from those cells despite their relatively large number placed in 3 parallel rows. Most of the outer hair cells are innervated by efferent fibers, reaching the inner ear from higher auditory centers. Furthermore, it has been shown that outer hair cells can actively contract or elongate their hairs. This leads to the very likely, but not yet proven assumption that Outer Hair Cells are mainly involved with the location specific modulation of the flexibility between basilar and tectorial membrane and therefore, the active modulation of sensory information by the higher auditory system.

The tonotopic organization principle is very prominent within the acoustical nerve. However, a single AN - fiber does not only respond to its best-matching-, or characteristic frequency. As can be observed in figure 4.10, their sensitivity to other frequencies can be described by asymmetric **tuning curves**, which remind on the traveling wave, moving from high to low frequencies. These tuning curves have been measured for many species and shall be displayed by example for the acoustical nerve of the Chincilla.

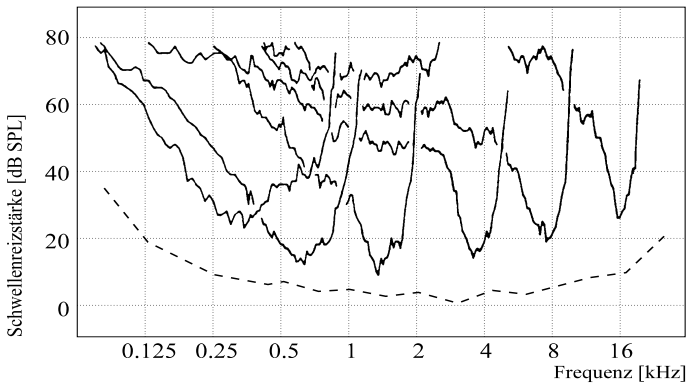

Figure 4.10: Tuning Curves of six different nerve fibers originating from different inner hair cells along the cochlea of the Chincilla after Ruggero [Rug92]

Beside the different characteristic frequencies and the tonotopic organization of the acoustical nerve, figure 4.10 also shows the asymmetric nature of the tuning curves preferring the lower neighboring frequencies against the higher ones. These tuning curves will be fairly well resembled by the frequency response of the filters, employed to model the cochlea and introduced during the next subsection.

4.2.2 Cochlea Model

In order to model the micro-mechanic principles of the inner ear, a simple but adequate method is needed to perform the main task of the basilar membrane - the frequency decomposition of sound. But this natural frequency decomposition has to fulfill at least four requirements:

1. A logarithmic frequency resolution between 50 Hz and 20 kHz needs to be combined with a time resolution below the range of milliseconds.

2. The time course of every frequency component, needs to be exactly preserved with regard to amplitude and phase.

3. All frequency components need to be available at every point in time and

4. the frequency response of each channel needs to reassemble the asymmetric tuning curves of figure 4.10 including the time lag of the traveling wave.

The classical engineering method to perform a frequency decomposition is the *Fourier Transformation* and it's realization as *Fast Fourier Transformation (FFT)*. However, it will, like all windows based methods, clearly fail to fulfill criteria 1. As long as a window of defined length is necessary for calculation, the inherent dilemma between frequency and time resolution will be prominent and will either cover the full range of frequencies **or** achieve the necessary time resolution of less than $1ms$, but never both. Also criteria 4 is not acomplishable by Fourier algorithms and criteria 2 can only be fulfilled by using the complex numbered Fourier spectrum, which is causing problems during later processing.

A second method to model the Cochlea has been evaluated during former work of the author [Zah96] and is based on Adaptive Resonance (AR) and Moving Average (MA) filters. Here digital filters adjust their filter parameters by feed back adaptation to the inherent frequency spectrum of the signal, this way excluding transient and variant noise components. Although those AR, MA or ARMA models fulfill the criteria 3 and 4 quite well and reach a better time resolution than FFT methods, their speed of adaptation is not fast enough to mirror signal components in the range of μs. Furthermore, they have problems to preserve phase differences as low as $10\mu s$ containing interaural time difference (ITD) information, essential for the auditory system.

A common method to model the cochlea by today, are separate, parallel digital filters with free definable frequency characteristics. The resonance-frequencies of those filters can bee separately defined and also positioned on a logarithmic scale (criteria 1a). Their time resolution depends on the filter order but reaches in case of 2'nd order filters and sampling frequencies of $44kHz$ down to $5\mu s$. This proves sufficient as time resolution for the cochlear mechanism (criteria 1b). Since every filter generates a separate time series, referring to a specific position along the cochlea, amplitude and phase of that specific CF are correctly preserved and represented in the time domain (criteria 2). Every filter generates an output value for every sampling point in time, therefore all frequency components are available at any point in time (criteria 3). Finally, the reassembling of tuning curves (4'th criteria) is fairly well, but not fully fulfilled. Of course, the frequency characteristic of every filter can be designed to be asymmetric preferring the lower

neighbors as shown in figure 4.10. However, they will all react at the same time and therefore miss the time-lag effect of the traveling wave within the cochlea.

This has led the author, to replace the parallel filters by a **Cascade of All Pole Gammatone Filters (APGT)**. These are special Gammatone Filters (GTF), with all zeros removed, and only one complex pair of poles resulting in the laplace transfer function

$$H(s) = \frac{K}{[(s-p)(s-p^*)]^N} \tag{4.1}$$

Here, the s-plane pole position is given by the complex number p and its conjungate p^*. K is the constant to adjust the unit gain to 1 ($H(0) = 1$). And N marks the order of the filter, set to 2 for this model.
In the cartesian parameterization of complex pole positions $p = -b + j\omega_r$ the APGF can be written as:

$$H(s) = \frac{b^2 + \omega_r^2}{[(s+b)^2 + \omega_r^2]^N} \tag{4.2}$$

Due to their frequency characteristic shown in figure 4.11, APGT Filters are especially suitable to model the asymmetric tuning curves of the auditory nerve fibers. As displayed, the removal of zeros removes the tail-gain scatter and sets the tail gain to zero instead of significant dumping as normally perceived in case of GTF and Differentiated All Pole Gammatone Filters (DAPGF).

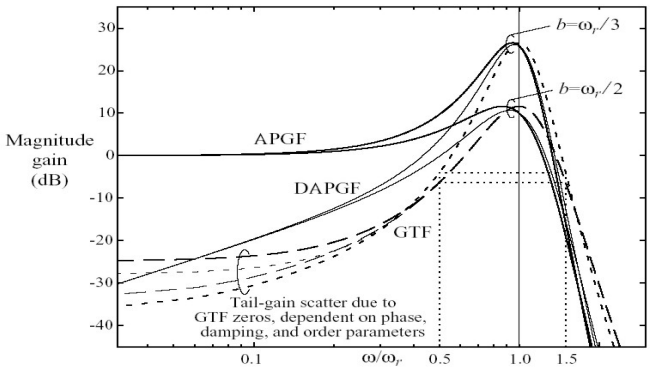

Figure 4.11: Comparison of GTF, AGTF and DAPGF for two different real parts of the pole location b

The zero tail gain of the APGT assures that the sound in this frequency range is not modulated when the signal passes the filter and therefore a cascade of those filters will, like in the cochlea, select only specific frequency components for resonance (for example at $b = \omega_r/2$) and cut of the signal afterwards. In fact the APGT realizes a combination of a second order lowpass filter with cutoff frequency just above CF and a second order bandpass filter function as a resonator for the frequency band around the CF. Therefore, if the signal moves trough a cascade starting with high resonance frequencies, the highest frequencies will resonate at the first filter and will then be cutoff as the sound moves on to the next one. For a more detailed description one might refer to a detailed report on these filters of Koehler [Koe98].

The APGT filters have been first mentioned by Flanagan [Fla60] to be helpful models for the basilar membrane. However, they only became a modeling standard based upon publications by

Richard F.(Dick) Lyon [LM88],[Lyo97] and the development and public distribution of simulation code within Malcom Slaney's "Auditory Toolbox" [Sla93]. Since then, they are often referred to as *Lyon Filters* as well.

For this model, the specific filter cascade shown in figure 4.12 has been designed and used throughout the experiments.

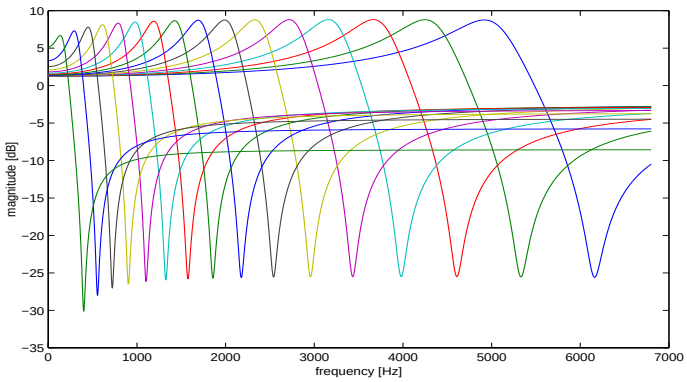

Figure 4.12: Frequency response of the 16 filters of the cochlear cascade with characteristic frequencies (CF) reaching on a logarithmic scale from 180 Hz to 5050 Hz.

The upper limit of 5050 Hz has been chosen, due to the fact that the sounds presented to the model have been recorded with a sampling frequency of 11025 kHz. According to the *Nyquist criteria* this results in a maximum sampled frequency of 5050 Hz. 180 Hz, as the lower boarder, results from the logarithmic positioning of the computationally desired 16 filters along the frequency range. If one interprets the resonance frequencies of the Lyon filters as characteristic frequencies of the auditory nerve fibers, shown in figure 4.10 on page 92, the similarity of the model to the natural system becomes obvious.

Another way to interpret the validity of that cochlea model is to look at it's responses to different simple and complex sounds as shown in figure 4.13.

As can be observed in figure 4.13 panel A-C, the filter with the closest CF always reacts best, but on the low frequency side there are quite some neighboring cells animated as well. This emulates the traveling wave within the cochlea fairly well, since the low frequencies also react with the typical cochlear time lag seen in the studies of natural cochleas.

Panels D to F impressively point out, how the cochlea codes modulation frequencies of the AM (Amplitude Modulation) or FM (Frequency Modulation) type. While the AM modulation, shown in panel D leads to rhythmic oscillation of all channel amplitudes, the modulation of the signal onto an 5 kHz FM carrier results in a second local elongation maximum at the first Harmonic (around 2 kHz). Finally panel F shows that this cochlea model proves capable even to code signals with contain both frequency and amplitude modulations. Although the pattern looks already complicated, it obviously contains all necessary components needed to decode the sound modulation and characteristics at higher auditory centers.

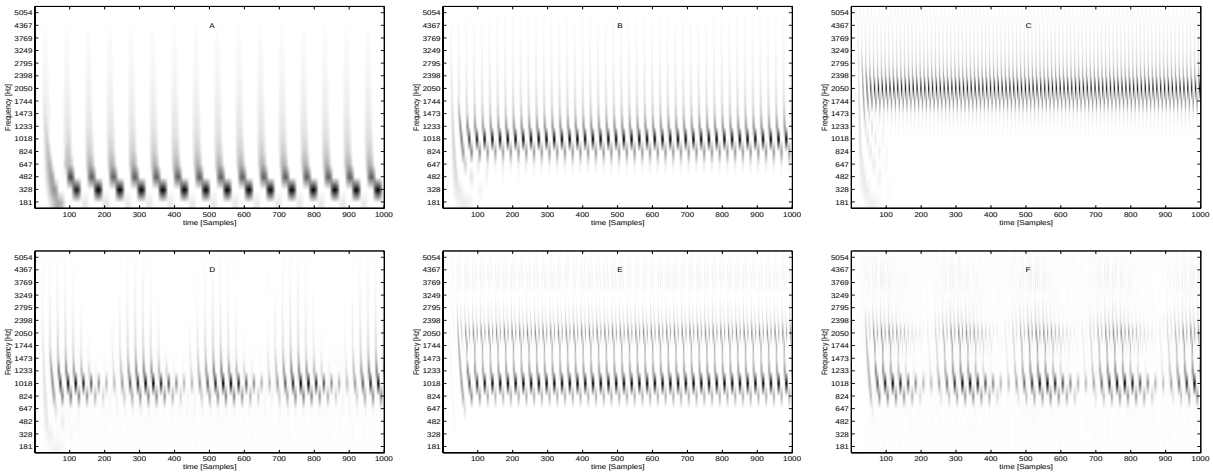

Figure 4.13:
Panel A - *model response to a 300 Hz pure sinus,* **Panel B** - *model response to a 1000 Hz pure sinus,* **Panel C** - *model response to a 2000 Hz pure sinus,* **Panel D** - *model response to a 1000 Hz pure sinus, AM modulated with 100 Hz,* **Panel E** - *model response to a 1000 Hz pure sinus, FM modulated to a carrier of 5000 Hz,* **Panel F** - *model response to a 1000 Hz pure sinus, FM modulated to a carrier of 5000 Hz, and AM modulated with 100 Hz.*

To summarize, the complicated mechanical properties of the Inner Ear cochlea have been modeled by two instances of All Pole Gamma Tone Filter Cascades with 16 logarithmic positioned filters on each site. They are accessible via the two model elements "Left Cochlea" and "Right Cochlea", shown in the figure 4.14, and mark the first stage of the proposed architecture.

4.2.3 Hair-Ganglion Model

As can be seen in figure 4.14, the first stage of the architecture, modeling the inner ear, contains two elements for each hemisphere. Beside the left and right cochlea filter cascades represented by the cochlea library elements, there is another block on each side named *hair-cell-ganglion complex.*

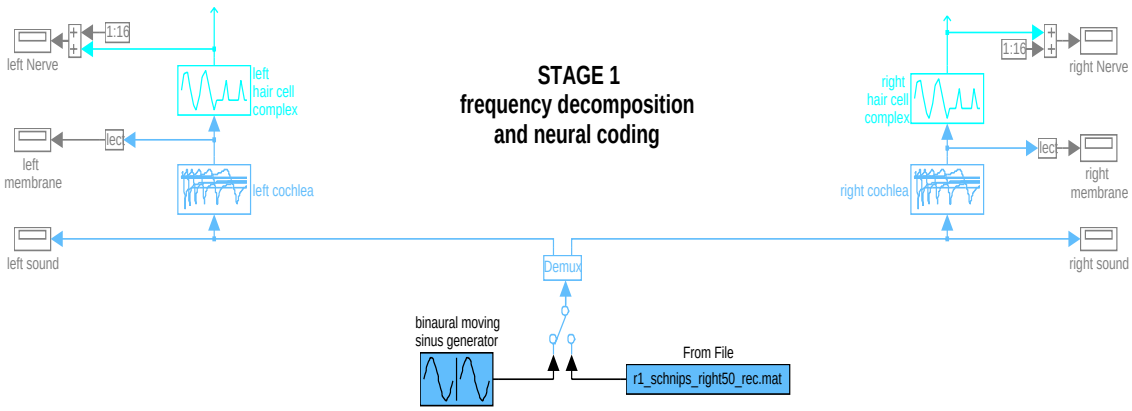

Figure 4.14: First stage of the model architecture, containing two instances of the 16 channel Cochlea filter cascades and two instances of the left and right hair-cell-ganglion complex

It has been described in subsection 4.2.1 that the true transduction of mechanical waves into electrical action potentials happens within the hair cell - ganglion complex. During this task, the function of the inner hair cells (IHC) is, to transform mechanical oscillations into electrical potentials and the task of the ganglion cells is, to generate the adequate chains of action potentials coding the parameters of those oscillations.

If one assumes that every hair cell potential resembles the sinusoidal movement of the basilar membrane at the specific location of that hair cell, the filter output, modeling this movement, can be seen as the inner hair cell potential. Additionally, it has been mentioned that due to the many hair cells along the cochlea, a nearly continuous solution of the frequency spectrum is achieved, which can be only roughly represented by the 16 frequency channels of the cochlea model. However, each of the filters represents it's channel fairly well. And since the hair cell - ganglion connection is a direct one, the filter output will not only be interpreted as the hair cell potential but also account for the input potential of the ganglion cells as described during the introduction of the hair cell ganglion model in chapter 3.

The specificity of the two hair cell ganglion complexes, shown in cyan color in figure 4.14 is, that within that structure, three ganglion cells are feeded by the same hair cell, i.e. the same filter channel. Each of these ganglion cells has a different firing threshold and therefore, the extended coding range, shown in figure 4.15, is achieved. This additional place code principle enables the model to efficiently and continuously code the entire dynamic hearing range within each of the 16 frequency channels.

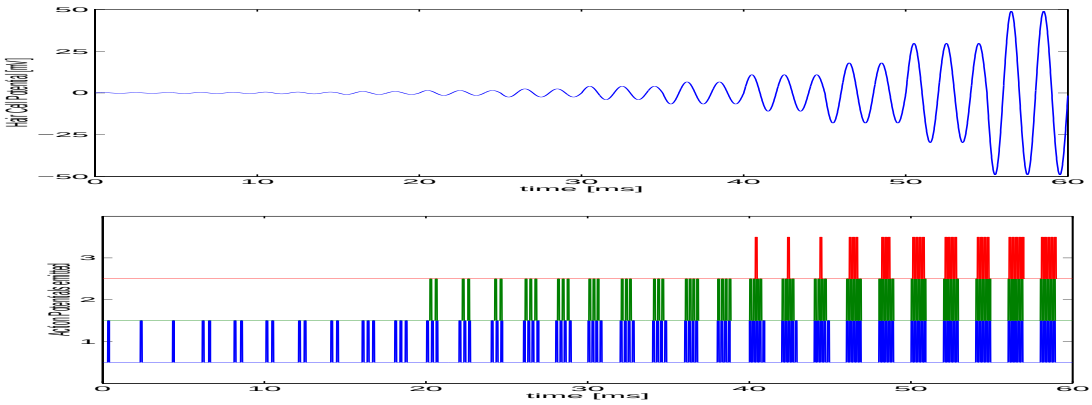

Figure 4.15: **upper panel** - *oscillating input potential with increasing intensity from 0 to 120 dB,* **lower panel** - *response of the 3 Ganglion Cells assigned to that frequency channel; blue cell with lowest firing threshold, red - Ganglion Cell with highest firing theshold.*

Beside the intensity coding shown, in figure 4.15, the spike chains of the ganglion cells also need to preserve the phase information of all frequency components. It's realization is another example of the simple but very effective natural auditory coding system. As shown, in figure 4.16, the interaural phase difference of an AM modulated sound arriving first at the left and later at the right ear, is well preserved in the spiking pattern of the referring frequency channels at the left and right inner ear. Since the first spike of each wave is always emitted at the same point of the phase cycle, the ganglion cells realize a **phase locking** up to the 5 kHz channel frequency in the model and about 1 kHz in nature. Above that phase locking limit, the cell will still always fire at the same point of the cycle but will leave out some of them, using again a population

code to reach a stable phase coding. Because population codes are always hard to simulate, they have been worked around, by cells with very short AHP periods, capable to spike with a 5 kHz frequency - as the highest frequency detected in this model.

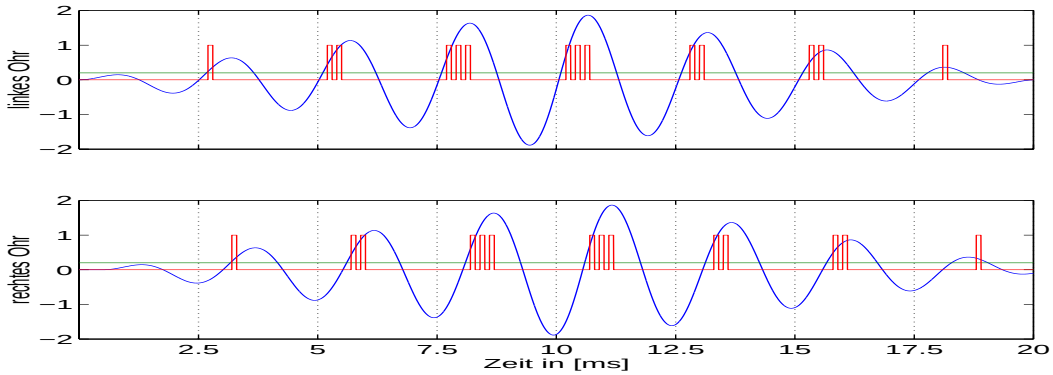

Figure 4.16: Coding of Interaural Phase Differences and Phase Locking during spike generation in the Hair-Ganglion Complex

As the amplitude of the hair cell potential is raised, of course the point of firing moves slightly toward earlier cycle times but since this effect is equal at both sites it will not influence interaural time delay coding. However, it will generate a phase coding of amplitudes called *latency coding*, where higher amplitudes cause earlier spikes and lower ones cause later spiking. This method of phase related intensity coding accompanies the intensity coding shown above and supports the decoding in the SOC, as described later. In general, the accuracy of phase coding goes down to about $4\mu s$ in nature [1] and can be followed down to $10\mu s$ by this model. The reasons for that is the Computation Time Step (CTS) already discussed.

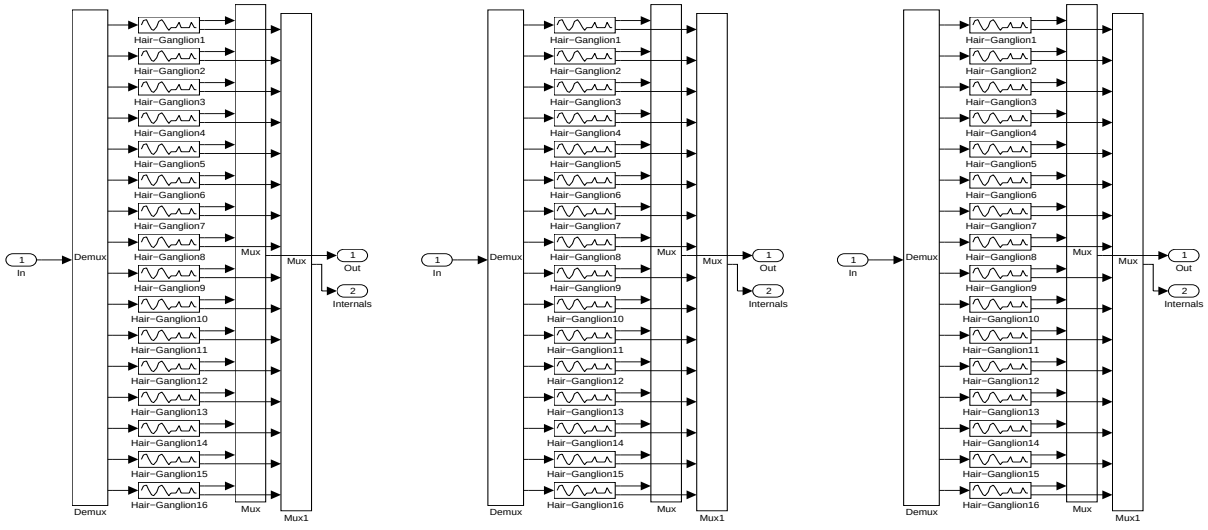

Figure 4.17: Inner Structure of the IHC-Ganglion Complex

[1] according to the newest results of J. Blauert mentioned by Alan Palmer from MRC Institute of Hearing, Nothingham during a talk at the M. Planck Institute of Neurobiology, Munich on January 13'th 2003

Overall, each IHC-Ganglion Complex model contains $3 \times 16 = 48$ ganglion cells, resulting in 96 ganglion cells for both hemispheres. Since every cell receives exactly one input from one hair cell this would have led to 96 separate synapses to be modeled. But since the result of this synaptic transmission is assumed to be a simple transfer of the hair cell potential to the soma of the ganglion cells it has been not separately modeled at this stage and the filter output is directly used as soma potential of the ganglion cells. The internal structure of the hair-cell-ganglion complex therefore is fairly simple as displayed in figure 4.17.

The output of the hair-cell-ganglion model within the 16 separate channels on each hemisphere, can be seen as the activity of the acoustical nerve fibers. Their firing pattern in response to a sinusoidal sound of 1 kHz moving from the left to the right side of the listener is shown in figure 4.18 including the sound itself and the movements of the basilar membrane.

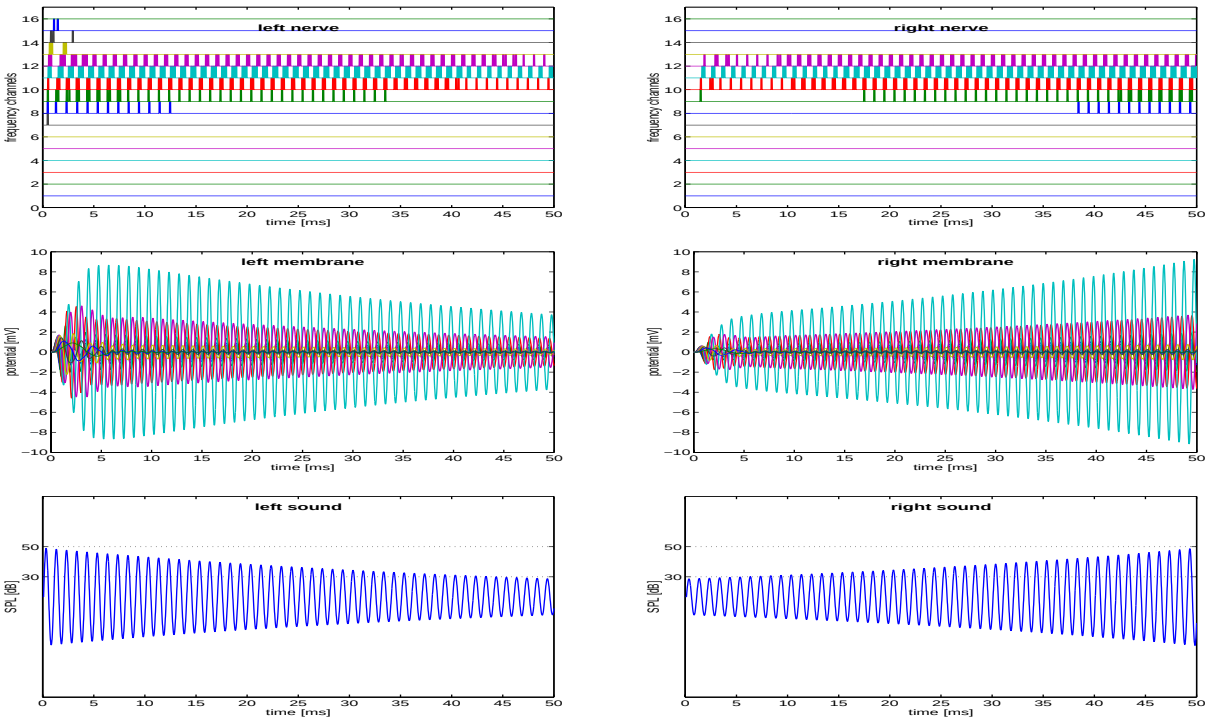

Figure 4.18: **lower panels** - *sound signals presented to the model at the left and right ear*
middle panels - *resulting hair cell potentials along the left and right cochlea*
upper panels - *sounds coded into spike chains of the left and right acoustical nerve*

As the sound source moves from the left to the right side, the SPL on the left ear decreases from 50 to 30 dB while the SPL on the right side increases accordingly - this way switching the interaural intensity difference from +20 to -20 dB with respect to the left side. Additionally, the Interaural Time Delay between left and right sound moves from $+200\mu s$ (left sound leading) to $-200\mu s$ right sound leading. This cannot be observed by the eye but becomes clearly vissible after zooming into the simulation result, possible for the experimenter at any point in time.
It can be observed that the higher the SPL of that pure tone, the larger is the amplitude response of the specific filter with the best matching characteristic frequency - here the one at 980 Hz in cyan color. Furthermore, as the SPL increases, not only the two neighboring filters (red and magenta) start moving but also the green and the blue channel becomes recruited in case of 50

dB SPL. This additional channel recruiting results from the asymmetric tuning curves which in turn results from the mechanical properties of the basilar membrane and marks an additional way to code the SPL (and IID).

Finally, the firing pattern of the acoustical nerves (shown in the upper panels of figure 4.18) displays the phase locking as well as the intensity coding by firing rate, as described before. The intensity coding, caused by recruiting of additional ganglion cells, is not observable in this schematic diagram, since only one (the middle) fiber of each channel is shown here for clarification. It has been displayed in figure 4.15.

After the detailed introduction of the first model stage, we will now turn to the next model stage containing the cells and interconnections of the cochlear nucleus.

4.3 Stage 2 - The Cochlear Nucleus

The cochlear nucleus (CN) is the first processing center of the auditory system. All acoustical nerve fibers end here. At the CN, the firing pattern of the nerve fibers is first time analyzed and several monaural features become extracted in the different regions of the cochlear nucleus. The characteristic features to be extracted are, however, limited to monaural cues, since due to the purely ipsilateral input to each of the CN's there is no contralateral information available at this stage of processing. In section 4.1 it has been shown that the timing pathway as well as the intensity pathway originate in the ventral part of the cochlear nucleus, but there are also direct projections, toward the inferior colliculus from the cochlear nucleus. To prepare the raw information, arriving from the auditory nerve for further processing within these different pathways and centers, the secondary auditory neurons of the CN exhibit specific response properties and form complex receptive fields (mainly in the dorsal part), contributing to the necessary feature extraction. Because this model is currently limited, to the evaluation of interaural intensity differences (IID) only those regions in the AVCN, where the intensity pathway originates have been modeled here. However, during the time course of this work, several models for onset detection, modulation frequency extraction and spectral solution enhancement have been developed by the author and are available upon request.

4.3.1 CN - Physiological Aspects

Within the Cochlear Nucleus, a number of sub areals can be distinguished, based upon different cell types with characteristic firing pattern. An overview on the physiological subdivisions and their typical cell types is shown in figure 4.19.

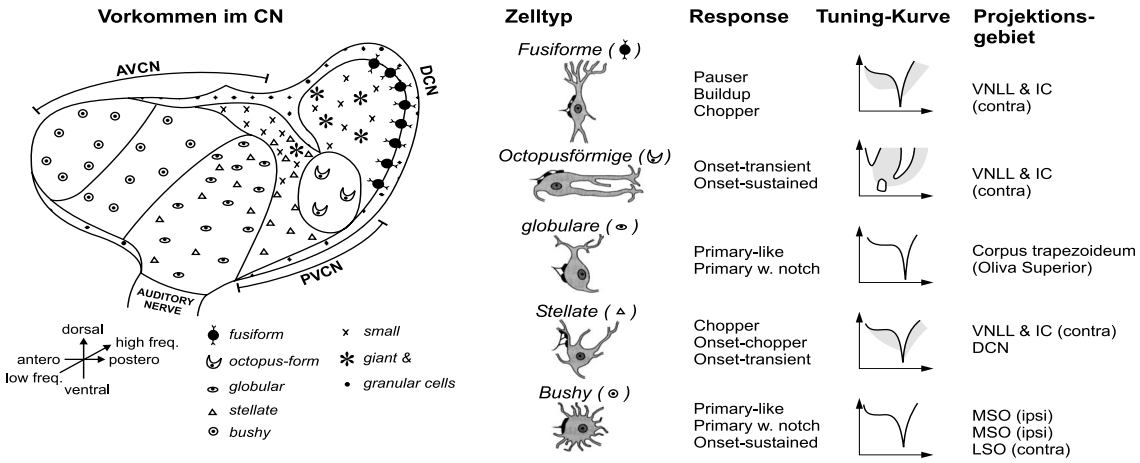

Figure 4.19: Cell types and response properties in the CN after Rouiller [Rou97] pg.17/18.

As can be seen, the eight main cell types of the CN in vertebrates are: fusiforme, octopus, globular, stellate, bushy, small, giant and granular cells. Bushy cells are the main population within the antero ventral cochlear nucleus (AVCN) while stellate and globular cells populate the postero ventral cochlear nucleus (PVCN). The octopus cells form a separate region within the PVCN close to the DCN. The DCN itsself is populated by small and giant cells in the central region and granular cells in the external layers. Finally, the fusiform cell types have been found to link the central DCN to the external DCN.

But the physiological forms of those cell types are not directly linked to their response properties. The post stimulus time histograms (PSTH) [2] displayed in figure 4.20 show quite different response types for the different regions of the CN.

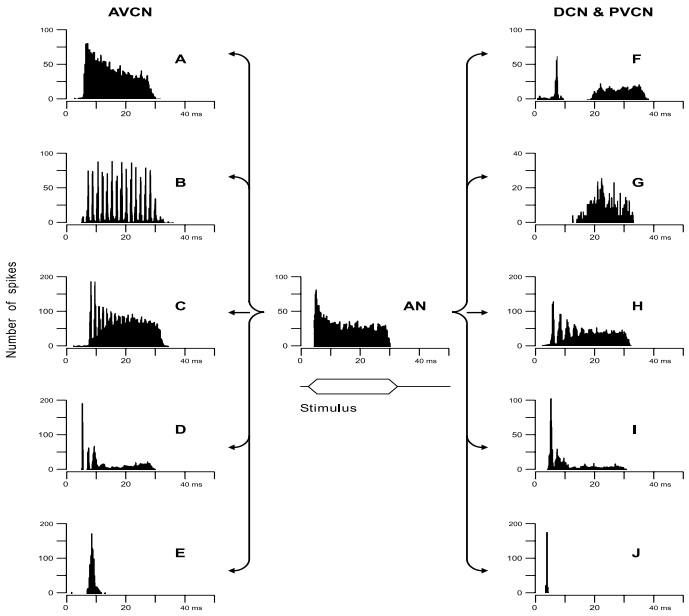

Figure 4.20: post-stimulus-time-histograms of typical excitation pattern in the CN in response to a 25 ms tone burst. For comparison the response in the acoustical nerve (AN) is shown in the center Typical response types in the AVCN are: A-primary like, B-phase locked, C-sustained chopper, D-onset chopper, E-pure onset. The PVCN and DCN cells exhibit: F-pauser, G-bildup, H-sustained chopper, I-onset sustained and J-onset transient responses. (by permission of [RA97] pg.131).

The main response type employed in this model is the primary like response of the bushy cells in the AVCN. Therefore it shall be described here with some detail:
The name primary like is derived from the fact that those cells respond very similar to the primary auditory nerve cells. That means, they exhibit a short phase of intense (phasic) firing during the onset of a sound, which is followed by a less intense (tonic) firing behavior as long as the signal remains stable and is statistically invariant. If the signal changes, the firing is again intensified. This response type is also called *phasic-tonic response* and within the CN it is found nearly exclusively in the AVCN. Some of the cells exhibit a short ($1ms$) gap (*a notch*) after the phasic burst in case of high SPL intensities. Primary like and primary like notch are the typical responses of the *bushy cells*.

The dorsal part of the CN, as seen above, contains more complex and chopper like response pattern. It is assumed that this is the location, where monaural spectral cues and modulation frequency decomposition happen to be initially performed. Most of it's projections reach the inferior colliculus directly, but since spectral and AM/FM cues are not subject to this study, the DCN is not part of that model.

[2] spike timing in response to a short stimuli, recorded from a cell assembly during several trials - sometimes referred to as peri stimulus time histogram

4.3.2 AVCN Model

As mentioned, only the antero ventral part of the cochlear nucleus has been included in the model architecture. Shown in figure 4.21 the second model stage containing the left and the right AVCN is directly feeded by the fibers of the acoustical nerve.

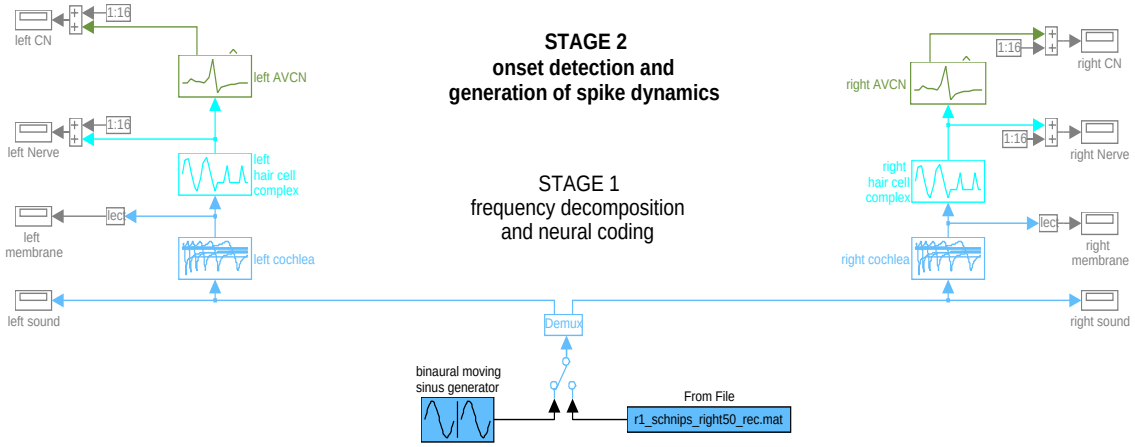

Figure 4.21: Structure of the first and second model stage including the left and right AVCN

The cells of the AVCN, exhibit a primary like response, which is achieved by employment of dynamic IF Neurons with a specific parameter set for the resource consumption element. (Threshold = 30 mV, Noise Variance = 0.0001, τ_{ahp} = 300 μs, ahp-max = 100 mV, Resource consumption: max = 0.2, τ_{auf} = 1 ms, τ_{ab} = 5 ms). However, at the level of the AVCN, another task has to be performed in the model which is, the three separate channels of the ganglion cells, coding the intensity of each frequency channel, need to be recombined without loosing their effects. This is done by a membrane element, performing the spatial integration of the three separate fibers. A set of synapses is employed, to realize an equally weighted spatio-temporal integration of all PSP, generated by the spikes of the auditory nerve. To simplify computation, these synapses are included into the cell model, in fact an "Extended Dynamic IF Neuron" is employed at this stage and the membrane element is positioned within the hair cell complex as shown in figure 3.28. Therefore the AVCN model itself contains only 16 instances of the extended dynamic IF neuron as shown in figure 4.22.

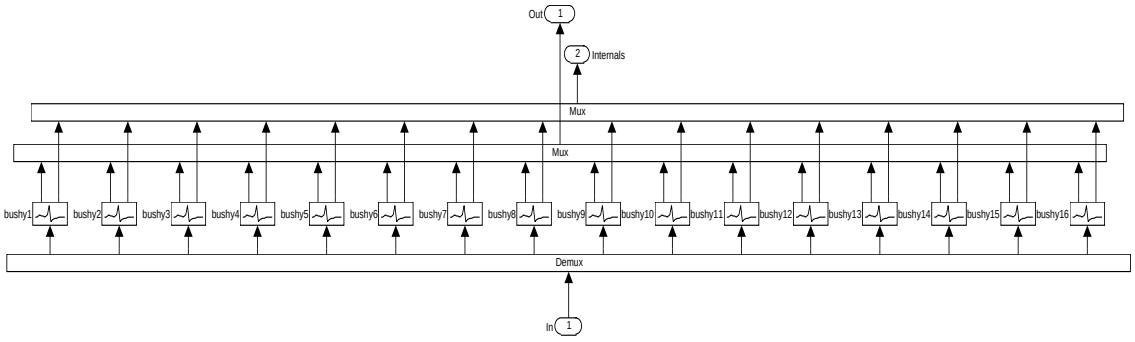

Figure 4.22: Internal structure of the AVCN model containing a separate bushy cell for each channel

The response of the AVCN model to the known sound, representing a sound source moving from the left to the right hemisphere, is shown in figure 4.23.

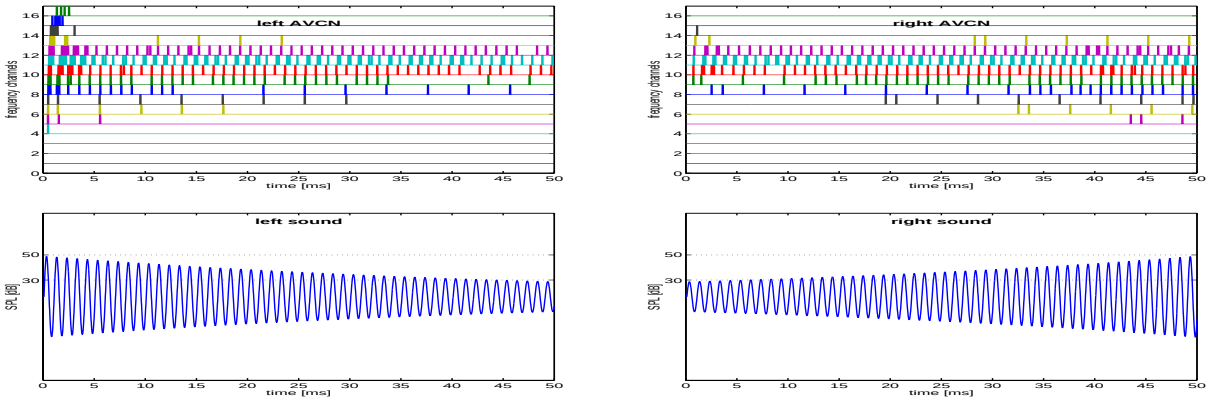

Figure 4.23: **lower panels** - *sound signals presented to the model simulating a sound source moving continuously from the left hemisphere (left louder than right channel) to the right hemisphere (right louder than left channel)* **upper panels** - *spike pattern of the left and right AVCN model*

As can be observed, the primary like response becomes only visible at the beginning of the signal, when all channels fire more intense during the first 4 ms. During the rest of the time course changes appear to be continuously and will not cause another phasic response. To illustrate the models capability more clearly, another example has been chosen and displayed in figure 4.24, where the sound signal consists of 3 parts with IID's switching stepwise in between.

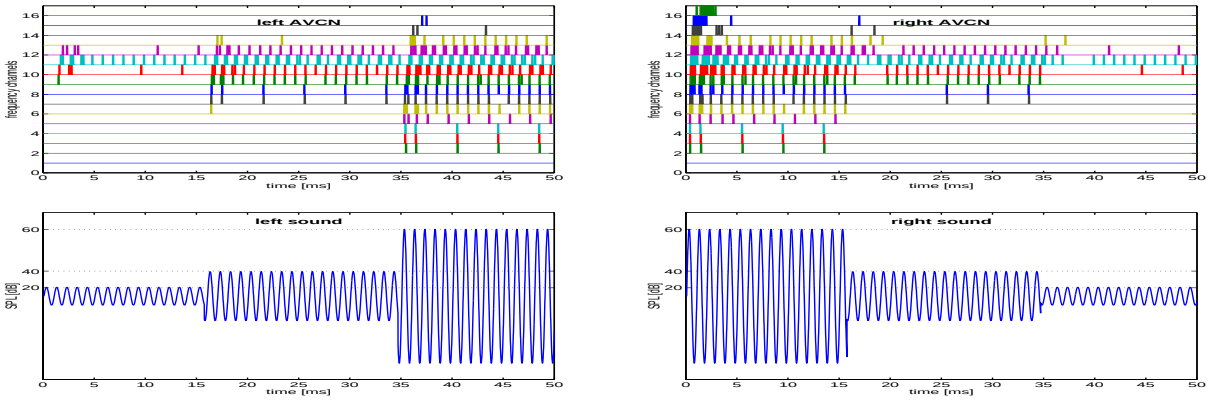

Figure 4.24: **lower panels** - *sound signals presented to the model representing three sound source locations: far right of the midline, at the midline and far left of the midline.* **upper panels** - *spike pattern of the left and right AVCN model*

At the left panels, one can now clearly identify the intense firing during the first 3-4 ms after the start of each new signal phase. Interestingly, the same change causes a "notch" at the right side, where the SPL has been decreased. This way the difference between the two hemispheres is even enhanced indicating the direction of the sudden changes more clearly.

Overall the simple dynamic principle of the primary like responses at the second model stage will be very useful to reproduce some specific properties of the precedence effect at later stages.

4.4 Stage 3 - The Superior Olivary Complex

The superior olivary complex (SOC) describes a cluster of auditory nodes in the lower brain stem perceiving its afferent inputs bilaterally from the ventral cochlear nuclei and it's efferent projections mainly from the inferior colliculus.
In contrast to the CN, the SOC is the first auditory processing region, where binaural cues become available and are processed. This connectivity predestinates the SOC to extract basic binaural cues like interaural intensity differences (IID) and interaural time differences (ITD). The localization task of the SOC has been confirmed by studies where the SOC has been selectively damaged, which demonstrated, that it is essential during the localization of sound sources. However, at this stage the localization task seems to be limited to detection of sound source directions in the azimuthal plane, rather than their exact localization in space (which would include the spherical space and the sound source distance).
As shown in Figure 4.25, the SOC contains 3 major nodes, the lateral superior olive (LSO), the medial superior olive (MSO) and the medial nucleus of the trapezoid body (MNTB). Further identified but less investigated regions are the lateral nucleus of the trapezoid body (LNTB) -not shown- and the several superior periolivary nuclei (SPN, MVPO, LVPO and DPO).

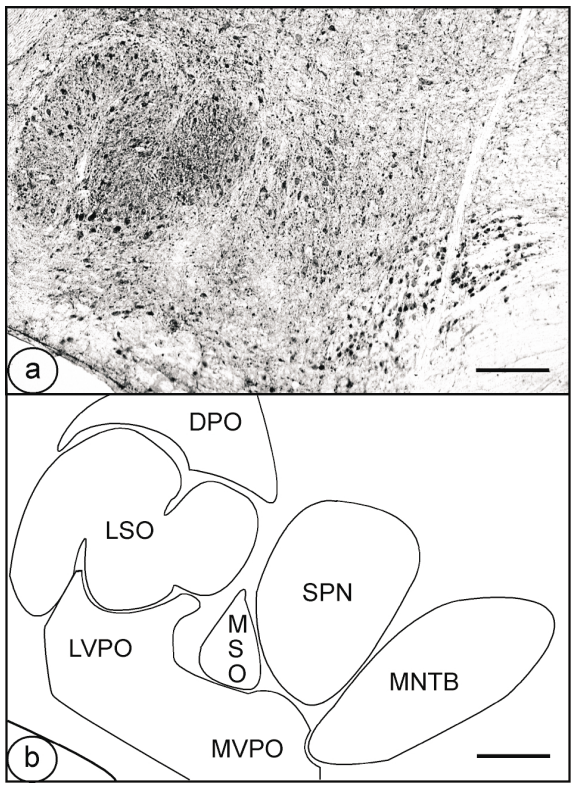

Figure 4.25: Cytological structure of the SOC in the rat, displaying the folded LSO, the drop shaped MSO, the MNTB with large cells and synapses as well as several SPN (by permission of C. Kapfer)

Since only the LSO and the MNTB are involved with the IID processing, this work will focus exclusively on these two parts of the SOC.

4.4.1 Lateral Superior Olive - Physiological Aspects

The LSO is one of the two nodes in the SOC, which are assumed to be responsible for the detection of sound source directions. In mammals it often exhibits a two folded shape, with strict tonotopic organization.
Its mainly fusiform cells are bipolar with dendrites expanding orthogonal to the tonotopic axis. About 3 quarters of the LSO cells are *bipolar principle cells*. The rest has been identified as *multipolar cells, marginal cells, small cells* and *class V cells*[HS86]. The main response type of LSO cells is most of the time chopper like and nearly independent of the sharply tuned characteristic frequencies.

Most of the LSO cells form the so called *EI units*, resulting from excitatory and inhibitory inputs generated by the two hemispheres. They receive their excitatory input directly from the bushy cells of the ipsilaterally AVCN. Their inhibitory inputs originate in the contralateral AVCN but cross the ipsilaterally MNTB before entering the LSO. The full overlapping of both for each frequency, results in interference phenomena of IPSP's and EPSP's and the typical response of EI units being sensitive to binaural intensity differences of specific frequency components. For example, a sound arriving from the left hemisphere will be more intense at the left ear than at the right ear. Therefore, excitation in the left LSO is stronger and arrives slightly earlier than inhibition, while excitation in the right LSO is weaker and arrives a bit later than inhibition - consequently only the left LSO will fire in this case. As the intensity difference increases, the firing rate will also increase, indicating the angle to the left hemifield. From here it could be concluded that if the sound arrives from the midline, excitation and inhibition to both LSO would just cancel each other out and both would remain silent. However, this is not the case, since excitation slightly exceeds inhibition and in case of a midline sound source location, both LSO will exhibit equal but some firing. If the sound source moves to the right hemisphere, the right LSO will increase firing, while the left one will finally stops to do so.

Most of the LSO-EI units are tuned to middle and higher frequencies, suggesting an important role during high frequency sound source localization, when the capabilities of localization by interaural time difference detection become limited. However, LSO neurons also react to ITD's. This is for one thing caused by the latency coding described in section 4.2.3 and is also due to the fact, that an early excitation cannot be fully removed by a late inhibition in the EI units. Both effects cause the LSO-EI cells, to exhibit at least some activity, as the ITD's become large enough.

LSO neurons project in strictly tonotopic manner directly to the inferior colliculus but send also strong excitatory projections to the contralateral DNLL and inhibitory projections to the ipsilateral DNLL [HA97]S.201 ff.[YLP96]. Those will be described in more detail during the next section.

Overall, LSO neurons are the main sites of the IID based sound source localization in the higher frequency range but also contribute to ITD based source localization over the whole frequency range. They are well developed in nearly all species but especially pronounced in species with a high frequency hearing range [HA97] S.205, [GZ96] S.737.

4.4.2 Medial Nucleus of the Trapezoid Body - Physiological Aspects

The MNTB is a relative small node within the SOC. It's cells receive their afferent inputs from the globular bushy cells of the contralateral AVCN and exhibit a precise tonotopy with low CF's in the lateral and high CF's in the medial regions. The main cell type of the MNTB are the *principal cells* exhibiting a simple, primary like response pattern. The MNTB cells project with strong inhibitory glycinergic synapses to the LSO and the MSO. This way, the MNTB realizes a temporally precise contralateral inhibition, essential to form the EI units of the LSO. The synapses within the MNTB are especially large and of fast and very effective nature. It can be assumed that they convert the arriving excitatory signals within only 100-200 μs into inhibitory ones. Beside the formation of EI Units in the LSO it has been revealed by Grothe et al. [GS93][BBM+02][KSSG02] that inhibitory inputs from the MNTB to the MSO play a critical role during ITD detection within the MSO. Here, the traditional model of Jeffress [Jef48] has been questioned, because the entire principle of ITD detection obviously depends on these inhibitory MNTB inputs and goes far beyond the well known delay line model. However, one of the major task of the MNTB cells still seems to be to provide a fast and efficient inhibitory input to the contralateral SOC.

4.4.3 LSO and MNTB Model

The third model stage, shown in figure 4.26, now adds the two LSO models to the architecture including the light green structures, representing the MNTB of both sides.

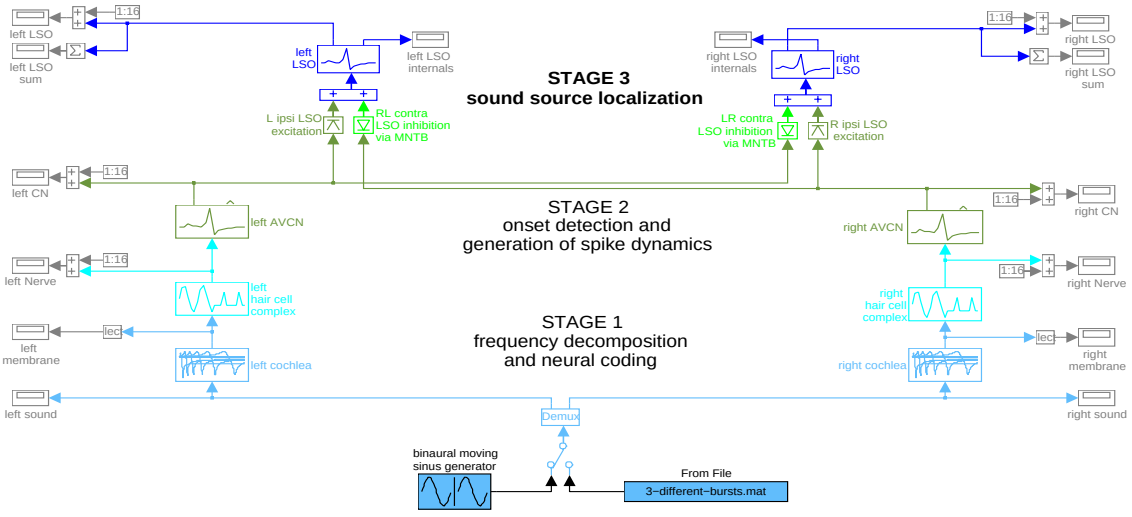

Figure 4.26: Model architecture containing the first 3 model stages including the LSO and MNTB

As can be seen, the dark green connections connect the AVCN ipsilaterally to the LSO cell membrane via the excitatory (raising triangle) dark green synapses. From the contralateral side the inhibitory (falling triangle) light green synapses represent the MNTB and project also to that membrane, where the IPSP - EPSP interference will happen. Each symbol contains an array of 16 synapses for each frequency channel, but set to identical parameters. These are for the excitatory projections: $synaptic\ weight = 0.24mV$, $\tau_d = 0.2ms$ *and* $\tau_r = 5ms$. The inhibitory MNTB synapses are set to $synaptic\ weight = 0.2mV$, $\tau_d = 0.2ms$ *and* $\tau_r = 5ms$. Here, the slightly more efficient excitation becomes visible in order to assure an LSO response for sounds originating from the midline.

The response of this LSO model to the sound source moving from the left to the right is again displayed in figure 4.27. Here, it can be observed that the two LSO responses counteract each other. After the initial phasic response, clearly indicating a sound arriving from the left hemifield, the firing intensity of the left LSO gradually decreases, while the response of the right LSO gradually increases as the sound moves towards the right. At the $25ms$ point, the sound orriginates from the midline and both LSO models act about equal. Particularly at the right LSO, the angle coding can be perceived very clearly. The further right the sound source is positioned, the higher is the spike intensity and the more frequency channels are recruited.

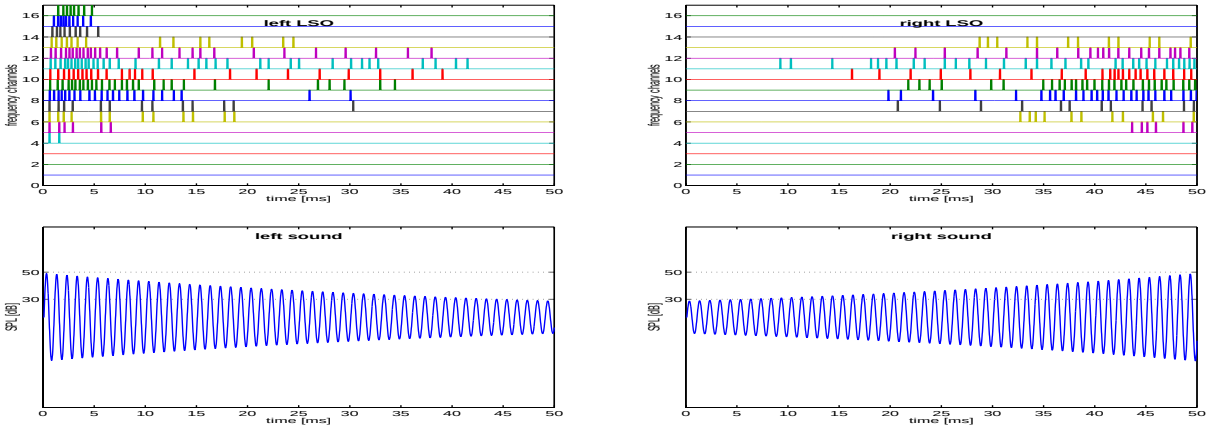

Figure 4.27: **lower panels** - *sound signals presented to the model at the left and right ear*
upper panels - *spike pattern of the left and right LSO model*

Beginning from the level of the LSO, the model contains two more display elements for each stage. These are visible in figure 4.26 as gray rectangles named “LSO sum“ and “LSO internals“. Their output is shown in figure 4.28.

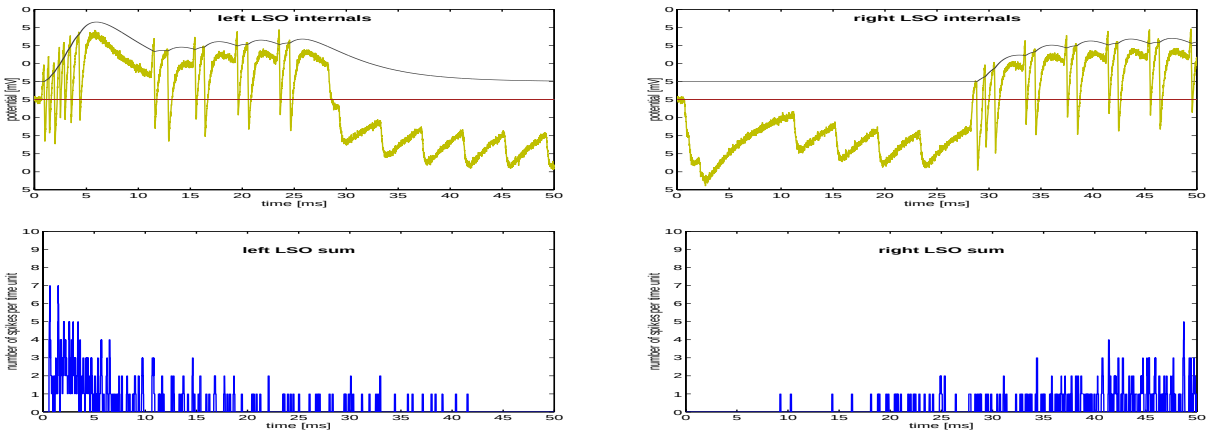

Figure 4.28: **upper panels** - *internal parameters - soma potential and firing threshold of the dynamic LSO cells representing the best matching frequency channel 13 of the left and right hemisphere*
lower panels - *spike count of the left and right LSO, summed over all frequency channels for each Simulation Time Step (STS)*

Here, the lower panels show the LSO sum of both sides, where the single spikes of all channels are summed at each simulation point, resulting in a PSTH like impression of the firing intensity. The upper panels display the time course of the soma potential (ochre) and firing threshold (black) of two corresponding LSO cells, representing the best matching frequency (BMF) channel of both sides.

In conclusion, the LSO model accounts for the decoding of IID's and based on this, the pure sound source localization in the azimuthal plane. However, there is no echo suppression at this stage and the LSO will detect every echo as a new sound arriving from a different direction. In order to overcome these limitations, a model of the DNLL is added in the next section.

4.5 Stage 4 - The Lateral Lemniscus

The lateral lemniscus (LL) has become subject to intensive studies just very recently. For a long time its role in auditory processing was unclear and remains a mystery to some extended even today. One of the commonly agreed roles of the lateral lemniscus is to transfer auditory information from the CN and the SOC toward the IC. However, intensive studies during the 80'th and 90'th of the last century revealed that there are obviously significant processing areas within the this auditory structure. The lateral lemniscus consists of three sub-nuclei shown in figure 4.29, the *dorsal nucleus of the lateral lemniscus (DNLL)*, the *intermediate nucleus of the lateral lemniscus (INLL)* and the *ventral nucleus of the lateral lemniscus (VNLL)*.

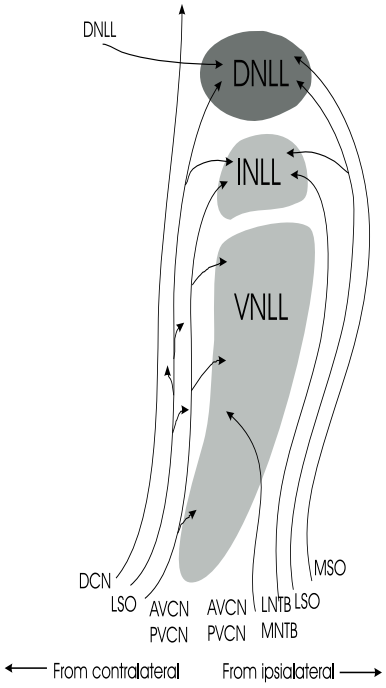

Figure 4.29: Structure of the lateral lemniscus containing a dorsal part (DNLL), an intermediate part (INLL) and a ventral part (VNLL)

However, most studies have only found a ventral (VNLL) part and a dorsal (DNLL) part. The third, intermediate cell group has been identified so far only by some studies in the cat, rabbit and bat. Presently, there is still very little data on humans, [Moo87] but it has been found that the VNLL is poorly, while the DNLL is well developed in humans.

The VNLL and INLL shall be mentioned here only briefly. In cats and rodents the VNLL is usually recognized as a single group of cells with mixed characteristics. However, in bats a lateral and a medial region can be distinguished, with columns of spherical cells in the lateral part, and multipolar and round cells in the medial part. The main response type of the VNLL is again primary like. It's tonotopic organization has been found to be 3-dimensional with multiple clusters throughout the dorsoventral extent of the VCLL. The shape, size, and location of the clusters suggest an interdigitation of clusters assigned to different frequency-band representations. However, at the same time an overall mediolateral distribution gradient was observed, with high frequencies represented medially and lower frequencies more laterally [MLB+98].

As shown in figure 4.29, the VNLL receives it's major inputs from the contralateral VCN, with a much smaller projection from the ipsilateral cochlear nucleus restricted to a relative small region within the VNLL. In the cat and the gerbil, VNLL neurons are known to send large diameter axons to the IC.

The INLL, in the cat, is composed of sparsely distributed multipolar cells and clusters of horizontal cells among the fibers of the lateral lemniscus [GBHM81]. The cells of the INLL are reported to have little similarity to those of the DNLL or VNLL and display mostly chopper like response properties. They receive their major inputs from the contralateral CN and LSO. In addition, ipsilateral input is received from the LSO and the MSO. The INLL also receives substantial input from the MNTB [GBHM81].

So far, little is known about the functional roles of the VNLL and the INLL and since they don't seem to be involved in the echo suppression task, they have not been modeled in this architecture.

4.5.1 Dorsal Nucleus of the Lateral Lemniscus - Physiological Aspects

The dorsal nucleus of the lateral lemniscus (DNLL) has been studied more intensely during the last years and is now believed to play a crucial role during echo suppression. It is therefore introduced with greater detail, including some physiological experiments.

The DNLL is composed of neurons distributed among the fibers of the lateral lemniscus just ventral and caudal to the IC. Regarding it's tonotopic organization, quite different results have been obtained by different experimenters. While Aitkin [AAB70][AIW84] saw a clear tonotopic organization, with dorsal neurons most sensitive to low frequencies and more ventral cells sensitive to higher frequencies in the cat, Yang [YLP96] found two subdivisions of the DNLL of the mustache bat, where the frequency representation in the posterior division was from about 15 to 120 kHz, whereas in the anterior division it was only up to 62 kHz. Bajo [BVdRR98] found a loose tonotopic organization in the rat, displaying a concentric pattern with high BF units located in the most dorsal and ventral parts of the DNLL and lower BF units in the middle part of the nucleus. More recently, Bajo [BMM$^+$99] suggested a laminar organization with rostrocaudally oriented lamina representing a specific BMF and resembling a flattened tube. He found the low frequency lamina located in the dorsolateral corner of the DNLL and the high frequency regions at the ventromedial surface of the DNLL. During the recordings of this study in the gerbils DNLL, a laminar structure, with low frequencies in the external and high frequencies in the central regions seemed to be present.

Whatever is the right hypothesis, it seem clear that there is a tonotopic organization of the DNLL, which has been also incorporated into the model by the separate projection of each frequency channel originating from the lower nuclei.

The current opinion on DNLL cell response types is also diverse. Markovitz [MP93]found 58% of DNLL neurons to exhibit a chopping pattern evoked by contralateral stimulation and distinguished EI cells with largely chopper response from EE cells, exhibiting more often primary like or primary like-with-notch response types. Furthermore, he found that the sustained responses of EE units to contralateral stimulation differs dramatically from their onset responses to ipsilateral stimulation. He concludes that the DNLL contains a heterogeneous population of neurons based on physiological responses to pure tones.

Yang [YP97] found 55% of the DNLL neurons to react with sustained responses (throughout the signal duration) and 45% onset neurons, (responding only at the beginning of the signal). Sustained and onset neurons responded differently to amplitude modulated signals. Signal intensities affected the phase-locking of sustained and onset neurons differently. Sustained neurons exhibited tight phase-locking only at low intensities, 10-30 dB above threshold. Onset neurons, in contrast, maintained strong phase-locking, even at relatively high intensities.

Similarly Kelly [KBK98] found three response categories within the DNLL: onset (57%), sustained (21 %) and onset-pause-sustained (22%). Most DNLL neurons fired multiple action potentials to a single click at the contralateral ear and the majority (77%) of DNLL neurons showed a monotonic increase in the number of spikes elicited by contralateral tone pulses of increasing SPL.

Based on these findings and observations in the Gerbils DNLL, the DNLL model cells have been tuned to exhibit a chopper like sustained response, ignoring the large portion of the pure onset cells. The reason for ignoring them is that they will not be able to contribute to a continuous localization of sound sources after the first wave front, which is the goal of this model.

A significant number of experiments during the last decade was also concerned with determining the percentage of EI units within the DNLL as well as their site of origin. In principle, there seem to be only two options. Firstly, EI units response properties could be simply transferred by the excitatory projections reaching the DNLL from the contralateral LSO with own EI units. Or secondly, EI units in the DNLL could be formed *de novo* by the interaction of contralateral excitation and ipsilateral inhibition from the two LSO in conjunction with contralateral inhibition from the opposite DNLL.

The results of many experiments concerning the creation of EI Units [PPWL92], the role of the GABAergic contralateral inhibition from the opposite DNLL [IvK96],[CKW99], [vAKK99] and the general input structure to the DNLL [YP94c],[YP94b], [YP94a],[KLv96] suggest the following:
EI response properties hold for a substantial part (88% [MP93], 74% [KBK98]), but by far not all units in the DNLL. In principle, they are simply transferred, via the commissure of probst from the opposite LSO, but their response properties are modified in the early and ongoing part by the glycinergic inhibition from the ipsilateral LSO and later by the GAGBergic inhibition from the contralateral DNLL. Kelly [KK00] suggests that the AMPA receptors contribute selectively to the initial component of binaural inhibition and the NMDA receptors to a longer lasting component.

The inhibitory circuitry between the two LSO and the two DNLL nuclei results in the previously described *push-pull effect*, causing a *persistent inhibition* of DNLL-EI units in case of preceding sounds arriving from the ipsilateral hemifield. A detailed description of the push pull circuit has been provided in section 2.3.3. Persistent inhibition in DNLL-EI neurons has been identified by Yang and Pollack [YP94c] or more recently by Pollak [PBP$^+$02],[Pol02]. A good review on the current knowledge on the DNLL connectivity and its functional role is provided once more by [PBK03].

4.5.2 DNLL Model

The primary task at this stage of the model is to provide a functional and physiological plausible model for the generation of persistent inhibition effects within the DNLL-EI units. It is therefore restricted to this cell type and again remains with the tonotopic organization seen in the AVCN and the LSO models.

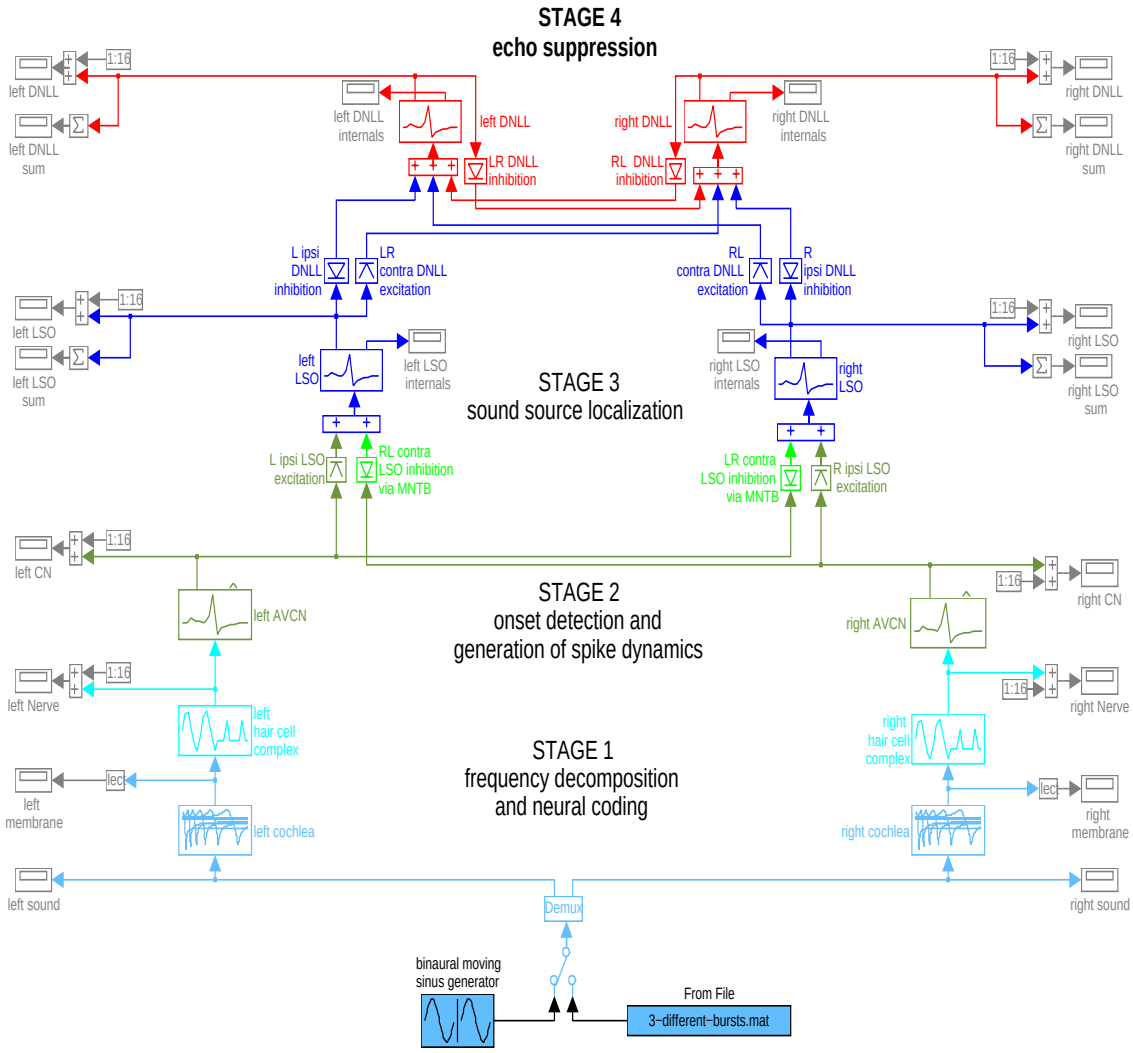

Figure 4.30: Model architecture containing the first four model stages including the model of the DNLL

Shown in figure 4.30, the DNLL receives excitatory input via the blue synaptic arrays (raising triangles) from the contralateral LSO. Glycinergic inhibitory inputs from the ipsilateral LSO, reach the DNLL cell membrane via the second blue synaptic array (falling triangles). Finally between the red structures of the DNLL cells contralateral GABAergic inhibitory projections are represented by the red connections, containing a inhibitory synaptic array on each side.

Both DNLL models are composed of 16 "Dynamic IF Neuron" NBL elements, each tuned to chopper response by the parameter set: *threshold at 5 mV above resting potential (0.05),noise variance = 0.0001, after hyper polarization - time constant 0.3 ms and maximum 50 mV, resource consumption: max = 0.5,* τ_{auf} *= 0.1 ms,* τ_{ab} *= 0.3 ms*

The most important question at this stage is to set the synaptic time constants in a physiological manner assuring the appropriate persistent inhibition response. According to Jack Kelly (by e-mail communication), the rise time of GABAergic IPSP within the DNLL ranges between 5 and 7 ms leading to time constant of 1-1.5 ms and the decay time is about 30 - 40 ms, leading to time constants 6-8 ms. Glycinergic IPSP in the DNLL are usually slightly faster and have been studied by Wu [Wu98] and Wu and Kelly [WK95]. Their results suggest, but may not prove, glycinergic IPSP durations of up to 25 ms and much shorter EPSP around 10 ms which where also less efficient. Rise time ranges in both cases between 0.5 and 1 ms. The parameters used in the model are based on these observations: *contralateral excitatory LSO input: synaptic weight:*$5mV$, $\tau_d = 0.2ms$, $\tau_r = 2ms$ *ipsilateral glycinergic LSO input: synaptic weight:*$10mV$, $\tau_d = 0.1ms$, $\tau_r = 5ms$ *and contralateral GABAergic DNLL input: synaptic weight:*$30mV$, $\tau_d = 0.2ms$, $\tau_r = 7ms$ (see Appendix A for overview on model parameters). Using this parameter set, the response to the moving stimulus is displayed in figure 4.31.

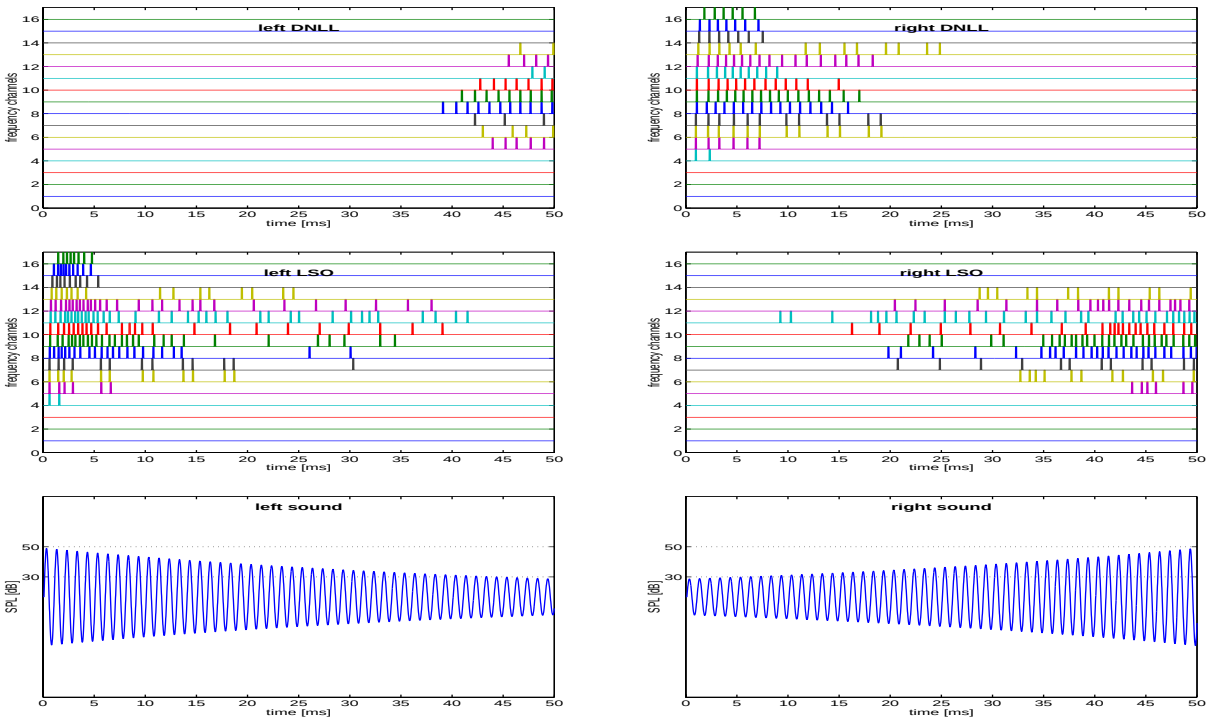

Figure 4.31: **lower panels** - *sound signals presented to the model at the left and right ear*
middle panels - *spike pattern of the left and right LSO model*
upper panels - *spike pattern of the left and right DNLL model*

As can be seen, the right DNLL, despite his chopper like response, fairly well resembles the spike pattern of the left LSO as long as the signal arrives from the left hemisphere. However, the left DNLL does not so with regard to the right LSO. It's response starts about $20ms$ later than the one of the right LSO. The reason is - persistent inhibition evoked by the period of ipsilateral leading sound from the left. Why the response of the left DNLL starts so late becomes evident, by looking at the internals of the channel 8 neuron of the both DNLL models in figure 4.32.

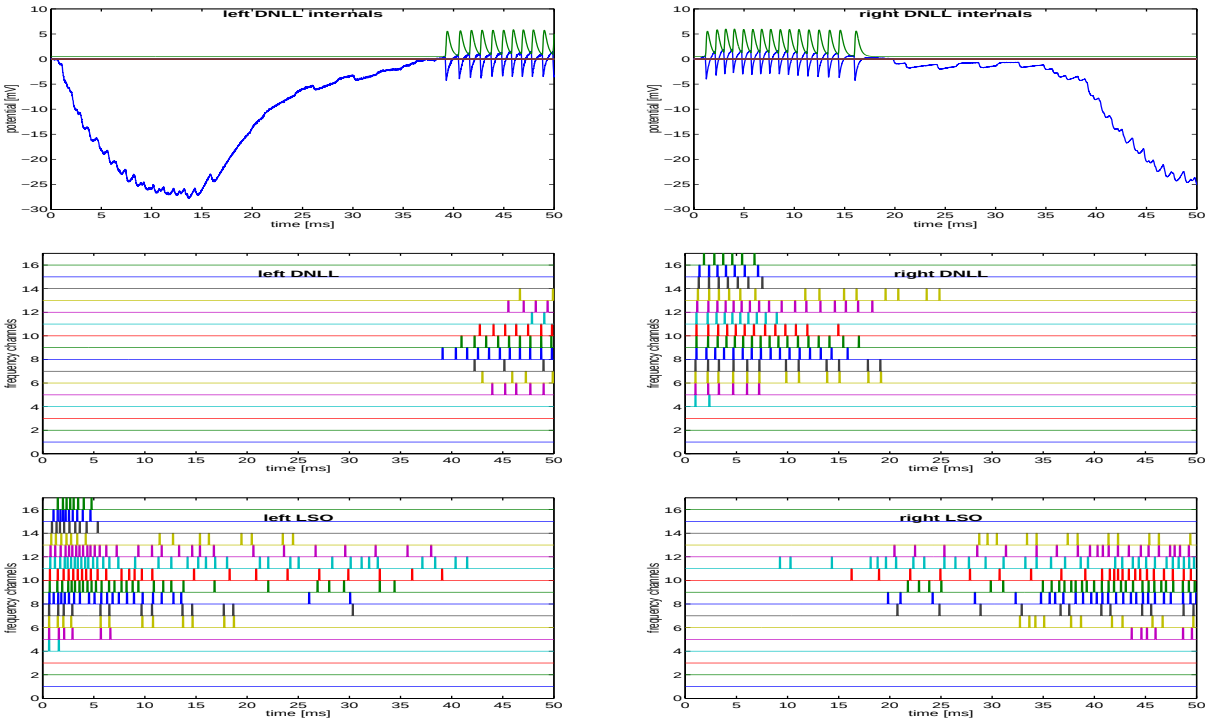

Figure 4.32: **left upper panel** - *soma potential and threshold of the left DNLL cell* **right upper panel** - *soma potential and threshold of the right DNLL cell* **middle and lower panels** - *responses of the LSO and DNLL models respectively*

As can be seen, both DNLL cells start at the resting potential. While the right DNLL cell receives only pulling excitation from the left LSO, it's left counterpart is pushed by inhibitory input from the left LSO and the right DNLL down to $27mV$ below resting potential (to about $77mV$ hyperpolarization). Here, it saturates due to the hyperpolarization limit of that cell. As the input condition changes gradually to favor the left DNLL, this has first to recover from that strong hyperpolarization and needs until the $40ms$ point to reach back to its resting potential and to exceed the firing threshold.
On the opposite site, the left LSO stops about after $15ms$ to further excite the right DNLL, and the DNLL consequently stops firing, but stays near the resting potential. At the $20ms$ point the right LSO starts gradually to fire and to hyperpolarize the right DNLL slightly, this remains until the $40ms$ point, when also the left DNLL starts firing and now strongly inhibits the right DNLL via it's contralateral inhibition. Now, the right DNLL is pushed downward toward the $-75mV$ hyperpolarization limit, while the left DNLL has been successfully pulled back and fires.

The push pull effect, shown in figure 4.32, is the key feature to generate persistent inhibition within the DNLL model and is realized here, by the specific connectivity of the DNLL and the exact timing of excitatory and inhibitory spikes within this part of the model. It will be further investigated in the next chapter 5 "Experimental Results".

4.6 Stage 5 - The Inferior Colliculus

The inferior colliculus (IC) is one of the most interesting and intensively studied areas of the auditory brainstem. It is the first auditory structure where both binaural and monaural cues as well as the timing and the intensity pathway converge. Regardless of their origin and character, all auditory signals to the auditory cortex will need to pass the inferior colliculus. Therefore the IC is one of the most important centers of auditory processing and signal evaluation. Since it is relatively easy accessible in many animals, a significant number of studies has been performed concerning its physiological structure and functional role. The evaluation of all those experiments would lead far beyond the scope of this work and will therefore be limited to those concerning the influence of the DNLL with regard to specific cell types within the central inferior colliculus and their excitatory inputs.

4.6.1 IC - Physiological Aspects

Based on the physiological structure of the IC, it can be divided into two major regions the central area (ICc) and the external area (ICx) (see figure 4.2 on page 84). The external area i sometimes subdivided into a dorsal and the truly central part [MRLB95]. But the main input is obviously received in the central region of the IC, directly or indirectly from the brain stem structures at the contralateral side.

The large number of different inputs and different feature representations suggests a quite diverse response pattern of the individual IC neurons. In fact, this is the case, when looking at individual cell responses in the IC. The differentiation of response pattern, first time introduced in the CN, reaches a new quality in the IC. Here, chopper type responses with build up and pauser features, long latencies, nonlinear intensity curves and tuning curves with complicated receptive fields have been identified. Variations of all those types are distributed throughout the IC with seemingly no topological order. It is important to notice that many cells exhibit variable response properties, depending on the specific acoustical scene or even the overall condition of the organism.

The only organization principle, agreed upon on so far, is the well known tonotopic organization of frequencies. However, in the IC it is rather a laminar structure with high frequencies represented more ventromedial and lower frequencies more dorsolateral [KLvAI98].

Despite the isofrequency lamina, a number of other maps have been identified by some experiments, but so far have not been agreed upon by others, using different animals or simply different experimental setups. These are: *sensitivity maps*, *maps of modulation frequencies*, *auditory space maps* [ER97],[HA97] and *IID maps* [PP93b]. Three of them are displayed in figure 4.33, based upon the review of Helfert and Aschoff [HA97] in 1997.

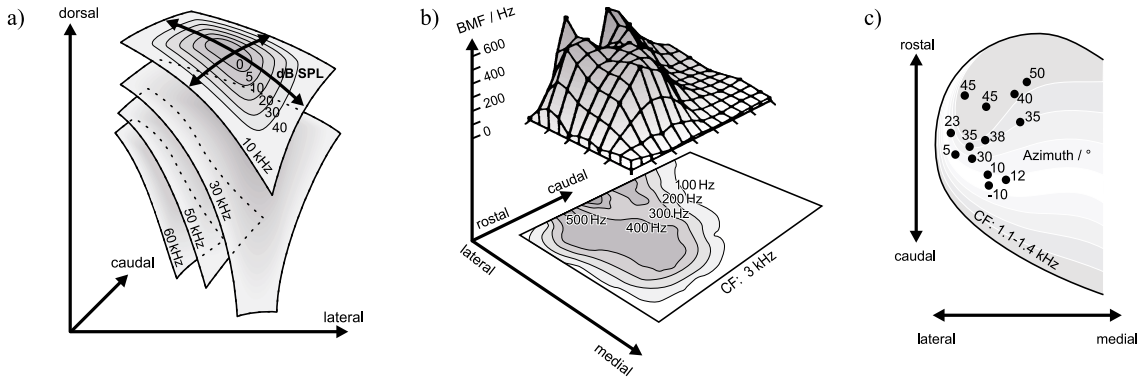

Figure 4.33: Feature maps of the ICc: a) iso-frequency lamina in the IC of mice, b) modulation frequency map within a frequency lamina in mice c) azimuthal map identified in the cats IC. (from [HA97] pg.287, 293 ff).

The afferent connectivity of the IC is comparably complex, receiving inputs from nearly all regions of the contralateral auditory brainstem. However, the connections of interest to this study are the well proven contralateral excitation from the AVCN, the driving inputs from the contralateral LSO and, most importantly, the bilateral inhibitory projections from both DNLL.

In the following, some experiments regarding the influence of the DNLL to specific IC units will be discussed with greater detail in order to justify the chosen model architecture.

Already in 1993 Park et al.[PP93b] studied the influence of GABAergic inhibition from the DNLL onto EI units in the IC by reversibly blocking GABA with biculline. They identified 70% of the IC cells to be conventional EI units and found that the collicular map of IID sensitivity is formed to a substantial degree in the colliculus by GABAergic innervation, which likely originates in the DNLL. This GABAergic innervation contributes to the establishment of the IID map, in at least three ways. The first way is the creation of the EI properties in some collicular cells through the convergence of excitatory and GABAergic inhibitory inputs. A second way occurs in other cells where GABAergic inputs adjust the neuron's sensitivity to IIDs. And a third way occurs in yet other cells, in which the inhibition produced by inputs from the ipsilateral ear was increased. Another implication of Parks study is that GABAergic inhibition within the colliculus appears to create or reinforce binaural facilitation in most collicular EI units. In a subsequent study [PP94] they added that in IC-EI units, in which the ipsilaterally evoked inhibition was reduced or abolished by bicuculline, the contralateral DNLL most likely was the source of the inhibition.

At the same time Pollack et al. [PP93c] suggested that GABAergic inhibition acts on collicular cells in two principal ways. The first way, is to modify the effects of the excitatory innervation and thereby shape the response features of collicular neurons. The second way, is to provide a regulated suppression of evoked activity. They propose that the suppression is situation dependent and may act to enhance the operating range of collicular neurons in situations of particular importance to the animal, such as during periods of selective attention.

A finding of special interest for this study has been published by Kelly and Li in 1997 [KL97]. Here, the inhibitory influence of the two DNLL to the rats IC has been separately removed by disabling first the contralateral DNLL and later the ipsilateral one. They found that although both DNLL projected to the IC [KLvAI98], an injection into the contralateral DNLL greatly reduced the response suppression produced by stimulation of the ipsilateral ear. Injection into the ipsilateral DNLL, however, had no effect. Injection into the ipsilateral SOC reduced the amount of binaural suppression, but the effect was apparent only in cases with surgical transection of the contralateral lateral lemniscus at a level below the DNLL. These data support the conclusion that binaural responses in the rat's ICC are shaped by inhibitory projections from the contralateral DNLL and contralateral excitatory projections from the LSO and the AVCN.

Klug et al. [KBP99]finally investigated the timing of excitation and inhibition within IC- EI units receiving inhibition from the DNLL. In 80% of the cells that were inhibited, the inhibition by the ipsilateral ear and contralateral excitation were temporally coincident. In many of these cells, the inhibition suppressed contralateral discharges and thus generated the cell's EI property in the ICc. Finally, in the majority of cells, the ipsilateral induced inhibition persisted for tens of milliseconds beyond the duration of the signal that evoked it. Thus inhibition by the ipsilateral ear has multiple components and one or more of these components are typically evoked in ICc neurons by sound received at the ipsilateral ear.

Based on these and other studies Pollack summarizes in 1997 [Pol97]:

1. that the DNLL shapes the binaural properties of many inferior collicular neurons,

2. that the inhibitory inputs to the DNLL allows it to act as a switch to the IC that can be turned on or off with appropriate acoustic stimulation, and

3. that when two or more stimuli are presented, each from a different region of space, the first stimulus can switch the DNLL to its off position. The consequence of the initial stimulus is that stimuli that follow shortly thereafter cannot activate the DNLL, and thus the binaural properties of those IC cells, which receive inhibition from the DNLL are changed.

The implications of this switching action are that the location of the initial signal is coded appropriately, whereas the coding of the location of the signal or signals that follow the initial signal is smeared, and consequently, those following signals cannot be accurately localized. In short, it is proposed that the DNLL plays a pivotal role in the way the locations of multiple sound sources are coded by the auditory system.

Since these experiments clearly support the hypothesis of DNLL cells to generate the EI properties of at least some cells in the central IC and modulate them in case of preceding sounds, the model architecture includes the connections mentioned above and models these specific IC-EI cells only.

4.6.2 IC Model

Adding the fifth stage to the model architecture finally leads us to the full model shown once again in figure 4.34.

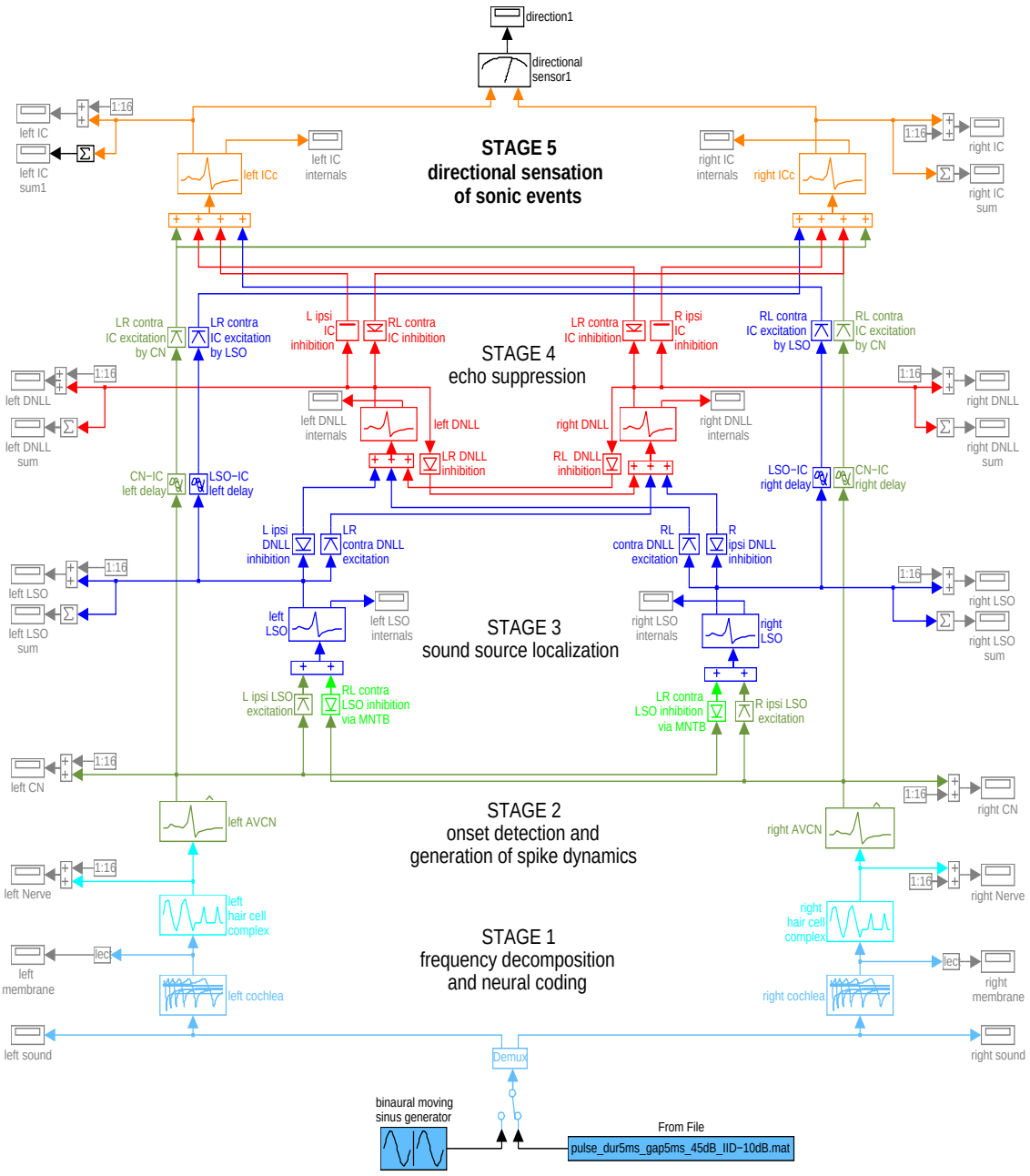

Figure 4.34: Full model architecture containing the five model stages including the model of the IC

As can bee seen , the orange IC model stage receives multiple inputs. First of all it is innervated by the green excitatory projections from the contralateral AVCN via a axonal delay. The reason for introducing a axonal delay element at this point is that the direct connection from the AVCN has to cross a significant physiological way through the entire brainstem, while the DNLL inputs arrive from areas right below the IC. Furthermore, the processing stages at the LSO and the

DNLL introduce each about 1-2 ms signal delay by the nature of the IF model. Since Klug [KBP99] found excitation and inhibition to arrive temporally coincident, an axonal delay set to 3 ms in both AVCN inputs, recreates this situation fairly well by compensating for the processing delays of the LSO and DNLL.

The synaptic properties of the AVCN inputs are seen to be quite efficient and their parameters have been set to: synaptic weight = 20 mV, τ_d = 200 μs,τ_r = 3 ms. The formerly employed dendritic decay element has been incorporated into the synaptic array, since at this point we assume a direct connectivity with no compartmental dendrite involved.

Secondly, the IC model perceives excitatory input from the contralateral LSO via the blue synapses. Here, the synaptic efficiency is set to be smaller (synaptic weight = 4 mV, τ_d = 200 μs, τ_r = 2 ms) and contains a delay element of 2 ms, reflecting the shorter distance between LSO and IC as well as the smaller processing delay, caused only by the DNLL.

Finally, the red synapses realize the GABAergic inhibitory inputs from both DNLL. Although the DNLL projects bilaterally to the IC model, the findings of Kelly [KL97], described above, clearly revealed the much more critical role of contralateral inhibition compared to the ipsilateral connections. Modeling this, the efficiency of the contralateral synapses has been set to high values of: synaptic weight = 50 mV, τ_d = 200 μs, τ_r = 5 ms, while the ipsilateral projections are set to: synaptic weight = 2 mV, τ_d = 200 μs, τ_r = 5 ms. This way, the main inhibitory influence of IC cells by the DNLL occurs contralaterally, as can be observed by the size of the synaptic triangles in figure 4.34.

The orange model of IC - EI units in the inferior colliculus only strives to duplicate the properties of the EI units, created de novo in the IC by DNLL inhibition and AVCN Excitation. This specific model will exhibit the appropriate response with nearly every physiological parameter set and has been realized by the use of "Dynamic IF Neurons" with chopper response and the parameters: threshold = 20 mV, noise variance = 0.0001, τ_{ahp} = 500 μs, ahp_{max} = 100 mV and resource consumption: max = 1, τ_{auf} = 300 μs, τ_{ab} = 500 μs.
The response of the entire architecture including the IC model to a typical stimulus, also used during physiological experiments is shown in figure 4.35.

The example of figure 4.35 exhibits several aspects of the echo suppression principle realized with this model.

The sounds presented to the two ears both contain 3 pulses of 5 ms duration each with 10 ms gaps in between. While the first pulse arrives from the left hemisphere (10 dB louder and 200 μs earlier at the left channel), the subsequent two are designed to arrive from the right hemisphere (10 db louder and 200 μs earlier at the right channel). It can be assumed that the first pulse represents the original signal and the later two from the right are echoes of that first pulse normally disturbing the localization. While the first pulse arrives within the echo threshold (at 20 ms) and should be suppressed, the second pulse arrives outside and should therefore not be affected.

According to the intensities of the pulses the left and right AVCN model will respond with more spikes and more channels involved if the sound is louder and vice versa.

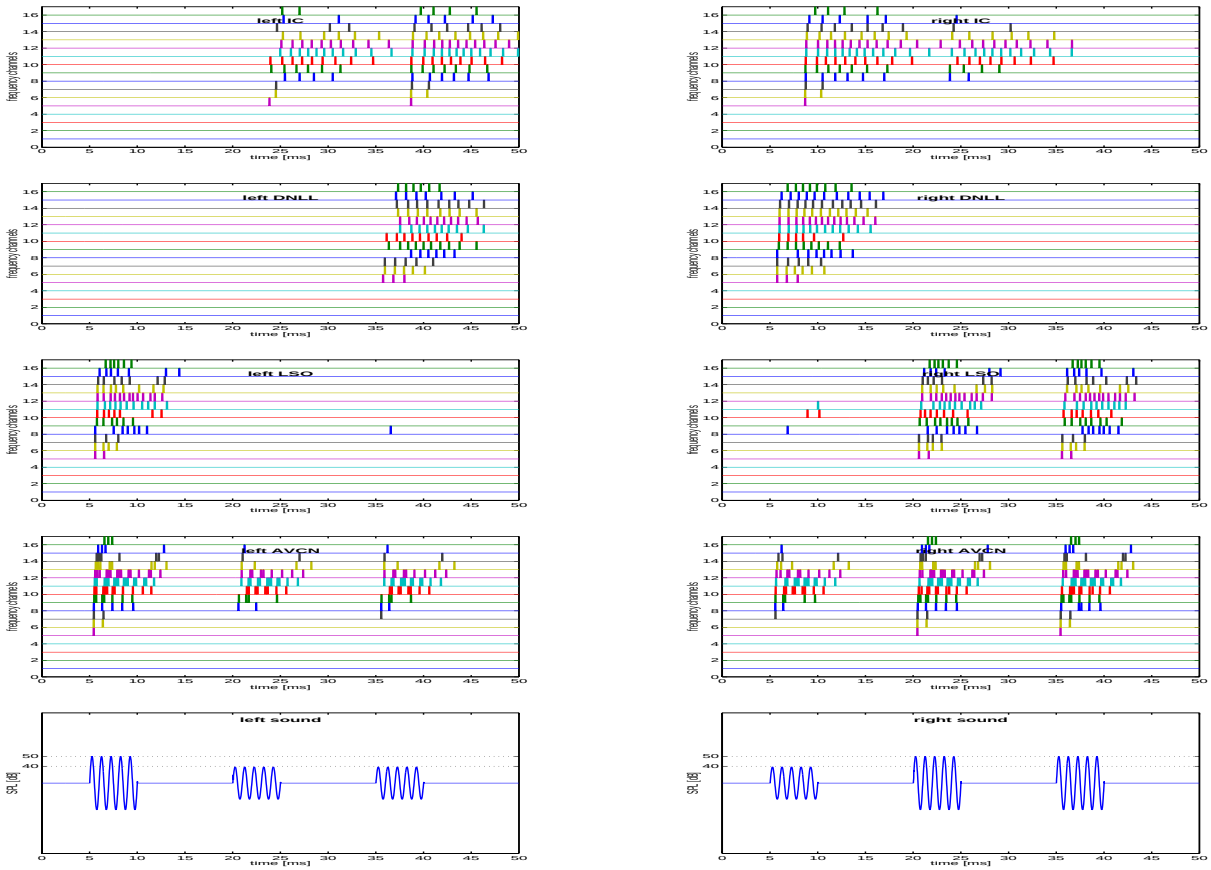

Figure 4.35: **bottom panels** - *3 sound pulses presented to the model, with switching IID of left 10 dB to right 10 dB* **second line panels** - *responses of the AVCN models reacting with a different number of spikes and channels according to the intensity differences* **third line panels** - *responses of the LSO models discriminating the direction of arrival* **fourth line panels** - *responses of the DNLL, where the second pulse generates no response due to persistent inhibition generated by the first pulse* **top panels** - *response pattern of IC model, where the first and the last pulse are dedicated to a specific hemisphere but the second one shows up at both IC's and therefore contains no directional information*

Based on these differences the LSO model can clearly distinguish the direction of arrival and consequently responds to the left pulse at the left LSO and to the two right pulses at the right LSO. Note that the response pattern to each of the pulses is fairly similar regardless of the hemisphere it is arriving from.

This is no more true at the level of the DNLL. Here, the first pulse is responded appropriately at the right DNLL (due to the contralateral excitation), but *the second pulse is not responded to at all!* The reason is that the first pulse (original sound) generates persistent inhibition within the left DNLL, and when the first echo arrives after 10 ms most channels are still inhibited and therefore do not respond. This again changes for the third pulse arriving after 20 ms and therefore outside the echo threshold. The third pulse is responded at the opposite DNLL, mirroring the response of the right LSO since the persistent inhibition has disappeared now. In other words, the DNLL suppresses the echo response as long as it arrives within a time frame of about 20 ms after the original sound.

The response of the IC model finally contains the second pulse response again. This has been expected, since it is driven by the AVCN responding to all three pulses at both sides. However, the inhibition applied by the DNLL generates also EI properties here, at least in case of the first and last pulse. During the first pulse, the right DNLL inhibits the left IC and this remains silent, while the right IC is not inhibited and responds to that sound from the left. However, during the second pulse, the situation is different. Neither one of the DNLL inhibits its target IC cells and the second pulse generates an equal response in both hemispheres. Hence, *The echo is perceived at the level of the IC, but doesn't carry any directional information any more.* Finally, during the third pulse, the left DNLL inhibits the right IC and the EI properties are seen to be restored in case of echoes arriving later than the echo threshold.

This mechanism, displayed by the response pattern to this fairly simple dynamic signal, has proven to be very robust, even for changing model parameters as long as the principles are maintained and they are not set outside the physiological range. It's functionality is based on connectivity aspects as well as the exact timing of single spikes. As a whole, it can duplicate a substantial number of psycho-acoustic experiments, as will be shown in the next chapter, and seems likely to account for the first stage of echo suppression within the natural auditory system. However, to make this useful for technical applications a technical output element - the directional sensor - is needed.

4.7 Directional Sensor

The task of this final model element, shown at figure 4.36 is, to generate a continuous signal out of the spike pattern, capable to function as a control unit for technical systems - for example the motors of a robot head asked to turn toward the direction of a sound source.

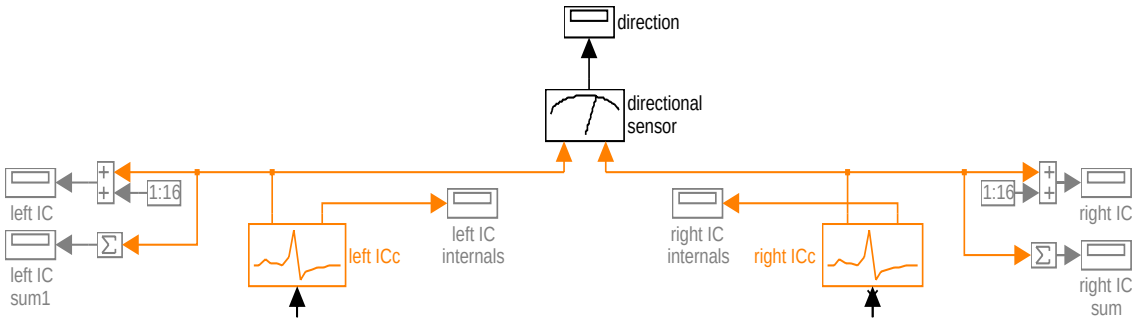

Figure 4.36: Directional sensor - generating the model output

Since this needs to function also in case of strong echoes, only the level of the DNLL or IC can be used to generate this output. Here, the IC level has been chosen, since at the level of the DNLL, only directional information is represented, while at the IC the full echo content is present, even if the directional information is suppressed. Therefore in any case, the IC models are activated and it is the difference in overall activation, caused by the inhibitory influence of the DNLL, which guides the sensation of direction within the output element.

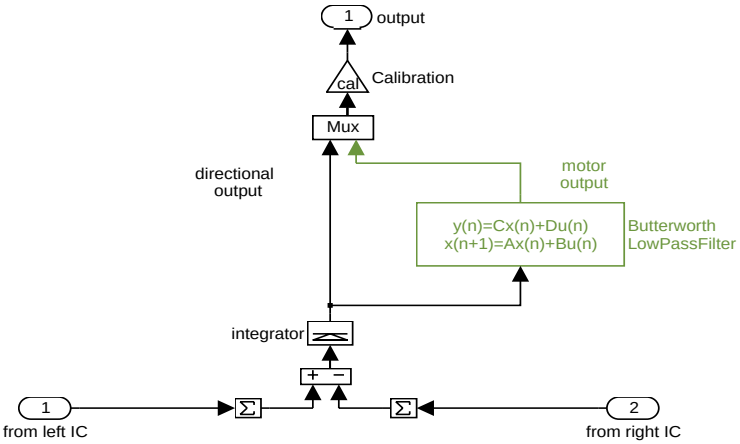

Figure 4.37: Elements of the directional sensor - generating the directional and the filtered motor output

The model structure used to generate the directional signal is shown in figure 4.37. It needs to perform two tasks: First, it has to integrate all spikes within the different frequency channels. Second, it has to generate a time continuous dumped signal, out of the discrete spike events. The structure used to perform these tasks is fairly simple.

It consists of two summing elements - summing the spikes of all frequencies at each hemisphere, a substractive element - generating the difference between the spike counts in both hemispheres -, an integrator - employing a static synapse model to generate the fairly continuous directional signal out of the positive and negative pulses and finally a calibration factor used to calibrate the model to the specific shadowing effect between the two microphones of the employed setup. By selecting a parameter set of $weight = 0.4$, $\tau_d = 500\mu s$ and $\tau_r = 70ms$ for the integrator and a calibration factor of 3.0, a fairly dumped signal is generated indicating sound source locations in the horizontal plane between -90 degrees left and +90 degrees right. This directional output signal is capable to follow a moving sound source, but removes statistical variations in the firing pattern.

In addition to the directional output, a second output signal - named motor output - is generated by filtering the directional output through a first order Butterworth low pass filter with a cut off frequency of 1 Hz. It generates a slower reacting but much more smooth output of the model. The reason behind is that in case of natural sound sources variations of the directional output are still high and the slight jitter will increase significantly (see section 5.3). In order to obtain a smooth signal, capable to steer a motor under real world conditions, this filtered motor output is generated. The employed SIMULINK standard element “transfer function“ is shown in figure 4.37 with green color. It generates the appropriate filter coefficients and simply filters the directional output by technical means. The relationship between directional output and motor output is shown in figure 4.38.

It can be observed that the motor output follows the directional output with a delay of about 20 ms. Therefore, the reaction time (minimal time to sense the correct position of a sound source) increases from 10 to about 30 ms. However, no motor will need to react faster in turning i.e. a camera toward a sound source and since the motor output moves within that 30 ms towards the sound source position, it will resemble the moving speed of a natural head fairly well.

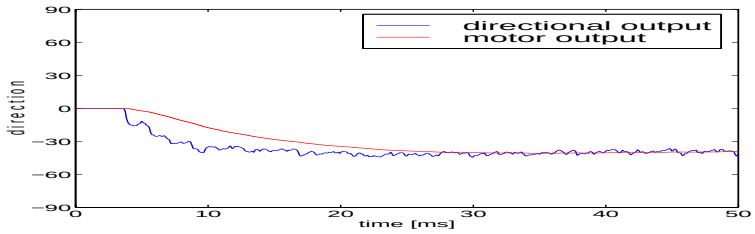

Figure 4.38: Relationship between directional and motor output in reaction to a synthetic sinusoidal signal arriving with IID= + 10 dB from the right

To demonstrate the sensitivity of the directional output, it's reaction to synthetic sinusoidal signals with IID's between -25 dB (sound from the left) and +25 dB (sound from the right) is shown in figure 4.39.

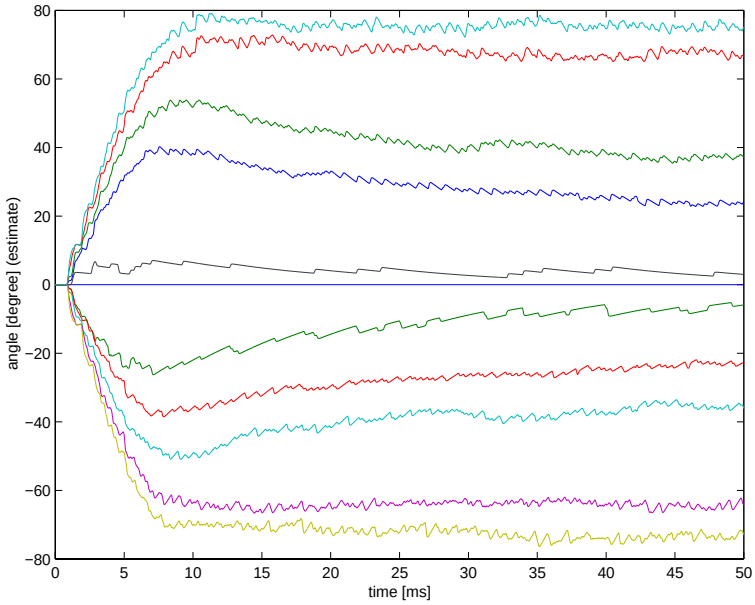

Figure 4.39: Output of the directional output element for sinusoidal signals with IID's reaching from $+25dB$ (signal from the left) to $-25dB$ (signal from the right) in steps of $5dB$

Here, it becomes visible that the directional unit proves to be quite sensitive. Already IID's of $5dB$ cause the unit to deflect into one direction, while at $0dB$ IID (and 0 ITD) there is no direction sensed (blue line at zero). Please note that the angle shown at the y-axis of this plot is only an estimate of the true direction, since the relationship between IID and angle depends on the real sound shadowing properties of the head (see section 2.1 Interaural Disparities). Since the signals here are of synthetic nature, with no real shadowing involved, the angle here and on the following figures is only for orientation, rather than a exact value. (The calibration factor was set to 2 in order to generate the displayed signals). The slight over-reaction during the first $30ms$ at small IID's, results from the abrupt onset of the synthetic signals, causing initial high frequency components but will not be present in case of natural signals.

More interesting than the directional sensitivity at this point, is the ability of the directional output to suppress echoes arriving within the echo threshold and stick to the direction of the first sound. Therefore, figure 4.40 shows the directional output in response to a sound signal containing an echo.

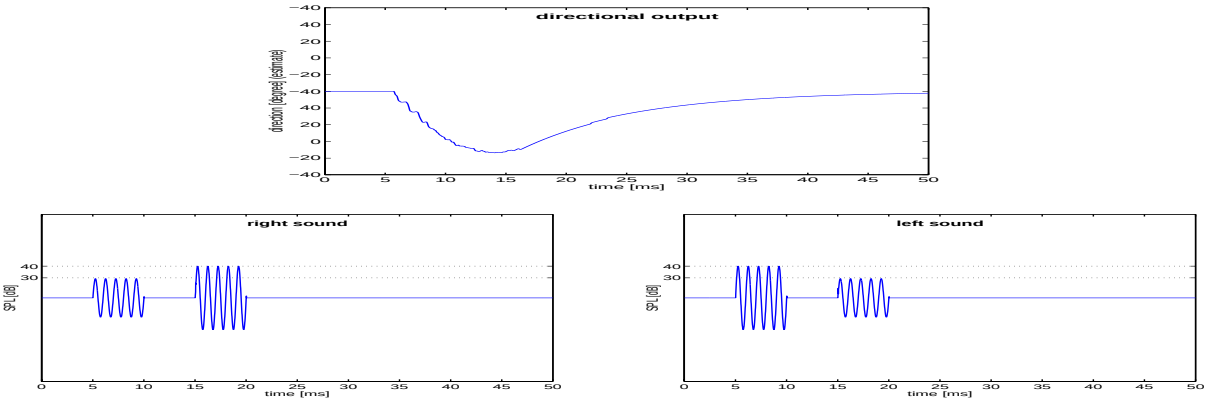

Figure 4.40: **bottom panels** - *5ms sound signal with artificial echo* **top panel** - *output of the directional unit suppressing the echo*

As can be seen, although the signal contains an echo arriving $5ms$ after the signal has ended from the opposite hemisphere, the output of the directional unit does not change direction since the directional information of the echo is suppressed by the persistent inhibition within the DNLL. Figure 4.41 on the opposite, displays the response under absolute identical conditions, except that the echo pulse arrives now 20 ms after the signal has ended and cannot be counted as an echo anymore. Consequently, the directional unit will sense it as a different sound arriving from a different direction.

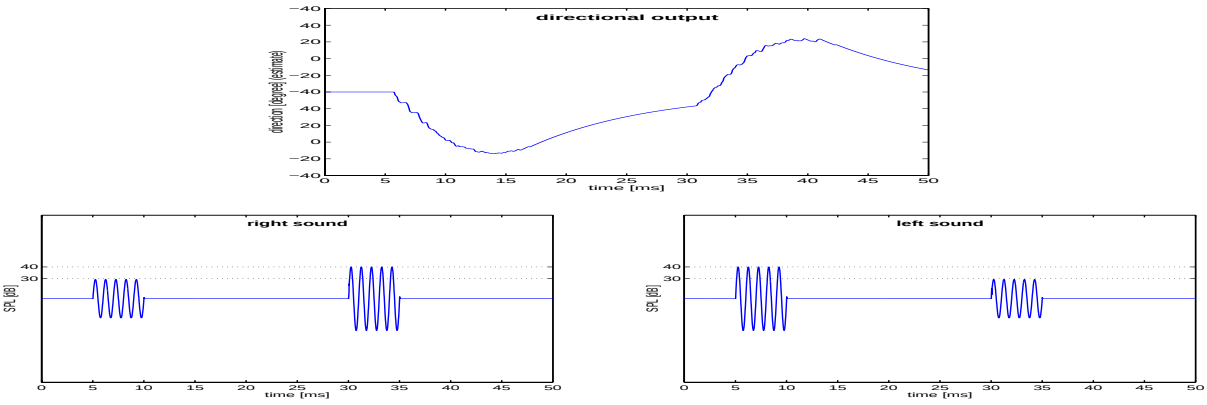

Figure 4.41: **bottom panels** - *5ms sound signal with two binaural pulses arriving with 20ms distance* **top panel** - *output of the directional unit sensing the second pulse as separate signal*

As shown by these two simple examples, the directional output seems capable to sense the direction of synthetic sound sources while suppressing echoes arriving from different directions within the time frame of the echo threshold.

The examples shown so far, employed rather simple and synthetic signals, designed to clarify the functional principle of this neural architecture for echo suppression during sound source localization. The next chapter will now describe systematic tests and physiological experiments carried out to evaluate the general performance of this model and to compare its output with living neurons in the gerbil.

Chapter 5

Experimental Results

The aim of this chapter is to prove the physiological relevance as well as the practical usefulness of the introduced architecture by several types of experiments. In the first section "Physiological Experiments", physiological recordings in the Gerbils DNLL in response to synthetic stimuli will be compared to the model's output in order to prove it's physiological compliance. The second section "Psychoacoustic Experiments", describes a number of simulations replicating psychoacoustic effects. Finally, the last section "Real World Experiments", displays the experimental results of real world experiments using sound stimuli, recorded in normal, and especially reverberating environments in order to fortify the usability of this architecture for technical systems under real world conditions.

5.1 Physiological Experiments

There were two major reasons to carry out a set of physiological experiments during this study:

1. to validate the hypothesis that echo suppression effects potentially caused by persistent inhibition , so far only recorded from bats, are also present in mammals like the Gerbil

2. to validate the model architecture by comparison of living cell spike responses to model cell spike responses under identical stimuli conditions.

The physiological experiments used a single cell recording technique in the Dorsal Nucleus of the Lateral Lemniscus (DNLL) of the Mongolian gerbil (*meriones unguiculatus*). The experimental method as well as the recording technique will be shortly outlined in the first two subsections. Afterwards, subsection 5.1.3 will display the recorded spike pattern from living cells and compare them to the obtained spike pattern of simulated cells in the model architecture under identical stimulus conditions.

Based on two experiments, originally developed by Pollack et al. [Pol97] for recordings in the DNLL of the mustache bat, it will be shown that persistent inhibition of DNLL-EI cells is most likely also present in the Gerbil and might be a common feature of mammals. Furthermore, the similarity between model-cell firing and living-cell firing in response to identical stimuli validates the models relevance, to duplicate and explain persistent inhibitory effects in the DNLL and support the hypothesis of their significant role for echo suppression during sound source localization.

5.1.1 Experimental Method

Single Cell auditory responses from 27 cells have been recorded in 6 adult gerbils. Out of the 27, 12 cells were found to exhibit stable EI response properties and in 8 of them significant persistent inhibition effects have been recorded.

The animals were anesthetized by an initial injection of 10 ml per 10 g body weight of general anesthetic solution (20% ketamine, 2,5% rompun and 77,5% NaCl). During surgery and recording sessions, the drugs were applied continuously by intramuscular injection (0.2 - 0.3 ml/h) until the animal was finally sacrificed by an doses of T61.

As shown in figure 5.1 middle panel, skin and tissue covering the upper part of the scull was cut and pulled aside laterally, in order to mount a metal rod onto the frontal part of the scull using UV sensitive dental-restorative material (Charisma, Heraeus Kulzer, Dormagen, Germany). The rod was used to fix the gerbils head in a stereotaxic device during recordings.

For electrode penetrations the dermis layers over the os parietale were carefully removed and a small elliptic opening was cut into the left cranium (left panel of figure 5.1). After removal of the dura mater within this opening, access to the brain was given to insert the 1 $M\Omega$ metal electrode. During the experiment, 0.8% NaCl Solution was frequently applied to the opening to prevent the brain from dehydration.

Figure 5.1: Animal recording setup **left panel** *Opening the left cranium of the gerbils skull* **middle panel** *Inserting a metal electrode after stereotaxic fixation by a metal rod* **right panel** *Sound attenuated recording chamber*

As shown in the right panel of figure 5.1, the animal was placed on a heating cushion ($39°C$) in a sound attenuated recording chamber. The animals position in the recording chamber has been standardized by stereotaxical landmarks on the surface of the skull (intersections of the bregmoid and lambdoid sutures with the sagittal suture in horizontal alignment). In order to properly adjust the electrode position relative to the brainstem, the animal was rotated rostrally by 60° (see gerbil brain atlas [LLV74]). Micro-manipulators were used to position the electrodes relative to landmarks on the brain surface used as reference point for all electrode penetrations.

Single cell responses were recorded using $1M\Omega$ Thungsten insulated metal electrodes (TM33A10KT) covered by a $10\mu m$ glass tube except at the non-insulated top. The electrodes have been inserted between 1900 and 2500 μmleft and 500 to 900 μm caudal of the intersection between the sagittal and the bregmiod suture.

Acoustic stimuli were delivered using a Tucker Davis Technology System III, comprising two 16 bit D/A converters (DA3-2; sampling rate 250 Hz), two anti-aliasing filters (FT-6; cutoff 100kHz) and two separate digital channel attenuators (PA4). They were delivered via a headphone buffer HB6 (Tucker Davis Technologies System II) and two Beyer dynamic speakers (model DT 900). The sounds of the left and right channel were dichotically presented to the animal via two 5 mm diameter probe tubes connecting the speakers directly to both pinae of the gerbil (see figure 5.1 middle panel). The earphones including the tubes have been calibrated using a 1/4 in microphone (Reinstorp VtS), a measuring amplifier(MV 302, Microtech, Gefell, Germany) and a waveform analyzer (Stanford Research Systems, SR 770 FFT network analyzer).

5.1.2 Recording Procedure

To search for acoustically responsive EI cells in the left DNLL, sinusoidal search stimuli of 100 ms duration and 3 ms rise/ decay time were presented to the right (excitatory) ear containing frequencies between 200 to 10000 Hz. The recording of action potentials and the stimulus generation was controlled by the TDT System III (see figure 5.2 left panel) and a customized commercial software (Brainware - for sensory Electrophysiology Version 7.0.2 with support for TDT system - Jan Schnupp, Department of Physiology University of Oxford UK). The electrodes have been moved by a remote control, using a motorized micromanipulator (Inchworm controller 8200 by Burleigh Instruments) from about 1500 μm depth (with regard to the brains surface) down to 6500 μm this way crossing the IC first and covering most of the DNLL structures found usually between 4000 and 6000 μm depth.

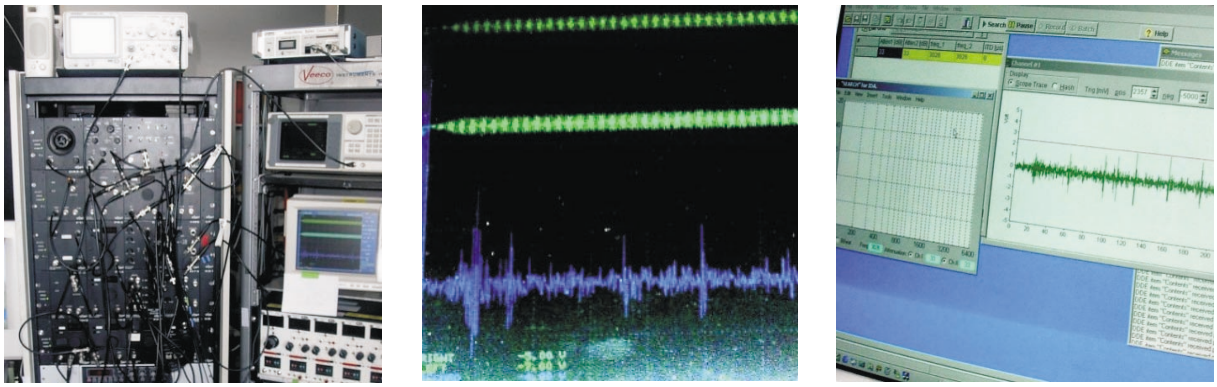

Figure 5.2: Animal recording system **left panel** *Tucker Davis Technology III system* **middle panel** *Stimuli and recorded spikes as displayed on the oscilloscope* **right panel** *Spike counting, evaluation and storage in the Brainware application*

The recorded signals where fed through a 50/60 Hz noise eliminator (Humbug, Quest Scientific) and a 0.7 to 3 kHz band-pass filter (spike conditioner PC1, TDT System III). During recording, the stimulus signals and recorded spikes have been displayed on a oscilloscope (see figure 5.2 middle panel) and digitalized by the two 32 bit A/D converters (AD2-3) of the Tucker Davis III System. From here, they were sent to a standard PC running the Brainware application (right panel of figure 5.2). Only action potentials from single neurons with a Signal to Noise Ratio (SNR) >5 were counted. Number, timing (relative to stimulus onset), shape and clustering of action potentials have been displayed and stored using the Brainware application.

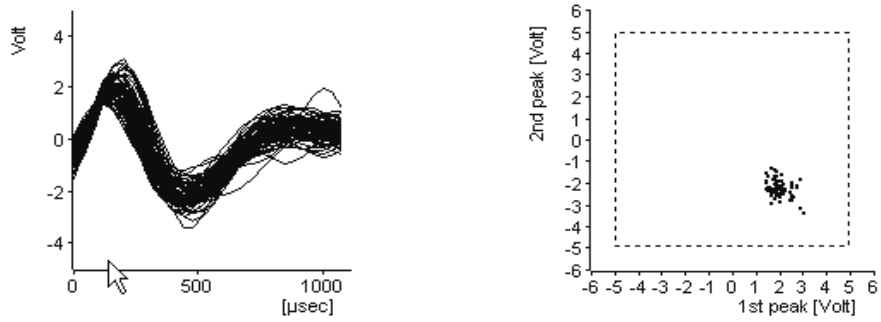

Figure 5.3: Shapes and clustering of recorded action potentials
left panel *Shape of action potentials - used to check response consistency*
right panel *Clustering of AP shapes - neuron where only be used if the 1'st and 2'nd peaks resulted in a sharp cluster as shown above*

The identification of auditory responsive single cells was based on the consistent shape of action potentials (APs) and the clustering of AP's (shown in figure 5.3), on the SNR (visually observed and measured at an oscilloscope) and on a subjective, auditory judgment of the recorded signal acoustically emitted by the speaker of the TDT System III. Within the 6 animals investigated, 27 contralateral excitable DNLL cells have been identified and recorded from.

To determine the Best Matching Frequency (BMF) and Response Threshold (RT) for each identified DNLL cell, tuning curves have been measured by varying stimulus frequencies between -2000 Hz and + 2000 Hz against expected BMF and attenuation between -20 to -80 dB. A typical example of the recorded cell response is shown in figure 5.4.

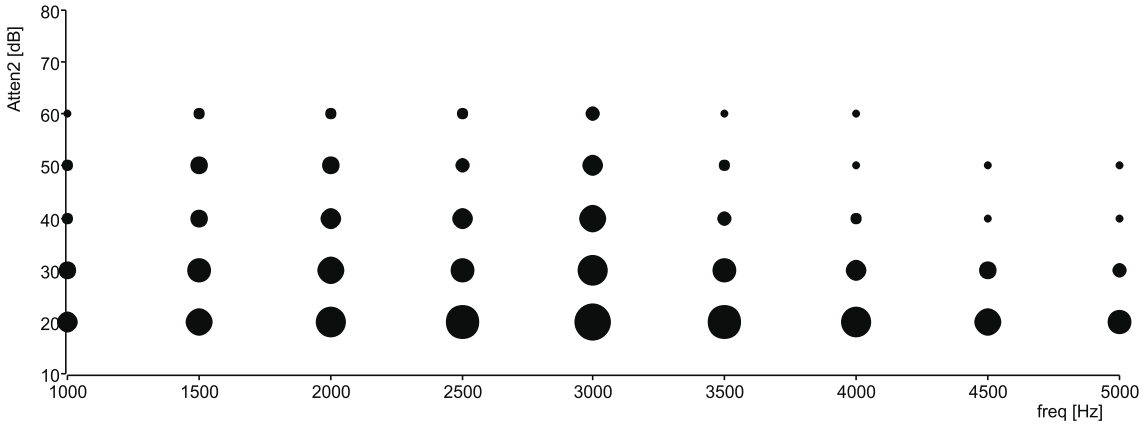

Figure 5.4: Recorded spike rate (circle width, max = 72 spikes/s) from a single DNLL cell to exclusively contralateral stimulation at varying frequencies and attenuation - revealing a BMF of 3 kHz and a RT of -60 dB

Here, the circle width represents the number of spikes during 25 presentations of each stimulus, revealing a Best Matching Frequency (BMF) of 3 kHz and a Response Threshold (RT) of -60 dB for that cell. Additionally, the asymmetric structure of the recorded tuning curves becomes visible when comparing responses to frequencies below 3 kHz to those above.

After determination of BMF and RT, the DNLL cells have been tested whether they belong to the EI type. Hence, if an ipsilateral presented signal is capable to inhibit the response to a contralateral presented stimulus, then this cell is of EI type. To determine the response type, the contralateral stimulus has been kept 20 dB above response threshold (i.e. right attenuation 2 was kept at 40 dB due to a response threshold of 20 dB in figure 5.5). Frequencies where chosen close to the BMF (i.e around 1300 Hz in figure 5.5). The ipsilateral attenuation 1 was varied between -10 and +50 dB with regard to the response threshold (i.e. between 10 and 70 dB for a response threshold at 20 dB in figure 5.5). This way, the resulting variations of interaural intensity differences (IID) range between -30 and +30 dB. One example of the recorded single cell responses is displayed in figure 5.5.

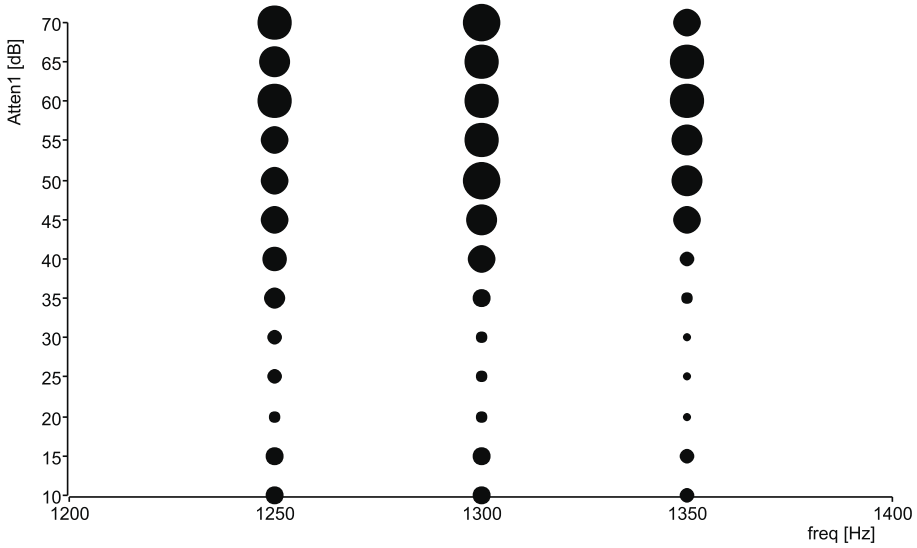

Figure 5.5: Recorded binaural response (circle width represents spike rate) of a single DNLL cell at three frequencies around the BMF of 1300 Hz. Contralateral attenuation was fixed to 40 dB and IID's varied from +30 to -30 dB by changing the ipsilateral attenuation (Atten 1) between 10 and 70 dB)

As can be observed, contralateral stimulation leads to strong responses as long as the ipsilateral signal is more attenuated than the contralateral one (Atten1 between 70 and 45 dB). If both attenuations are equal at 40 dB (IID=0), the number of emitted spikes decreases. In case of negative IID's (ipsilateral stimulus is less attenuated than the contralateral one - Atten1 between 35 and 10 dB), ipsilateral inhibition exceeds contralateral excitation and the spikes nearly vanish.

To distinguish the cells with EI behavior, a general rule was established: If the spike rate during binaural stimulation with IID's of -30 dB decreased at least by 50%, compared to purely contralateral stimulation, the cell was counted as EI type and selected for further investigation. Out of the 27 identified cells, 12 have been of the EI type - a 44% share somewhat below the observations of Kelly [KBK98], who found 74% of DNLL cells with EI responses. The reason might be either the small sample size, not claiming general significance, or a generally less developed EI type in the Mongolian Gerbil compared to the rat's brain Kelly recorded from. Nevertheless, the observed results suggest that a significant number of the Gerbils DNLL cells exhibits excitatory/inhibitory binaural response properties.

5.1.3 Results of the Physiological Experiments

The goal of the physiological experiments performed during this study was to find evidence for the existence of persistent inhibition within the Gerbils DNLL and to compare natural cell responses to those of the model. In order to achieve this goal, an experiment originally developed by Yang and Pollak [YP94a] to determine persistent inhibition in the mustache bat, was reproduced and applied to the gerbil as well as the neural model architecture.

The employed synthetic stimulus was specially designed, to show the inhibitory effect immediately after the end of an inhibitory stimulus - the persistent inhibition. It contained three pulses of 10 ms duration, consisting of sinusoidal signals at the BMF of the neuron. The left and right channel have been presented dichotically (separate but at the same time - via headphones) to the gerbils ears during the experiment and to the two ear models during the model simulation. Out of the three pulses, only the first one was binaural (stimulating both ears), while the second and third pulse were purely monaural signals, stimulating the contralateral (excitatory) ear only. It is important to mention that under these conditions, and the accepted assumption of contralateral excitation and ipsilateral inhibition, only the first pulse could cause inhibitory effects.

The expected result is that the neurons response to the second and third pulse is identical if there is no persistent inhibition. If persistent inhibition is present, the response to the second pulse should be partly or fully suppressed as a result of inhibitory effects immediately after the end of the first pulse. Consequently, the response to the second pulse should be strongly suppressed, if it follows immediately after the first pulse. It should be less suppressed as the gap between the pulses increases and the second pulse moves outside the period of persistent inhibition.

From a perceptual viewpoint, the first pulse can been viewed as the acoustic signal and the second and third one as echoes, arriving within or outside the echo threshold.

The experiment was performed with all of the 12 EI cells, while 4 cells did not show significant suppression to the second pulse, the other 8 gave quite clear and similar results. The obtained response of such a single cell in the Gerbils left DNLL - with BMF at 3 kHz and Response Threshold at 40 dB - is shown in figure 5.6. During the experiment, the IID of the first pulse has been varied between -20dB (inhibitory signal 20 dB louder than excitatory signal) and +20 dB (excitatory signal 20 dB louder than inhibitory signal). The experiment has been replicated 100 times in random order under identical conditions and the spike response times have been slotted into 1 ms slots for displaying reasons only.

Each of the blue columns in figure 5.6 displays the averaged number of spikes recorded during 100 iterations and therefore the spiking probability during this specific time slot of 1 ms. Hence, the sum over all columns within one stimulus pulse could be calculated to determine the response probability of that specific cell to a specific pulse.

The green areas represent the excitatory stimulus pulses (at the right ear) in duration (10 ms for each pulse) and intensity (40 dB for each pulse). The red areas on the other hand represent the inhibitory stimulus (at the left ear) in the same manner applied with varying intensities between panel A, B and C.

Finally, the yellow areas mark the expected effect of persistent inhibition assuming decay time constants of 20 ms. In should be noted that while the red, green and blue areas represent real presented or obtained values, the yellow areas have been added only as an explanation and do not represent recorded signals.

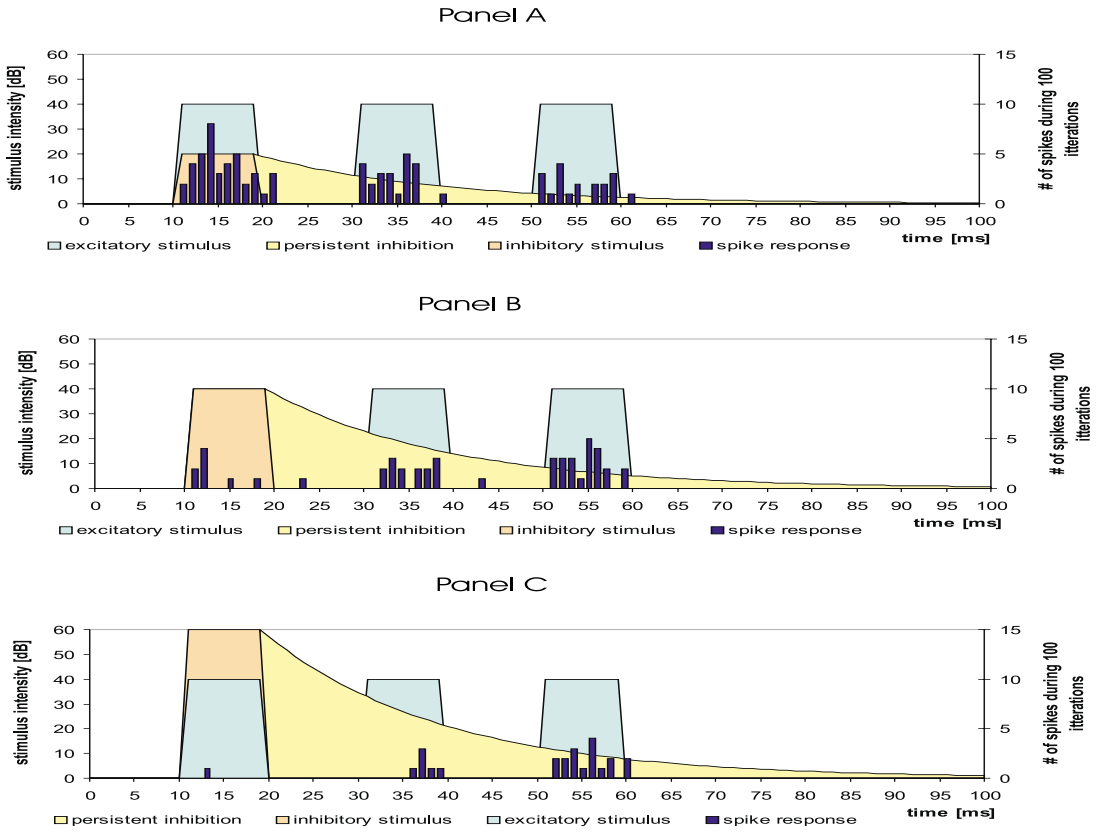

Figure 5.6: Recorded spikes from a single DNLL cell to a binaural leading pulse with IID's varying from +20 (panel A) to -20 dB (panel C) followed by two contralateral pulses at 40 dB after a gap of 10 ms

Shown in *Panel A*, the response to the three pulses is nearly identical, if the first binaural pulse is excitatory in nature - the excitatory signal at the right ear is significantly louder than the inhibitory signal at the left ear. The slightly stronger response to the first pulse has been perceived in all experiments with pulse gaps below 30 ms and seems to result from intrinsic recovery parameters of the living neuron.

Panel B displays the response of the same cell to the same signal, except that the first pulse exhibits a IID of 0 dB - the excitatory signal at the right ear and the inhibitory signal at the left ear have an identical SPL of 40dB. It can be observed that the cells response to the first pulse is diminished but not fully suppressed. The response to the second pulse is only slightly diminished possibly due to the assumed persistent inhibition indicated by the yellow area.

The reaction to a leading inhibitory pulse is displayed in *Panel C*. Due to a 20 dB louder signal at the left ear compared to the right ear, the neuron's response to the first pulse is almost completely suppressed. But more interestingly, although there is no inhibitory stimulus during the second pulse, the response is still strongly suppressed and significantly delayed (no onset spikes to the second pulse). The only explanation for this phenomena seems to be a persistent inhibition, resulting from the first pulse lasting more than 10 ms after its end and influencing the response to the second pulse. Finally, the response to third pulse is nearly identical to panel A indicating that the cell is responding normally if the persistent inhibitory effect has vanished.

Overall this experiment shows the existence of inhibitory effects in the DNLL cells of the Gerbil, 10 ms after the end of the inhibitory stimulus and therefore indicates the existence of persistent inhibition in the Gerbil.

To validate the model architecture, the same stimulus has been presented to the model after converting it from a .wav to a .mat file. The stimulus was presented 100 times to the model and the spike times of the corresponding BMF channel (channel 11) have been counted and slotted in the same way as the biological recordings. The obtained results are displayed in figure 5.7.

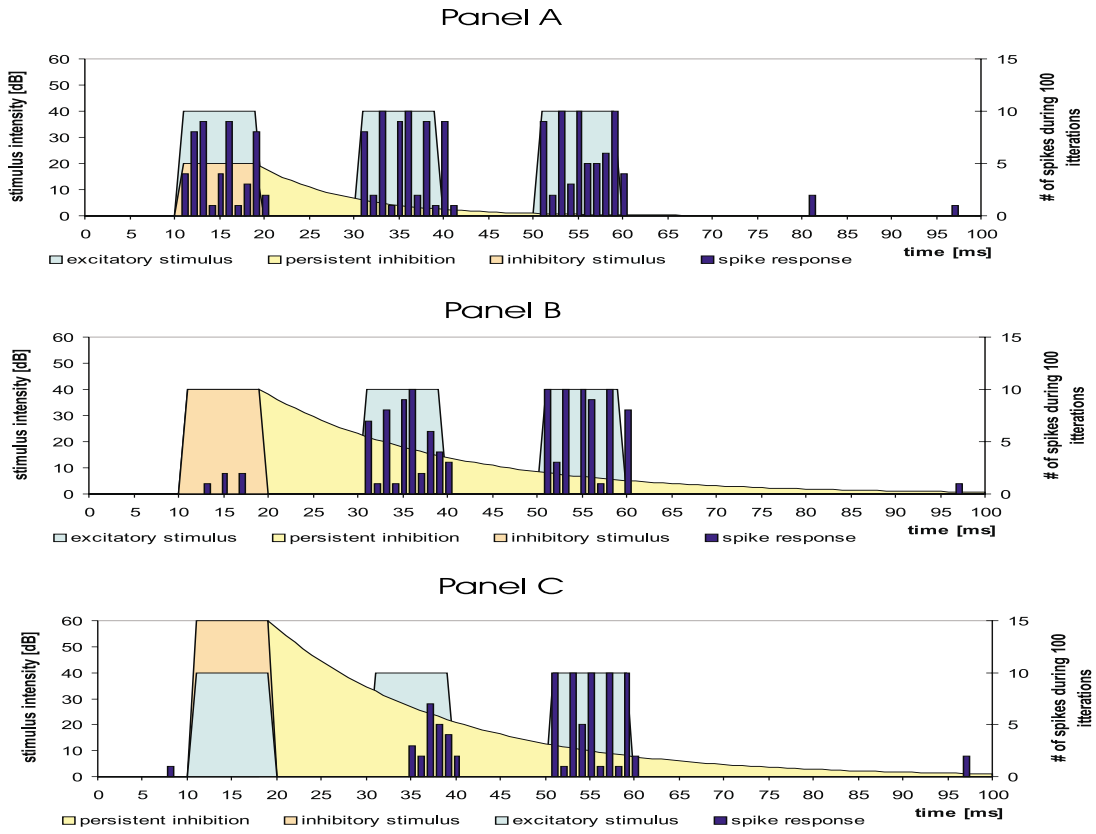

Figure 5.7: Simulated spike response of a single DNLL cell model (channel 11 - within the left DNLL of the model architecture) to a binaural leading pulse with IID's varying from +20 (panel A)to -20 dB (panel C) followed by two contralateral pulses at 40 dB after gaps of 10 ms

Despite some scaling differences the principally identical response of the model and the cell are clearly to observe. Analog to the cell response, the model response to the three pulses is nearly identically after an excitatory leading pulse in *Panel A*. In case of identically excitatory and inhibitory sound pressure levels during the first pulse (Panel B), there are only a few spikes in reaction to the first pulse and the response to the second one is slightly diminished. Most importantly, after an inhibitory leading pulse in *Panel C*, the response to the second pulse is strongly diminished by persistent inhibition, while the third one is not much affected.

The overall higher spike rate of the model compared to the living cell is most likely due to the more deterministic and responsive firing behavior of the cell model, resulting from it's simulation on a digital CPU. Nevertheless, statistical firing behavior, changing the response pattern between the different runs of the same model with the same signal has been achieved by increasing the

noise term of the DNLL models from 0.00001 to 0.002. This causes visible variations of the firing times during the three pulses but cannot completely override the concentration of most spikes in specific slots separated by the after hyperpolarization time constants of 2 ms. As can be observed in *Panel C*, the deterministic response character vanishes as the inhibitory influence increases and will be no more perceivable in case of complex dynamic inputs.

Despite these small limitations this experiment proves that the designed model architecture can duplicate the spiking behavior of living cells in the Gerbils DNLL and can be seen as a valid model of this specific subset of the auditory brainstem.

In order to determine the **duration of persistent inhibition** in the Gerbils DNLL and its correct modeling, another set of experiments developed by Pollak et al.[YP94a] has been carried out at both, the Gerbil and the model. Here, the IID of the first pulse has been kept stable at -20 dB but the distance (gap) between the pulses has been varied between 5 and 30 ms. This way, the duration of persistent inhibition can be determined by the amount of suppression to the second pulse as it moves away from the first pulse.

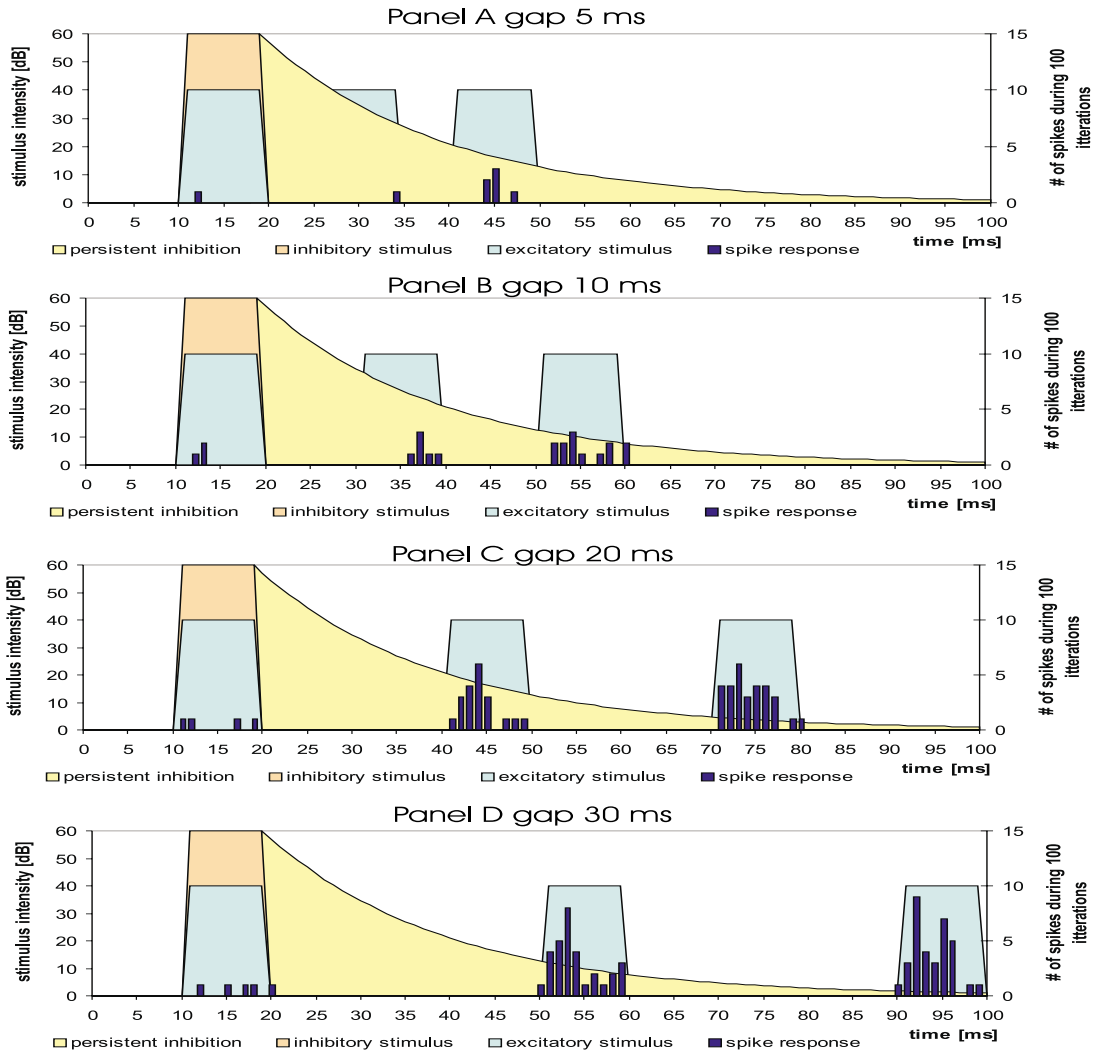

Figure 5.8: Recorded spike response of a single DNLL cell to a binaural leading pulse with IID's of -20 dB followed by two monaural contralateral pulses of 40 dB with varying gaps between 5 and 30 ms

As can be observed in figure 5.8, the influence of the yellow shaded persistent inhibition may last much longer than 30 ms after the end of the first pulse (as indicated by the yellow shading). However, its suppressing influence to the second pulse ends around 30 ms after the end of the signal and is comparable to those observed by Pollak et al. As shown later, the suppressive effect of persistent inhibition depends as well on the IID (strenght of inhibition) of the first pulse as on the intensity of the following pulses but might be limited to a maximum around 30 - 40 ms according to the model.

If the pulses follow each other with short distances of 5 ms, as shown in *Panel A* of figure 5.8, persistent inhibition will not only completely suppress the neurons response to the second pulse, but it will also significantly diminish the third pulse response. This suggests that the inhibitory influence is still effective during the third pulse 20-30 ms after the end of the first pulse.

As the gap increases to 10 ms (*Panel B*), the second pulse causes only some response, but the cell responds quite normally to the third pulse. Note that the overall low response rates in both cases may result from the overall short duration of the pulse train and the intrinsic properties of a single cell, thus preventing it from such fast reactions. Nevertheless, an assembly of neurons as usually acting to perform the same task in the brain, might be quite more reactive and exhibit a constant behavior.

As the gap further increases towards 20 ms (*Panel C*), the response to the second pulse is still somewhat suppressed compared to the third pulse. However, the cell responds normally to the third pulse - by now outside the reach of the assumed persistent inhibition.

Finally, a gap of 30 ms between the pulses, as shown in *Panel D*, results in nearly equal responses to the second and the third pulse and supports the observation that after 30 ms, persistent inhibition has decreased too much, to further suppress the cells response to an excitatory stimulus of 40 dB. However, a less intense stimulus might still be suppressed. The intensity-dependency of echo suppression has been investigated more systematically during the psycho-acoustical experiments in the next section.

Overall, the experimental results obtained by Pollak et al. during single DNLL cell recordings in the mustache bat have been supported by recordings in the Gerbil, suggesting a persistent inhibition of DNLL cells with suppressive effects lasting more than 20 ms after the ending of an inhibitory stimulus. This might therefore be a general feature in mammals and fits well to the psycho-acoustically obtained echo thresholds.

Beside the approval of this hypothesis, the aim of this experiment was to show that the designed model architecture also duplicates the duration effect and therefore might be a valid model to suggest further behavioral aspects of DNLL cells under different dynamic stimulus conditions.

In order to test the model, again the stimuli used with the animal have been applied to the models input and the simulations have been repeated under the same conditions as in the first experiment. The recorded spikes in response to the different stimuli are displayed in figure 5.9.

It becomes visible in figure 5.9 that the overall responsiveness of the model cell is somewhat higher than the responsiveness of the living cell. Nevertheless, the model fits well to the overall activity of the DNLL cell assembly, involved in the echo suppression task and clearly exhibits similar spike timing as the single living cell.

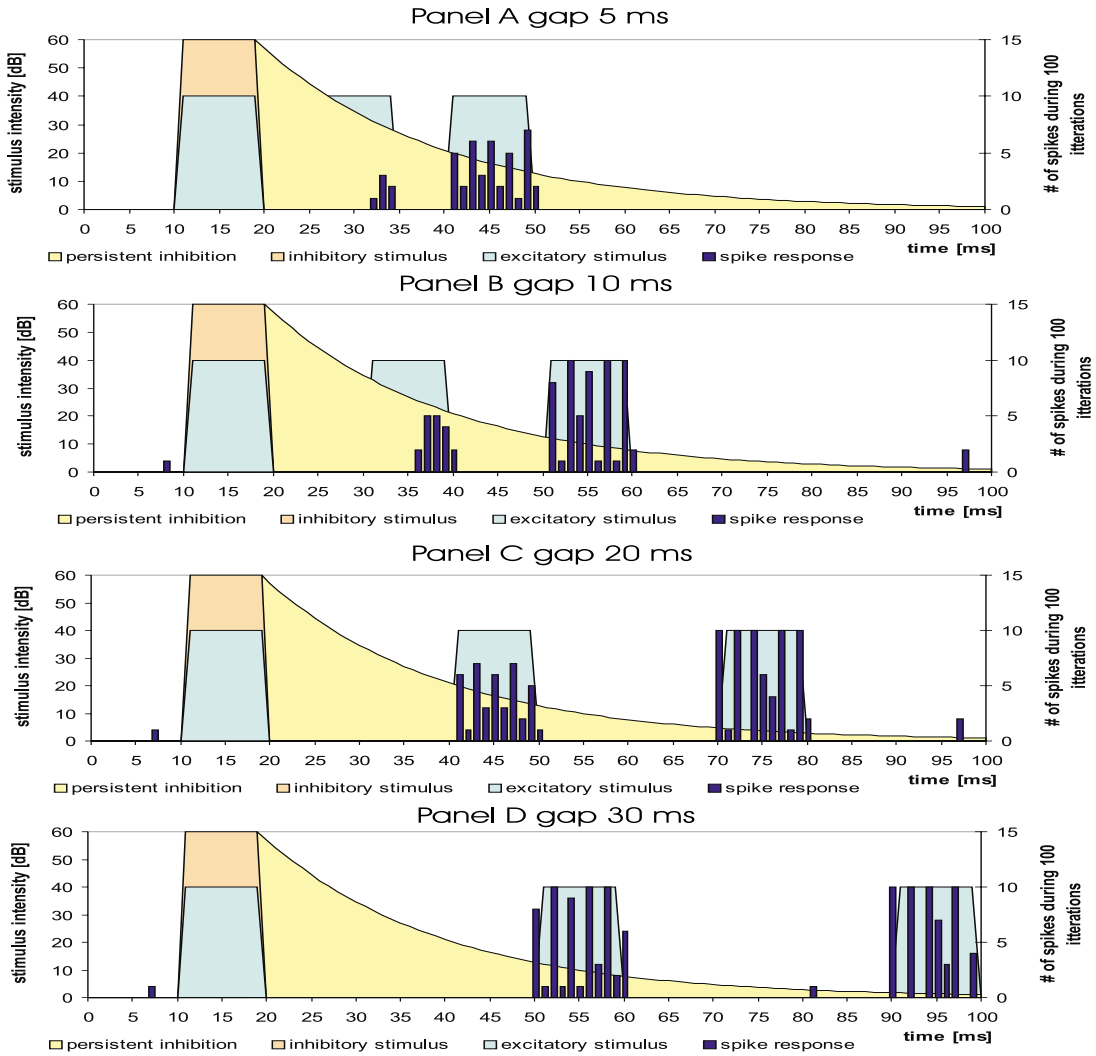

Figure 5.9: Simulated spike response of a single DNLL cell model to a binaural leading pulse with IID's of -20 dB followed by two contralateral pulses at 40 dB after gaps between 5 and 30 ms

By evaluating the observed spike pattern of *Panel A*, the strong suppressive influence of the first inhibitory pulse to the second pulse 5-15 ms, after the end of the first pulse, is clearly visible. Furthermore, the inhibitory influence on the third pulse 20-30 ms after the end of the signal is visible by a 30% decrease of spike rate.

As the gap between the pulses increases further, a similar behavior as in the recording experiment, is achieved. While a gap of 10 ms still results in strong suppression of the second pulse, this gradually disappears as the gap raises above 20 ms. In the bottom panel D, responses to the second and third pulse are no more influenced by the leading signal. Please note that this is shown for identical signals in terms of intensities and frequencies. The picture can be different, if for example the "echo" pulses have lower or higher intensities than the "signal" pulse or vice versa.

5.1.4 Summary

First of all, the results of the two physiological experiments have validated the hypothesis of persistent inhibition as a feature of specific cells in the Gerbil's DNLL and support the hypothesis of persistent inhibition as a general feature of mammals. Nevertheless, the statistical significance of 8 cells is fairly small and might need to be extended by further recordings and different experimental setups. Since the aim of this thesis was to find a valid model, capable to explain, duplicate and predict the specific response of DNLL cells to acoustical echoes, experimental efforts have not been extended at this point, but will be continued in the future, based on specific stimuli conditions with model-predicted response pattern.

Second, the biological relevance of the model architecture has been validated by a direct comparison of spike pattern between the biology and the model. The model might now be employed by biological experimenters to predict and explain living cell responses under specific conditions (i.e. a moving sound source), based on the models output. This way, compliance to natural effects, as well as the level of detail in the model, could be improved and further increase its value for the exploration of neural cell functions in the auditory brain stem. The value to the neuromorphic engineer arrives from the opposite site. As the employed dynamic structures obviously proved their capability to duplicate a dynamic biological cell behavior, they can now be used as building blocks for technical models, aiming to replicate other effects of biological information processing. This way they could help to achieve technical systems, capable to cope with real world tasks, currently restricted to living neural systems.

5.2 Psycho-Acoustic Simulations

While the physiological experiments of the first section have been concerned with the recording and duplication of single cell spike responses at the microscopic level, the psycho-acoustic simulations of this section focus on the duplication of more macroscopic perceptual effects, revealed by the psychoacoustic experiments introduced in section 2.3.2. Based on a number of controlled acoustic stimulations, psycho-acoustic experiments explore specific perceptual effects, like the precedence effect and the echo-threshold, resulting from the complex interaction of auditory information within the entire auditory brainstem and cortex. In this thesis the aim was to test, if the developed model of the auditory brain stem can duplicate some of those psycho-acoustic effects during simulation of typical experimental setups and stimuli.

Several hypothesis on the role of persistent inhibition and the interaction between the Lateral Superior Olive (LSO), the Dorsal Nucleus of the Lateral Lemniscus (DNLL) and Inferior Colliculus (IC) have been discussed in section 2.3.2. In summary, 6 psycho-acoustically perceived features of echo suppression at the level of LSO-DNLL-IC interaction have been identified in section 2.3.2. These are:

1. Echo suppression is present under dichotic as well as real world free field conditions.
2. Intensity dependency - lead signals with high intensities cause stronger echo suppression. Very intense echoes are less suppressed.
3. ISD dependency - the Inter Stimulus Delay (ISD) between signal and echo determines the perception of echo direction.
4. Duration dependency - Longer durations of the lead signal cause longer echo suppression with upper limits around 20 ms.
5. Effective echo suppression occurs also if the sound originates from the midline of the auditory field.
6. Echo information is preserved even during suppression of directional information.

The aim of this section is to prove that the designed model architecture is capable to duplicate these effects at the macroscopic level, based on the detailed biological principles and internal connectivity described in chapter 4. Therefore, only the model output (directional sensor), the presented stimuli and in some cases the summed firing pattern at the IC will be shown, representing the outside view onto the model. Nevertheless, for all experiments, every single cell model and its intrinsic parameters, like dendritic potential, soma potential and firing threshold, were simulated and could be displayed if needed to elucidate the macroscopic effects. The employed model was exactly the same as during the first (physiological) set of experiments.

Since it is impossible to duplicate all of the specific psycho-acoustical experiments of the last 60 years, the examples in this section will focus on principle effects rather than duplicate the conditions of specific experiments. The stimuli used will be as simple as possible and will vary only in one parameter in order to explain the different features step by step.

The following subsections will describe the stimuli and experimental setups, as well as the macroscopic results, of a number of different experiments, designed to prove the capability of the designed neural architecture to model the features 1 to 6 listed above.

5.2.1 Feature 1: Echo Suppression under dichotic and free field conditions

The stimuli of this section are still of synthetic nature, hence they are designed at the computer rather than being recorded from real world sound sources. In contrast to section 5.1, they are entirely based on natural signals, instead of pure sine waves. The base for all of them was a recorded human male voice, vocalizing the letter "aaa" over a period of 400 ms. This signal was recorded by a Sennheisser microphone K6 and pre-amplified by a Toshiba SY-C12 amplifier before being digitalized by a Cirrus Logic Crystal Audio Codec Sound Card and recorded by the Cool Edit Pro V1.0 Sound Software. The recorded waveform is shown in figure 5.10. Based on this recording (and subsets of it) a number of stimuli were designed and presented under dichotic as well as free field conditions.

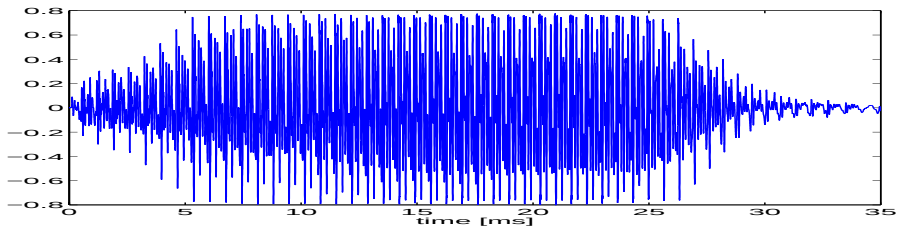

Figure 5.10: Waveform of a recorded human male voice intonating the vocal a

For free field experiments, two pulses of 20 ms length and equal intensity were selected from the recorded sound and send to two Altec-Lensing speakers. While the first pulse was send exclusively to the right speaker, the second pulse was only send to the left speaker, resulting in two pulses switching from the right speaker position to the left speaker position.

Figure 5.11: Experimental setup for the recording of free field signals using a Teddy with an ear base of 20 cm and two microphones (right panel) positioned in front of two speakers at angles of +-25 degrees and a distance of 1 m (left panel)

The experimental setup itself is shown in figure 5.11 and was composed of a Teddy with a head base (ear distance) of 20 cm - similar to the base of a human head. Since the head of the Teddy creates a fairly natural acoustic shadow (see section 2.1), the sound was perceived louder at the right ear if originating from the right speaker and louder at the left ear if originating from the left speaker. Both speakers have been positioned at an angle of 25 degree right and left from

the midplane at a distance of about 1 m from the Teddy. This way, the experimental setup resembles the conditions of a human listener perceiving a lead signal from the right followed by an echo from the left. The ears of the Teddy have be equipped with two Sennheiser K6 capacitor microphones (see figure 5.11 - right panel). Therefore, the perceived sound could be recorded using the Toshiba SY-C12 stereo pre-amplifier and the Cool Edit Pro Software in recording mode. While the designed stimulus (containing a 20 ms pulse followed by an echo from the opposite direction) was played by the speakers, the microphone signals where recorded simultaneously, digitally stored and later on presented to the left and right input of the model. The two recorded microphone signals are displayed in the upper right panels of figure 5.12 and clearly exhibit the expected IID's of about 10 dB generated under free field conditions.

In case of dichotic experiments, the left and right signal were directly designed based upon short pulses of 20 ms duration out of the vocal record in figure 5.10. Here, the IID of -10 dB for the first pulse and +10 dB for the second pulse were generated by digital scaling of the waveform. (see figure 5.12 upper left panels). The signals designed this way, were directly fed into the left and right ear model, similar to a presentation over ear phones.

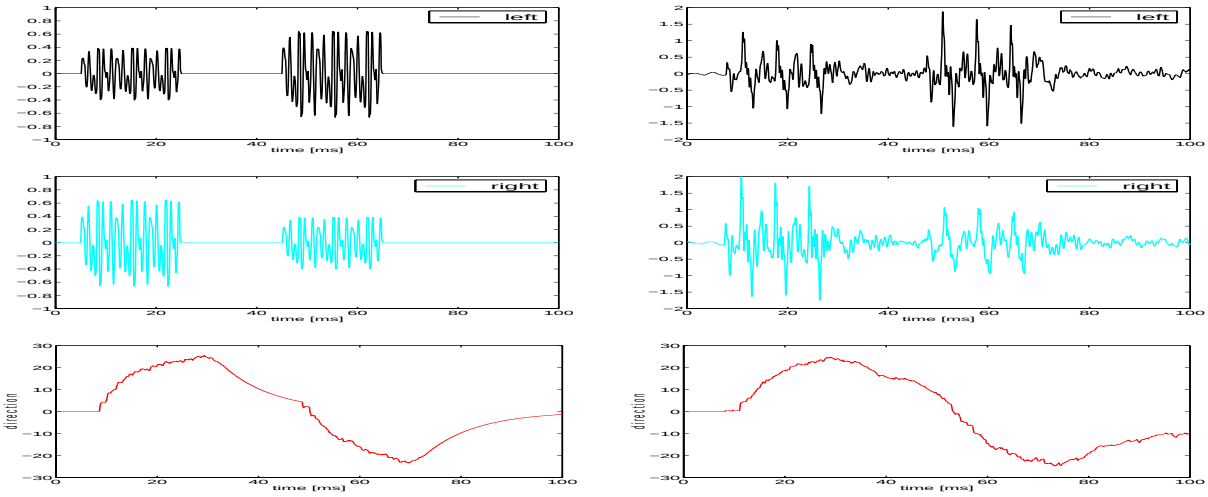

Figure 5.12: Directional sensation under dichotic and free field conditions
upper left panel - *left dichotic stimulus* **middle left panel** - *right dichotic stimulus*
upper right panel - *left free field stimulus* **middle right panel** - *right free field stimulus*
lower left panel - *directional output of the model in response to the dichotic stimuli above*
lower right panel - *directional output of the model in response to the free field stimuli above*

At the lower panels of figure 5.12 the directional model output is shown in response to the signals above. Despite the different waveforms of the synthetic signals (left panels) and the free field recorded signals (right pannels), the similar IID's result in both cases in a very similar sensation of direction. The first pulse is perceived to arrive from 25 degrees right (+25) and the second one is perceived to originate from 25 degrees left (-25). It becomes obvious here, that directional sensation is working under dichotic as well as free field conditions. In this case there is no echo suppression, since both pulses are perceived as separate signals, different in direction. This is correct, because the delay between the two pulses is 40 ms and therefore the second pulse occurs outside the echo threshold. The continuous output signal is generated by the integrative component of the directional output element, as described in section 4.7.

Under the conditions of figure 5.13 this has been changed. Here, the inter pulse distance has been shortened to 1 ms and therefore, the second pulse (echo) occurs 1-21 ms after the end of the first pulse (signal), hence, inside the echo threshold. While the upper and middle panels again exhibit the according left and right acoustic stimuli, the directional output at the lower panels now only senses *one direction - the one of the first pulse*. Here, the directional information of the second pulse (the echo) is suppressed, although there are cells within the IC model reacting to this pulse, as will be shown later in this section. Comparing the left and the right panel of figure 5.13 it becomes visible that echo suppression not only works under dichotic conditions but is also effective if the signal is recorded in the free field with natural IID's.

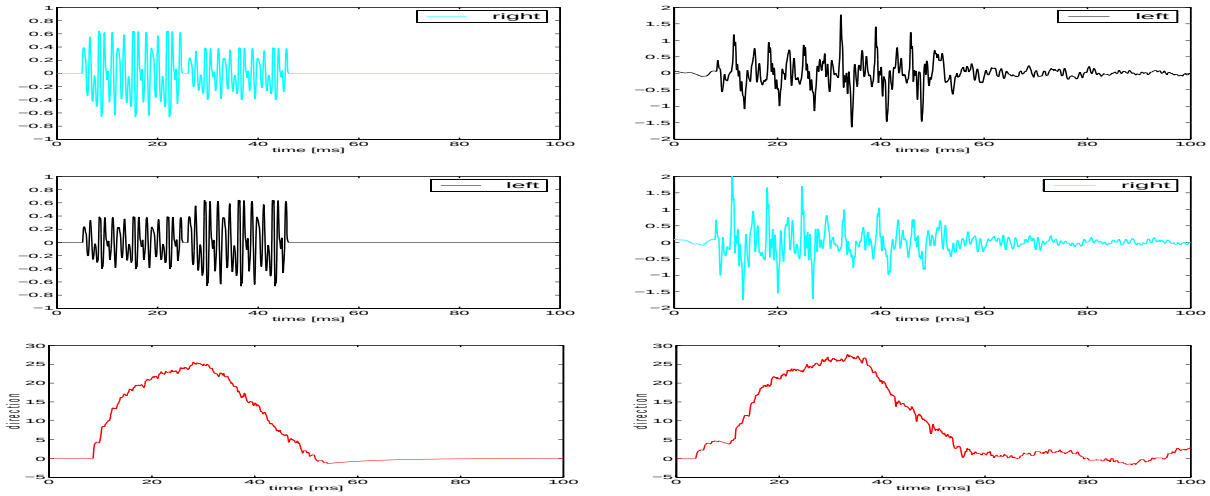

Figure 5.13: Echo suppression under dichotic and free field conditions
upper left panel - *left dichotic stimulus* **middle left panel** - *right dichotic stimulus*
upper right panel - *left free field stimulus* **middle right panel** - *right free field stimulus*
lower left panel - *directional output of the model in response to the dichotic stimuli above*
lower right panel - *directional output of the model in response to the free field stimuli above*

By these simple experiments, 3 important features of the model architecture have been shown:

1. *The model can correctly sense the direction of a signal if it arrives outside the period of echo suppression* (echo threshold) and would therefore be perceived by a human listener as a separate sound arriving from a different direction,

2. *the model can suppress the directional information of the same signal if it arrives within the period of echo suppression* and the human listener would then only perceive one signal arriving from the direction of the first sound, and

3. *the model can suppress echoes with "designed" directional information and dichotic presentation as well as echoes with "real" directional information, recorded under free field conditions.*

Additionally, the directional output of the model has been empirically tested by the experimenter and was found to be in compliance with the normal listeners perception.

5.2.2 Feature 2: Intensity dependence of echo suppression

5.2.2.1 Dependency on signal intensity

To test the compliance of the model with the psycho-acoustically proven *dependency of echo suppression on the intensity of the signal*, similar stimuli were employed. They also consisted of two 20 ms subsets of the recorded waveform and were digitally designed to exhibit an IID of +10 dB for the first pulse and -10 dB for the second pulse. The silent gap between the two pulses was kept constant at 10 ms, making the "echo" appearing 30 ms after the signal - just partly within the echo threshold. The two signal channels shown at the left panels of figure 5.14 have been presented to the model under dichotic conditions only.

The varied parameter in this first set of experiments is the absolute intensity of the signal (first pulse). This varied between 80 dB (40 db louder than the echo) and 20 dB (20 dB less intense than the echo). Since the IID of the signal was always kept at +10 dB, the directional perception should not be affected, while the amount of suppression imposed on the echo should be much larger in case of a loud signal (80 dB) than in case of a muted signal (20 dB). The presented stimuli and the model output for selected signal intensities are displayed in figure 5.14.

As can be seen at the top row of panels, a very intense signal will suppress the directional sensation of the echo, even if the echo lasts as long as 40 ms after the start of the signal. As the signal intensity decreases (second and third row of panels), the directional output more and more senses a second signal from the opposite direction. If the intensities of signal and echo are equal - as in the third row of panels - the echo is already sensed, partly as a separate signal due to the fact that it lasts longer than the echo threshold of 30 ms. If the signal is less intense than the echo - a rather unusual situation under natural conditions shown in the bottom panels - the echo becomes perceivable as originating from a different direction. Interestingly, even if the echo is as much as 20 dB louder than the signal (bottom panels) there is still some suppression to the echo response, preventing the output from reaching the expected -25 degree position.

From another point of view, the obtained results can be interpreted as a shortening of the echo threshold from more than 30 ms in case of the very intense signal to less than 10 ms in case of the very weak signal. This coincides with the findings of Damaske [Dam71], who used noise burst of 10 ms duration and obtained very similar echo threshold variations (see page 37).

It should be noted that the general delay of 5 ms, visible between input and output of the model, is caused by the transmission delays of the several stages of neurons and resembles fairly well the psycho-acoustically and physiologically reaction delay, between cells in the inner ear and the inferior coliculus (IC).

Overall, the experimental results shown in figure 5.14 display the principle capability of the model to resemble the psycho-acoustical phenomena of echo suppression dependency on the absolute intensity of the signal.

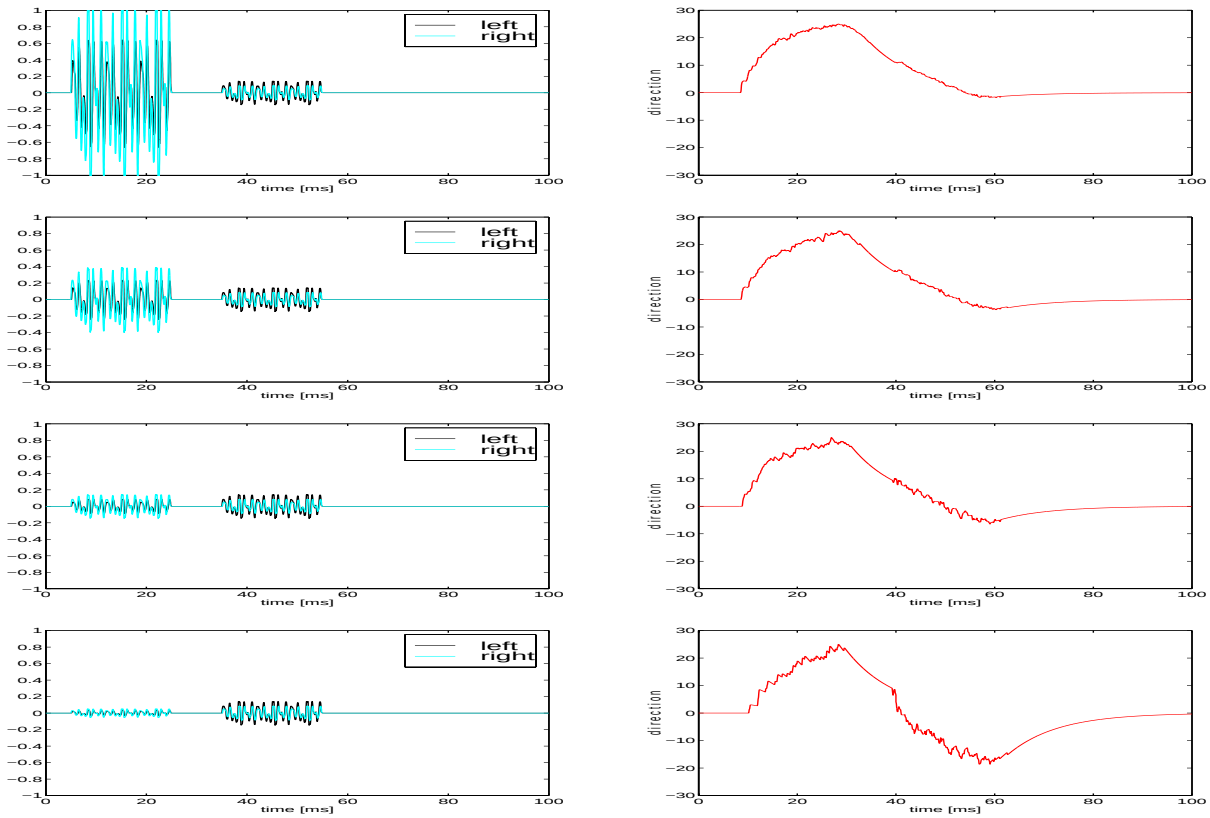

Figure 5.14: Echo suppression dependency on signal intensity - LEFT PANELS: presented stimuli: right signal (cyan color) always 10 dB louder than left signal (black color) - echoes arrive 10 ms after the signals end with intensities of 40 dB and IID's of +10dB RIGHT PANELS: model output in response to the signals displayed at the left. **Top row** - *signal intensity 80 dB - 40 dB more intense than echo* **Second row** - *signal intensity 60 dB - 20 dB more intense than echo* **Third row** - *signal intensity 40 dB - equally intense as echo* **Bottom row** - *signal intensity 20 dB - 20 dB less intense than echo*

5.2.2.2 Dependency on echo intensity

Having shown the dependency of echo suppression on signal intensities, the next set of experiments is concerned with the compliance of the model with the psycho-acoustically determined **dependency on echo intensities**. To test this dependency, similar stimuli have been employed. Once more they consisted of two 20 ms voice pulses with IID of +10 dB for the first pulse (signal) and -10 dB for the second pulse (echo). The silent gap in between the pulses was also kept at 10 ms, making the echo appear partly within the "normal" echo threshold of 30 ms.

The varied parameter in this set of experiments was the absolute intensity of the echo. Here, it has been changed between 90 dB (40 db more intense than the signal) and 10 dB (40 dB less intense than the signal). Since the IID of signal and echo has always been kept at +10 dB / -10 dB, the directional sensation should not be affected. However, the echo suppression should be much larger in case of a muted echo (10 dB) than in case of an intens echo (90 dB). The presented stimuli and the directional output of the model for selected echo intensities are displayed in figure 5.15.

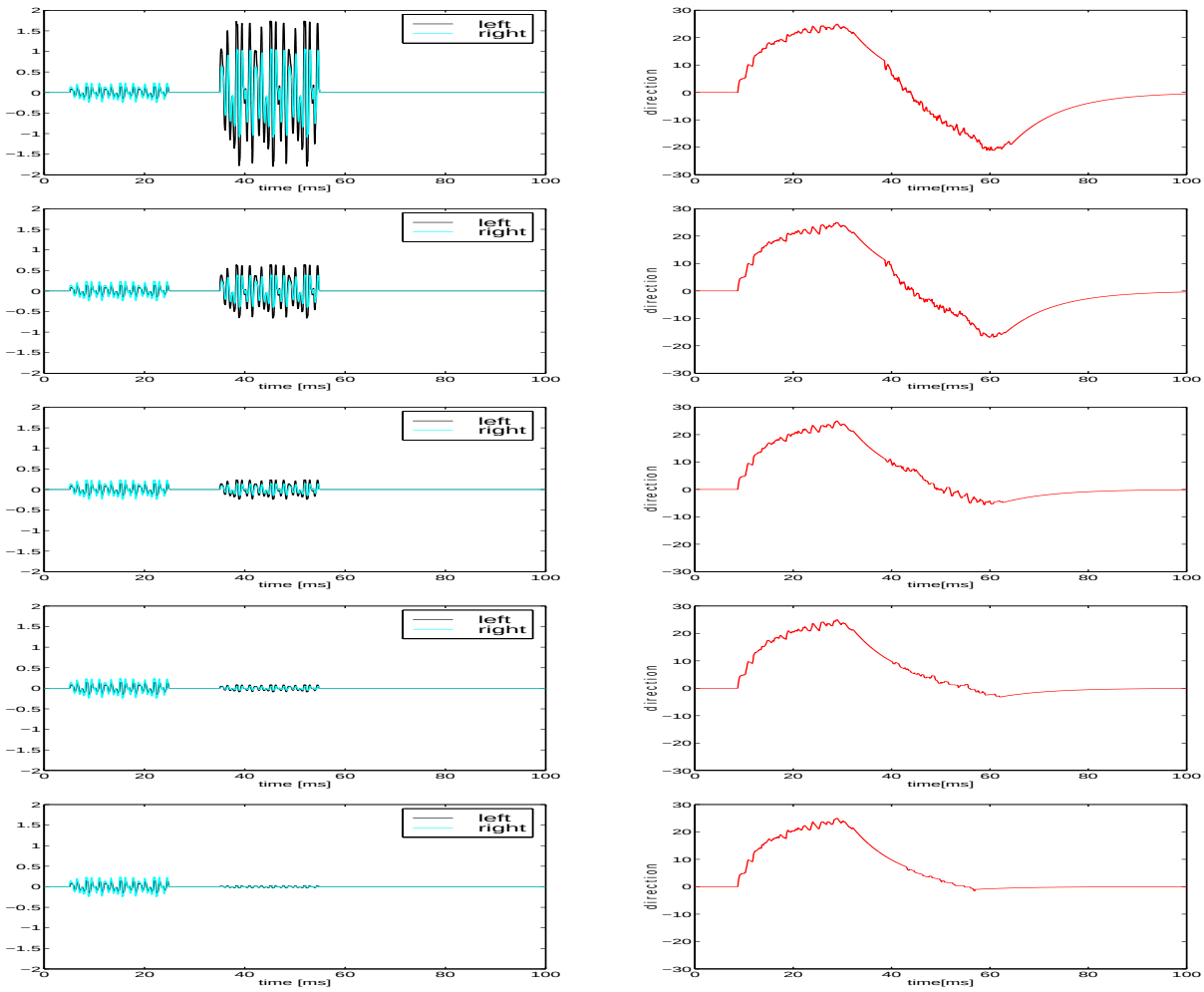

Figure 5.15: Echo suppression dependency on echo intensity - LEFT PANELS: presented stimuli: signals (first pulses) with IID's of +10dB and 50 dB SPL - echoes (second pulses) with IID's of -10dB with right (cyan) channels always 10 dB louder than left (black) channels. Echoes arrive always 10 ms after the signals end with intensities between 90 dB (top panel) and 10 dB (bottom panel), decreasing in steps of 20 dB RIGHT PANELS: model output in response to the stimuli shown at the left

By evaluating the models output it becomes clear that the effect of an intense echo is just opposite to the one of an intense signal. In case of a very intense echo (top panel), persistent inhibition, caused by the signal, is not strong enough to prevent the system from sensing it's direction and perceiving it as a separate acoustic event. As the echo intensity decreases - from top to bottom - the models response is more and more suppressed, resulting in a completely suppressed directional information at echo intensities of 10 dB in the bottom panel of figure 5.15. This behavior can again be interpreted as a change of echo threshold from less than 10 ms at the top to more than 30 ms at the bottom, depending on echo intensity. The results comply with the main findings of Lochner and Burger [LB58] using 25 dB speech signals as discussed at page 36.

Overall, the results of the two sets of experiments shown in figure 5.14 and 5.15 display the capability of the model, resembling the dependency of echo suppression on the absolute intensity of the signal as well as the echo.

5.2.3 Feature 3: Dependency of echo suppression on Inter Stimulus Delay

The most prominent feature of echo suppression is its dependency on the distance between signal and echo, described as inter stimulus delay (ISD). The perceptual effects that were replicated here, have been introduced in section 2.3 and summarized in table 2.1. Depending on the inter stimulus delay, three effects named *Summing Localization, Echo Suppression (Precedence Effect) and Discrimination Suppression* have been identified by many psycho-acoustical experiments. This subsection explains the results of three sets of experiments, displaying the capability of the model to replicate these effects.

5.2.3.1 Summing Localization

It has been mentioned in section 2.3 that at ISD's of 0 ms between two identical stimuli arriving from different directions, they are perceived to origin from one common source at a virtual location just in the middle between the two sources. This well known **stereo effect** is caused by summing localization. The models capability to duplicate this psycho-acoustic effect will be shown by the first set of experiments. It has also been mentioned, that as the ISD increases from 0 up to 1 or 2 ms, the direction of the virtual sound source is perceived to move toward the direction of the leading source. It will be shown here that the model is capable to sense ISD's as low as 50 μs and indicates a small move of the virtual sound source away from the midline toward the location of the leading source.

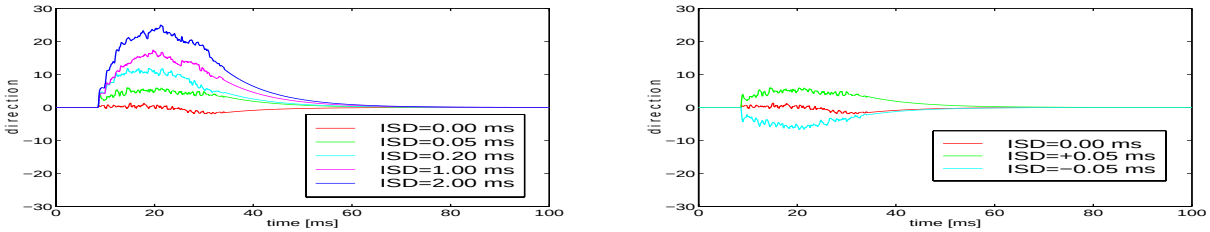

Figure 5.16: Summing localization - dependency on inter stimulus delay
LEFT PANEL: sensation of virtual sound source locations - moving from the midline at ISD = 0 ms towards the real location of the leading source at ISD=2 ms
RIGHT PANEL: sensation of virtual sound sources works in both direction

Figure 5.16 shows the model output for ISD's ranging between 0 an +2 ms in the left panel and -0.05 ms to +0.05 in the right panel. The employed stimuli were taken out of the same subsets of the recorded human voice, as during the previous experiments. The difference is here that the two pulses are no more separated in time, but only in space. Hence, the IID of one pulse has been kept at +10 dB, simulating a source about 25 degrees to the right and the IID of the second pulse has been kept at -10 dB, indicating the according source at the left hemisphere. During the design of the left and the right stimulus, both pulses have been added to each other with specific, but very short delays (ISD's). This way, the physical summing effect of two sounds originating from different directions in space was generated in a controlled manner. The resulting left and right signals last only 20 ms if overlapping with no delay (ISD=0), or 20.05 ms at ISD = 50 μs for example. Logically, at ISD=0, there will be no intensity difference if two identical signals with IID of +10 and -10 dB are added without time lag. The listener as well as the model perceives the virtual source right in the middle at 0 degree (red line at figure 5.16).

In case of an ISD = +50 μs, the duration of the left and right stimulus are again equal, but the intensity difference will be + 10 dB for the first 50 μs, 0 dB for the main part of the signal and -10 dB for the last 50 μs of the stimulus. Nevertheless, this short lead IID will cause the listener to perceive the sound to originate slightly from the right as does the model (green line at figure 5.16).

As the ISD increases further, the model senses the direction of the entire sound more and more towards the location of the leading signal. Finally, at ISD's of 2 ms, it senses the correct location of the leading source as the direction of the entire sound, even if the lagging part originates from the opposite direction. This value fits fairly well with the psycho-acoustically determined lower boundary of the precedence effect, or upper boundary of the summing localization effect.

In conclusion, the model architecture is capable to replicate the stereo effect, although it was originally designed to duplicate the precedence effect, evaluated in the next subsection.

5.2.3.2 Precedence Effect

During the second set of experiments, the ISD has been further increased, while all other signal features and the design method have been kept the same. The simulated result is that for ISD's between 2 an 20 ms the entire signal is continuously sensed as originating from 25 degree right (the location of the leading sound source). The stimuli as well as the results of these experiments are shown in figure 5.17.

Visible in figure 5.17 are two psychoacoustic phenomena. First, as long as signal and echo overlap in time, the absolute intensity of the stimuli increases for the period of overlapping and causes the listener to perceive a louder signal, as usually experienced in well designed arenas or theaters (see upper three panels). However, this does not influence the correct sensation of direction.
Second, although being more intense and generated from two different sources in space, the signal is still sensed to origin from only one direction - the one of the leading sound source. The directional information of the echo is completely suppressed, even if it is fully separated in time as shown in the bottom panel (ISD = 20 ms).

The situation displayed in figure 5.17 assumes an echo (lagging pulse) with equal intensities as the signal (70dB). This is a rather unusual situation in normal live and somewhat represents a worst case scebario. Under natural conditions, echoes will normally be less intense than signals and will cause the echo threshold to increase - extending the upper boundary of the precedence effect. This principle effect, discussed in section 5.2.2 has been tested for the entire set of signals in figure 5.17 - the upper boundary of the precedence effect was found to extend up to 30 ms if echoes are 20 dB less intense than signals and even up to 40 ms in case of very weak echoes with absolute intensities, 50 dB lower than the signal.

In general, the model proves its capability to duplicate the precedence effect, by suppressing the directional information of echoes with ISD's between 2 and 20 ms.

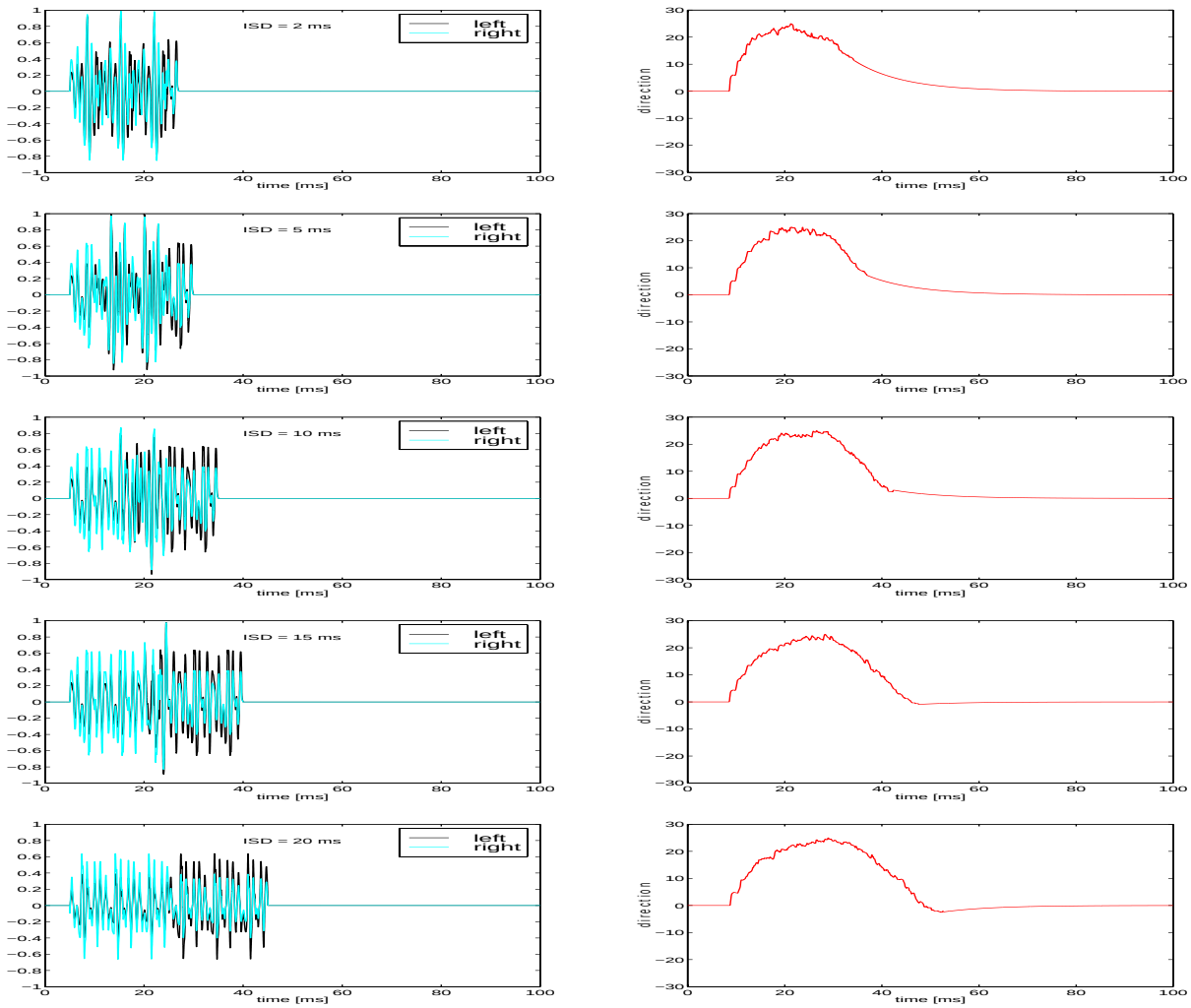

Figure 5.17: Echo suppression dependency on inter stimulus delays (ISD) increasing form 2 ms at the upper panel to 20 ms at the bottom panel LEFT PANELS: presented stimuli: right (cyan) and left (black) channels have been designed by summation of a 20 ms lead signal with IID of +10dB (from the right) and a legging echo with IID on -10 dB (from the left) both with constant intensities of 70 dB RIGHT PANELS: model output in response to the stimuli shown left

However, in nature there is no abrupt end to this effect. As the ISD increases further, the direction of the echo appears to be sensed more correctly the larger the inter stimulus delay becomes. As can been observed in figure 5.18, the effect totally disappears (in the worst case condition of equally intense echoes) only at ISD's around 40 ms. Figure 5.18 displays the directional sensation of echoes with ISD's between 20 and 50 ms, enhancing the effects of echo suppression during the transition between full directional suppression (top panel) and full sensation of echo location (bottom panel).

In conclusion, the model is capable to realize the psycho-acoustical precedence effect at Inter Stimulus Delays between 2 and 20 ms and to sense the correct location of the echo (as a separate acoustical event) at ISD's above 30 ms.

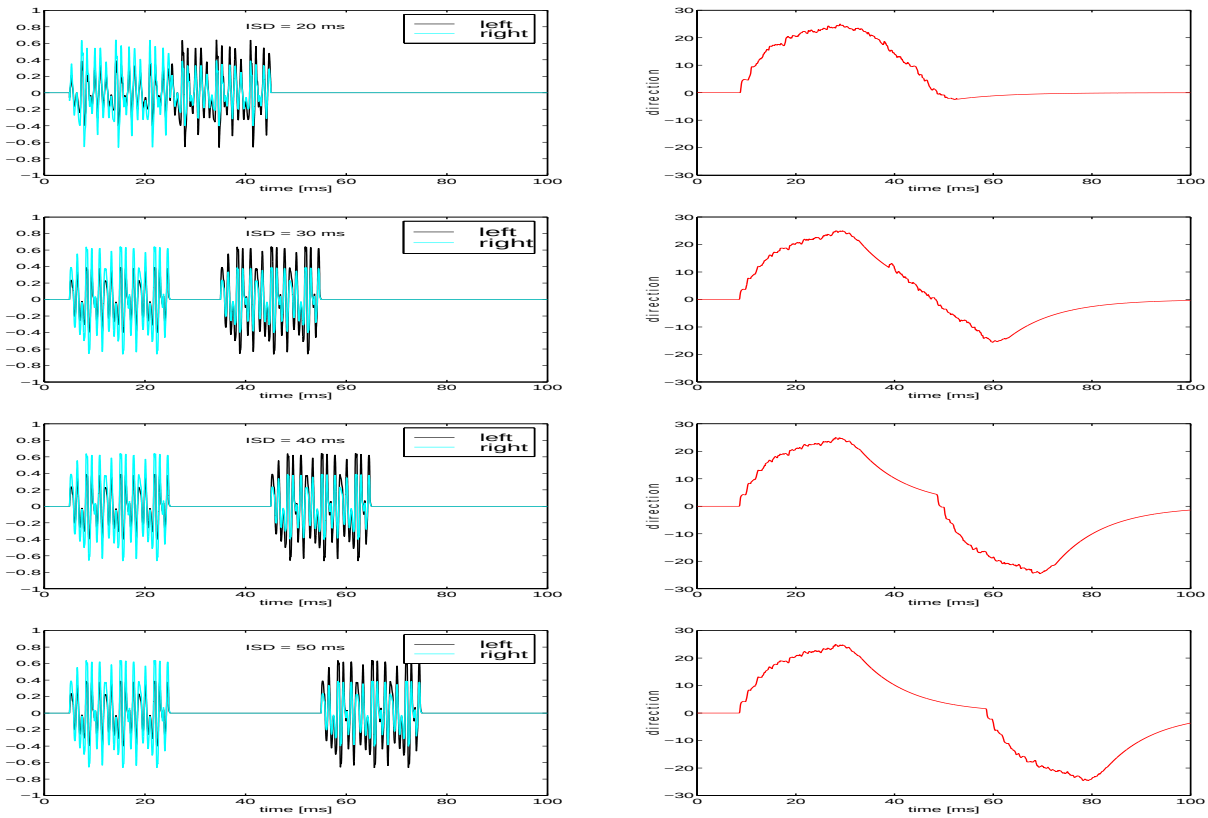

Figure 5.18: Echo suppression dependency on inter stimulus delays (ISD) between 20 ms (top panel) and 50 ms (bottom panel) LEFT PANELS: presented stimuli right (cyan) and left (black) channels have been designed by summation of a 20 ms lead signal with IID of +10dB (from the right) and a legging echo with IID of -10 dB (from the left) both with constant intensities of 70 dB RIGHT PANELS: model output in response to the stimuli at the left

5.2.3.3 Discrimination suppression

As mentioned in section 2.3, the precedence effect can be divided into two parts. First, the *Period of Complete Echo Suppression* - where only one (joined) sonic event is perceived to originate only from the location of the lead signal - and second, the *Period of Discrimination Suppression* - where two separate sonic events are perceived but only seem to originate from the location of the lead signal.

The second period is the one of interest in this subsection. Since the output of the model represents the directional sensation only, a more detailed view on the activity within the models of the left and right inferior colliculus might help to determine, under what ISD conditions two separate spike pattern are transferred to the higher auditory centers. Because the model can track the activity of single cells during any simulation, the summed activity of all frequency sensitive cells in the right and left IC has been chosen to represent the spatio-temporal pattern of the left and right IC model. It is obtained by simply summing the spikes of all 16 frequency channels at every point in time (time resolution = 10 μs). The obtained results for ISD's between 0 and 40 dB are displayed in the left panels of figure 5.19. The right panels show again the directional output of the model.

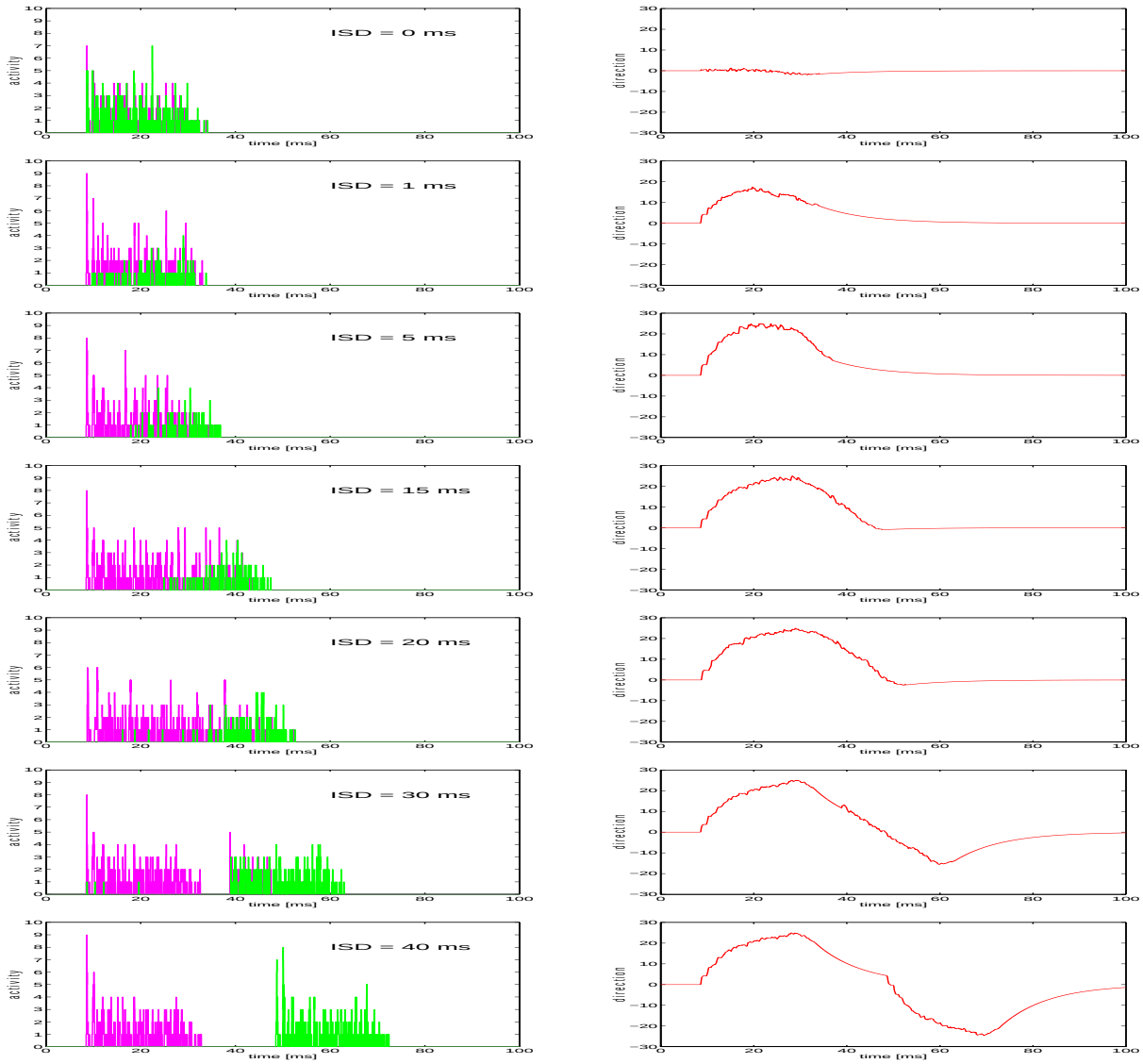

Figure 5.19: Discrimination suppression depending on inter stuimuls delays between 0 ms and 40 ms
LEFT PANELS: spike activity in the right IC model (green) and the left IC model (magenta)
RIGHT PANELS: model output in response to the signals

By looking at the spike pattern in the left panels it is easy to observe that at very small ISDs, the firing pattern of the right IC (green) as a reaction to an echo arriving from the left hemisphere almost completely coincides with the firing of the left IC (magenta) caused by the signal from the right. At ISD=0 (upper panel) even the firing intensity matches, since neither one of the IC's is inhibited. This is caused by the fact that the two signals at the right and left ear sum to exactly the same signal and, therefore, do not contain any directional information capable to excite the DNLL. At small ISD's (i.e. 1 ms in the second panel) the right IC is significantly inhibited by an active left DNLL and its firing intensity decreases. However, since both firing pattern still coincide nearly completely, this will prevent any higher auditory centers from the distinction of two separate events - the animal or human will only perceive one (joined) sonic event from the direction of the lead signal. This is obviously true during the entire summing localization period and beyond up to an ISD of 15 ms.

If the inter stimulus delay reaches about 15 ms, the right and left firing pattern is nearly separated in time. Now, the higher centers will first be excited by the left IC and only later by the right one. This may cause the brain to perceive two separate sonic events. However, since the right DNLL does not fire (caused by persistent inhibition) the left IC is not inhibited during the second (echo) part of the signal and therefore continues to fire. Consequently, no directional information is sensed and the brain perceives the two sonic events to arrive from the direction of the signal only. This situation respresents the discrimination suppression effect. Especially the 4'th and 5'th row of panels in figure 5.19 (ISD between 15 and 20 ms) display the ability of the model to sense two separate sonic events, but only one direction.

As the ISD increases further and above the echo threshold, the firing of the left IC during the second part of the signal increases more and more and finally completely suppressed at ISD=40 ms (bottom panel). Now, the firing pattern in response to the two sonic events are not only separated in time but also equal in intensity. Therefore, they are perceived as two separate sonic events arriving from two different directions.

The general hypothesis is that the timing of the response pattern to the same signal in the left and right IC causes the perception of sonic events, while the intensity difference of those firing pattern (modulated by persistently inhibited DNLL cells) accounts for the perception of different directions. Since the timing strictly follows the acoustical pattern and the intensity is inversely modulated by the DNLL, directional perception is generally longer suppressed than the perception of sonic events.

Although the nature of sonic events perception is not discovered by now, the hypothesis stated above might be valid for further evaluation based on the model architecture as well as defined physiological or psycho-acoustical experiments at the living brain.

Summarizing this subsection on ISD dependency of directional perception, it can be stated that:

- *The model architecture is capable of duplicating the summing localization effect of directional perception at ISD's between 0 and 2 ms*
- *The model architecture can duplicate and explain the major features of the precedence effect by suppressing directional information of echoes arriving 2 to 30 ms after the signal.*
- *Based on the interpretation of the intrinsic firing pattern in the IC model, a hypothesis on the distinction between complete echo suppression and discrimination suppression has been formulated.*

Since the aim of this subsection is to test specific psycho-acoustical features, only synthetic signals, with defined features, have been employed to generate the displayed outputs. However, similar experiments have also been carried out, based on free field recordings, and in principle these have not generated any different results.

5.2.4 Feature 4: Longer signals cause longer echo suppression with limits around 20 ms

The feature to be investigated here is the dependency of echo suppression on the duration of the signal. Based on the summary of many psycho-acoustical experiments concerned that topic Blauert's textbook [Bla74] at pg.184 states that:

- the shortest echo thresholds are observable around 2 ms for short clicks of 1-5 ms duration and
- for longer, slow raising speech signals, the echo threshold increases and remains relative constant around 20 ms

The evaluation of the model architecture in this section will, therefore, be limited to these two landmarks, accepting that many other experiments would be necessary to prove a general compliance of the model's behavior with nature.

To evaluate the signal duration dependency of echo suppression, a set of special stimuli has been designed. Once more, they consisted of a leading signal and a lagging echo represented by subsets of the recorded male voice with signal IID's of +10 dB (25 degree right) and echo IID's of -10dB (25 degree left). In order to visualize the duration of echo suppression, this time the echo lasted much longer than under normal conditions and always continued till the end of the simulated 100 ms period. Furthermore, the echo always started right at the end of the signal, hence, the ISD has always been identical to the signal length. These somewhat artificial conditions, have been chosen to enhance the persistent inhibition of the echo after the signal ended and no further inhibitory stimulus was present.

In order to determine the duration of the echo suppression period, the end of the signal was used as starting point and the "break through" point of the echo was used as end point of this period. The "break through" point was defined as that point in time, when the long lasting echo was first time sensed - i.e. the directional sensor identified an acoustic event from the opposite (negative) site. However, as can been seen in figure 5.20, if the echo is sensed for the first time, the correct echo location is not yet identified and the sensor may need an additional time frame of 15 - 20 ms to reach the correct direction. Since the mentioned psycho-acoustical experiments define the echo threshold as the point in time, when the echo is sensed for the first time, regardless whether it is perceived at the correct position, the zero crossing of the directional model output was used to determine the end of the echo suppression period during these experiments.

To enhance the visibility of the suppression period this time the stimulus as well as the directional output were combined in one drawing. To make the stimuli visible in this drawing, they have been scaled by the factor 10 for display reasons only. Furthermore the intrinsic delay of 5 ms between model input and output was compensated by a delay of 5 ms applied to the stimuli. Under these conditions, figure 5.20 presents the achieved results for signal durations between 0.5 and 50 ms.

As can be observed in the left panels of figure 5.20, for short signals (clicks) with durations up to 5 ms the echo threshold significantly shortens below 20 ms. However, even at clicks of only 0.5 ms length it never reaches the 2 ms mentioned by Blauert and stays around 5-6 ms. The most likely reason for this it the lack of volatility in the directional output element. As described in section 4.7, this element does not model a biological structure, but has been designed to generate a technical signal, usable to control motors etc. In order to achieve a stable signal the pulsing spikes have been smoothed by a PT_2 element with decay times of 70 ms.

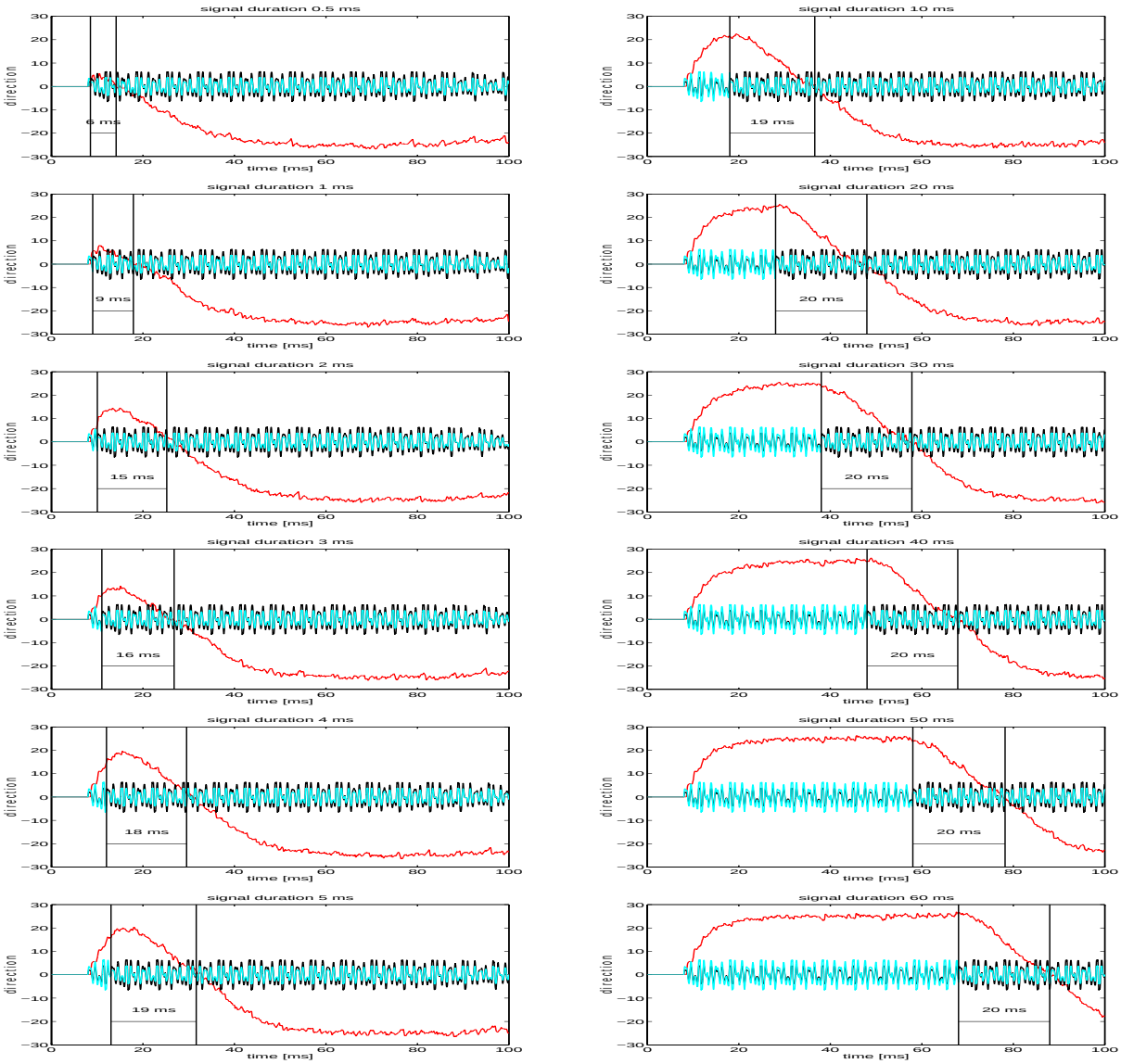

Figure 5.20: Echo suppression dependency on signal duration
LEFT PANELS: signal durations between 0.5 and 5 ms cause echo thresholds below 20 ms
RIGHT PANELS: echo thresholds remain constant at 20 ms for signal durations above 10 ms

This generates a fairly smooth output signal but limits the volatility of the output and, therefore, the minimal echo threshold. However, a slightly increased echo threshold for very short clicks will not limit the applicability of the designed architecture to technical systems - it might even enhance its value, since for short clicks, the aim is rather to suppress than to sense the echo.

The right panel of figure 5.20 shows that for longer speech signals - remember that the stimuli were designed out of a recorded human voice - the echo threshold remains at 20 ms and does not increase as the signal duration increases above 20 ms. The explanation for this phenomena is easy to observe when investigating the internal potentials of DNLL cells causing echo suppression based on persistent inhibition. Figure 5.21 exhibits the internals of a single cell in the right DNLL (channel 11 with BMF at 1.3 kHz).

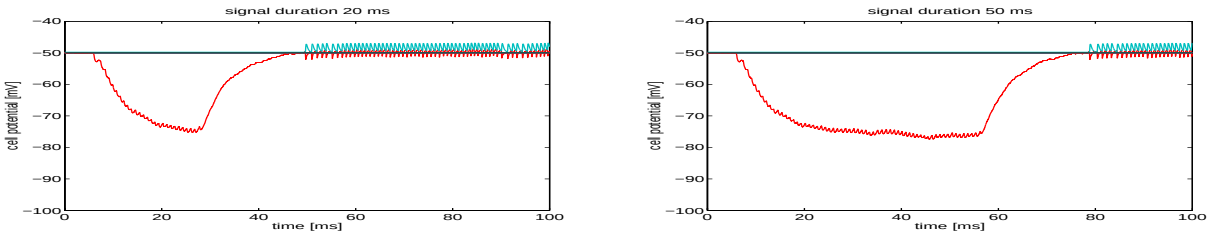

Figure 5.21: Persistent inhibition dependency on signal duration
LEFT PANEL: soma potential (red line) of a single cell in the right DNLL in response to a signal of 20 ms duration followed by a long lasting echo
RIGHT PANEL: soma potential (red line) of a single cell in the right DNLL in response to a signal of 50 ms duration followed by a long lasting echo

It is clearly to observe that this cell is hyperpolarized during the signal part of the stimuli. This is due to the fact that a signal arriving from the right causes the right LSO as well as the left DNLL to fire and both inhibit the right DNLL. During the first 20 ms of the signal, hyperpolarization is still increasing, until it reaches it's maximum level of possible hyperpolarization at -77 mV and remains at this level regardless how long the signal lasts. If the signal ends, the cell potential returns to its -50 mV resting potential. The time needed for this depends on the decay time constant $\tau = 7ms$ (see section 4.5 - DNLL Model) as well as the hyperpolarization level reached at this point in time. Since the reached level is equal for 20 ms signals as well as 50 ms signals, the time needed to return is also equal and the firing threshold of that cell will be reached with the same time lag (period of persistent inhibition) in both cases. If the DNLL cells start to fire after this period of about 20 ms it immediately inhibits the referring cell in the left IC and prevents it from further firing. Finally, the macroscopic result is that the the right (un-inhibited) IC keeps firing, while the left one stops to do so. Consequently only at this time the directional sensor moves toward the direction of the echo source.

This brief excursion into the details of single cell potentials explains not only the independence of echo threshold from signal durations above 20 ms but also the shortening of this threshold if the signal duration is below 20 ms. This situation is shown in figure 5.22.

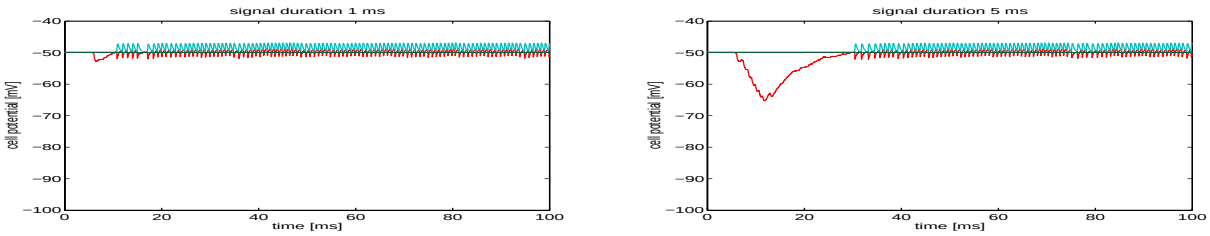

Figure 5.22: Persistent inhibition dependency on signal duration in case of short pulses LEFT PANEL: soma potential (red line) of a single cell in the right DNLL in response to a pulse of 1 ms duration followed by a long lasting echo RIGHT PANEL: soma potential (red line) of a single cell in the right DNLL in response to a pulse of 5 ms duration followed by a long lasting echo

It can be seen in the left panel of figure 5.22 that in case of a 1 ms pulse, the cell is hyperpolarized only by a little and quickly returns to its resting potential. It starts to fire already 4 ms after the signals end and the mechanism described above is activated. As the signal duration increases to 5 ms - shown in the right panel - the cell gets mre hyperpolarized and it takes as long as 15 ms after the signals end until the cell starts to fire. Since the saturation level on -77 mV hasn't been reached yet, the period of persistent inhibition still remains below 20 ms.

This simple biological principle causes the entire system to exhibit so many of the psycho-acoustically determined features and accounts for the underlying mechanism, realizing the echo suppression in this model.

To summarize this set of experiments it can be stated that:

- *The first statement of Blauert is only partly fulfilled by this model, since echo threshold shortens in case of short clicks of 1-5 ms duration, but the shortest observable echo thresholds stays around 5-6 rather than 2 ms.*
- *The second statement of Blauert is completely fulfilled by the model, since for longer speech signals the echo threshold increases and remains stable around 20 ms.*

Of course, this short set of experiments cannot claim to prove a general ability of the model to duplicate all of the psycho-acoustical duration dependency effects of echo suppression. However, for the aim of this study, the number of experiments to investigate this dependency have been limited, as it is not critical for the application of the model to robots or other technical systems for echo suppression during sound source localization.

5.2.5 Feature 5: Echo suppression in case of sounds from the midline

It has been argued by Litovsky [Lit97] that the suppression of echoes evoked by signals arriving from the midline of the listener depends on spectral cues rather than binaural cues (see page 44). The reasoning behind is that in case of midline signals there are no binaural cues to be detected within that signal, so the LSO as well as the MSO will not become activated and can not contribute to the suppression of echoes arriving several milliseconds later from left or right directions. The experiments performed in this subsection, will display why this is not covering the whole story, and why signals without binaural cues can suppress echoes as well.

In order to explain the capability of the model architecture to suppress echoes in case of sounds originating from the midline, the reaction of the several model stages to a pure sound without echo is shown in figure 5.23.

If the sound arrives from the midline it can be assumed that it contains identical waveforms and timing at the left and the right ear. Therefore, an identical 50 ms subset of the well known voice sound has been presented to both ear models.Although the two sounds do not contain binaural cues they activate both cochlear nuclei (not shown) and both LSO models (4.row) in an identical manner. While the EI cells of the LSO usually react distinctively to sounds arriving from the left or right hemisphere (see section 4.4.3), they both will exhibit some activation in case of a midline sound, since ipsilateral excitation slightly exceeds contralateral inhibition. The result is

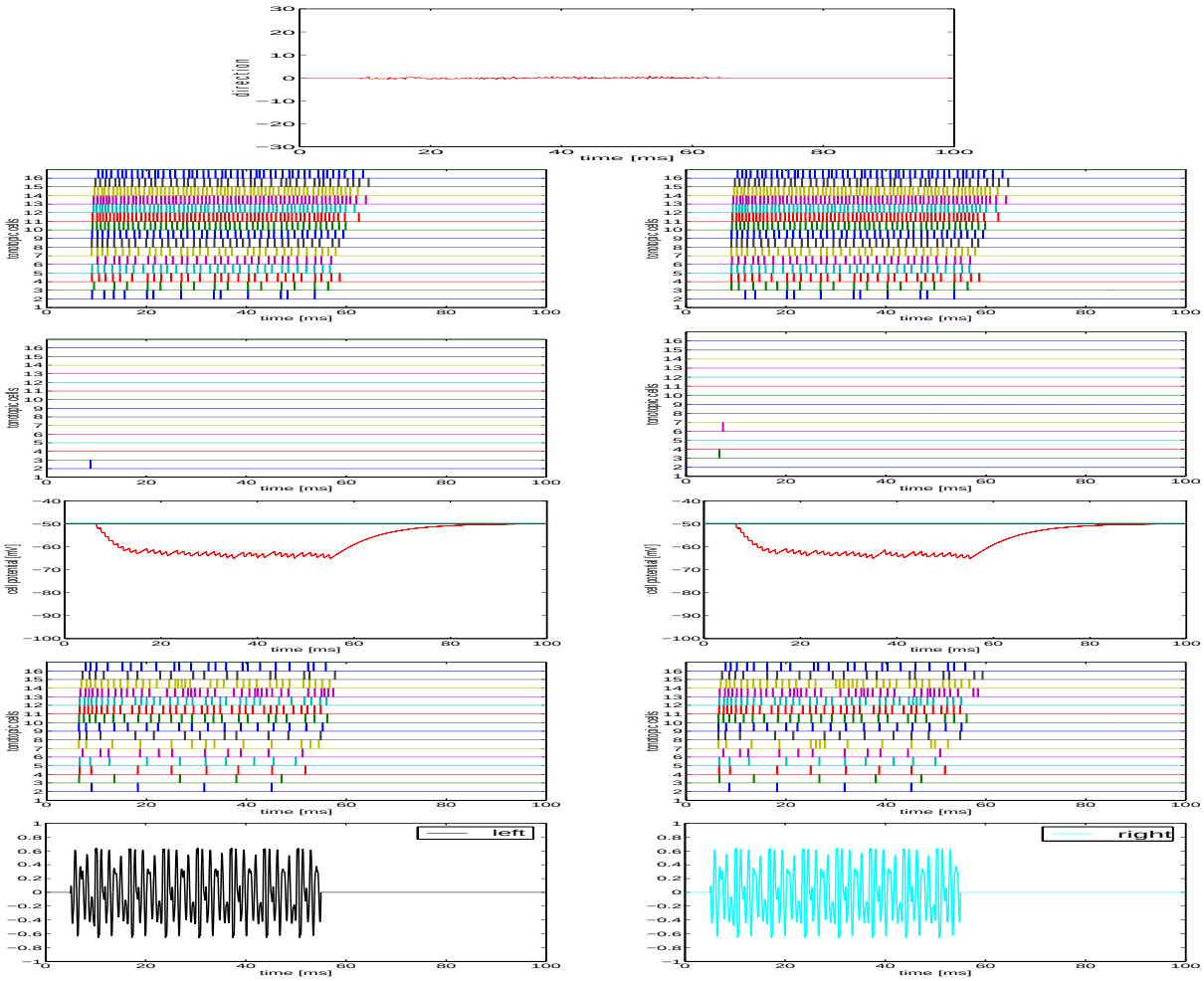

Figure 5.23: Model activity in response to a signal arriving from the midline - bottom to top:
1. row: left and right stimuli - identical in case of midline sounds
2. row: spike pattern of the left and right LSO - due to stronger excitation than inhibition
3. row: hyperpolarized soma potential of corresponding left and right DNLL cells (channel 11)
4. row: silent reaction of the left and the right DNLL - due to hyperpolarization
5. row: identical spike pattern of the left and right IC - no inhibition is present
top panel: directional model output - indicating a midline sound

that although no binaural cues exist, there is a slight activation of nearly all frequency channels in both LSO models. This activation causes a hyperpolarization of the subsequent DNLL cells (see 3.row and 4.row in figure 5.23). Since their ipsilateral excitation is less effective than their contralateral inhibition (see section 4.5.2), the DNLL cells become slightly hyperpolarized down to about - 60 mV as shown in figure 5.23 - 3.row of panels.

It has been perceived by Litovsky et al. that neither one of the cats DNLL's responded to a sound from the midline. This also holds for the model as shown in the 4.row of panels in figure 5.23. From that perception they concluded that the DNLL can't be involved in echo suppression in case of midline signals. However, what can't be seen during pure spike recordings is, if the cells are hyperpolarized as hypothetically concluded from the model architecture. To prove this hypothesis it would be a helpful experiment to record the soma potential rather than the spike response of DNLL cells under the condition of identical left-right stimulation.

If this hypothesis is true, the hyperpolarized DNLL cells are prevented from firing and will not inhibit either one of the IC cells. Consequently, they both will react in the same way, as shown in detail at the 5. row of panels in figure 5.23. The top panel displays that the directional model output will sense the correct midline position of the sound, despite heavy activation of both IC models.

Figure 5.23 introduced the basic case of a single voice signal from the midline. What happens if this midline sound generates an echo with Inter Stimulus Delay of 20 ms is shown in figure 5.24.

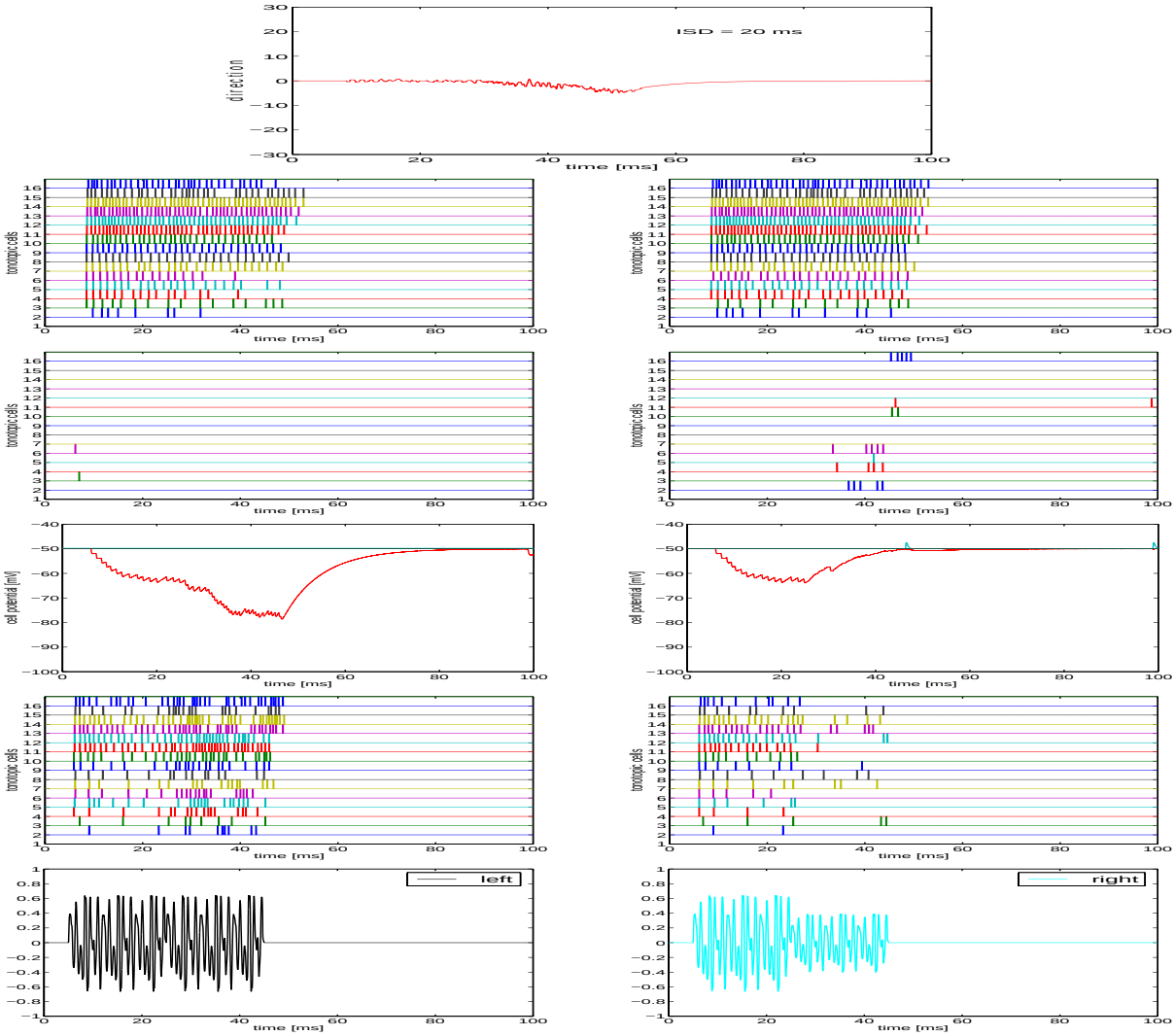

Figure 5.24: Activity within the model in response to a midline signal, followed by an echo from the right after 20 ms - BOTTOM PANELS: left and right stimuli - identical during the signal and with an IID of +10 dB during the echo part 2.ROW: spike pattern of the left and the right LSO 3.ROW: soma potential of corresponding left and right single DNLL cells (channel 11) 4.ROW: spike pattern of left and right DNLL with persistent inhibition of the right DNLL 5.ROW: spike pattern of left and the right IC, inhibited during the very end of the stimulus TOP PANEL: directional model output

The main difference between figure 5.23 and figure 5.24 is the fact that only the first 20 ms of the stimulus arrive from the midline (are identical), the second part in figure 5.24 represents an echo arriving form 25 degrees right and therefore contains an IID of +10 dB.

The reaction of the different model stages in figure 5.24 clearly exhibits, how and why the directional information of the echo is mainly suppressed and an echo is only sensed at the very end of the stimulus. During the first 20 ms of the stimulus everything is identical to the situation of figure 5.23. But when the echo arrives after 20 ms, DNLL cells at both hemispheres are hyperpolarized to about -60 mV. Since the echo part contains binaural information, the left LSO reacts much stronger that the right LSO indicating a sound arriving from the right. However, the right DNLL cells, supposed to fire, are hyperpolarized and need more than 15 ms to return to its resting potential. Only at the very end of the stimulus the soma potential of the DNLL cell has overcome hyperpolarization, caused by the initial midline signal and starts to fire. Looking at the entire structure, only some frequency channels start to fire. As expected, the left DNLL cells are further hyperpolarized down to the maximum level of -77 mV.

The result of this persistent inhibition of DNLL cells is visible in the reaction of the IC models (5.row of panels in figure 5.24). Here, the spikes of the left IC are not diminished up to the 40 ms point. It is only at the very end (outside the echo threshold), when the number of spikes in the left IC decreases compared to that of the right IC. Therefore, the directional sensor moves slightly toward the right direction, but suppresses the directional information of the echo as long as 15 ms after the signal has ended (equal to an ISD of 35 ms).

This interim situation, just exceeding the echo threshold, has been chosen to exhibit the principle of echo suppression in case of signals arriving form the midline. To test the general ability of the model to completely suppress the directional information of echoes caused by sounds from the midline, the experiments of subsection 5.2.3 have been repeated with signals of IID = 0 (arriving from the midline) and echoes with IID = +10 dB for increasing ISD's between 1 and 40 ms. The results are displayed in figure 5.25.

As can been seen for echoes arriving up to 15 ms after the midline signal (top 3 rows of panels) the directional information of the echo is completely suppressed. At ISD's between 20 and 35 ms (4. and 5. row of panels) the echo is sensed to arrive from the right but not yet at the correct location. Only at ISD's of 40 ms the echo is sensed as a separate signal from its correct location (25 degree right).

This set of experiments clearly shows that echo suppression is also possible for sound signals arriving from the midline of the listener. The internal model parameters shown in figures 5.23 and 5.24 explain the reason and the underlying hypothesis of hyperpolarized DNLL cells due to non-symmetric EI units at the level of the LSO.

However, comparing the results of figure 5.25 to the ones of figure 5.16 at page 146 it becomes clear that echo suppression in response to sound from the midline is somewhat less effective than if the signal arrives from the opposite direction. While the echo threshold for sounds from the midline ranges between 15 and 20 ms, the one simulated in figure 5.16 was determined between 20 and 25 ms.

This is due to the fact that stand alone inhibition (only by the ipsilateral LSO) will not cause the DNLL cell to hyperpolarize as strong as if both, the ipsilateral LSO and the contralateral

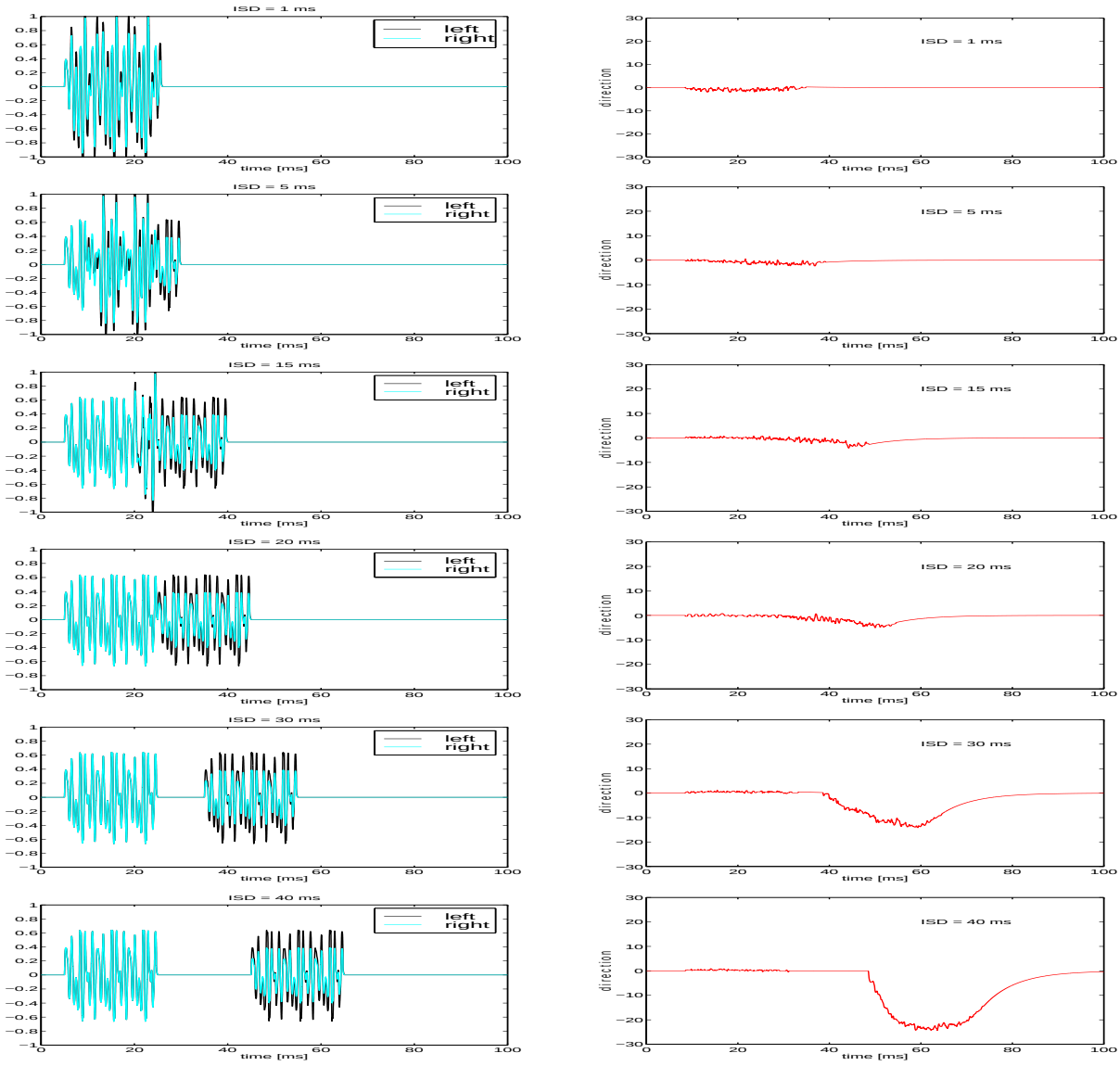

Figure 5.25: Echo suppression by signals from the midline for varying ISD's between 1 and 40 ms LEFT PANELS: left and right stimuli resulting form the overlap of a 20 ms midline signal with an echo from the left RIGHT PANELS: reactions of the directional model output in response to the stimuli at the left

DNLL inhibit the cell. Of course, for technical applications the synaptic parameters could be changed setting the relation between the two components to any number, but for the purpose of this study, biological relevance has been kept in mind and the parameters have been chosen not to hyperpolarize any cell stronger than by 27 mV (down to a soma potential of -77 mV).

In conclusion, the suppression of echoes caused by sounds from the midline of the listener is possible and can be explained by binaural interactions within a model of the auditory brainstem. It's effectiveness is slightly less than in the case of signals arriving from either one of the left or right hemispheres.

5.2.6 Preservation of spectral information during echo supression

The last psycho-acoustical feature investigated in this section is whether the model is capable to preserve the spectral information of echoes, while their directional information is completely suppressed.

It has been observed by the experiments of Fitzpatrick et al. [FKBT95] that specific (early high) IC cells in the rabbit responded to echoes, although they where not perceivable as separate acoustic events. These experiments as well as the fact that echoes are used to determine the distance of the source and the shape of the room, even if they are not perceived as separate sounds lead to the assumption that the echo information still causes appropriate responses in the IC, while their directional information is suppressed.

In order to test this feature, a somewhat artificial stimulus has been created and presented directly to the model input. It consisted of the usual 20 ms subset of the recorded human male voice with an IID of +10 dB but an artificial echo, consisting of a muted pure sinus of 1.3 kHz with IID of -10dB. The presented stimuli are displayed in figure 5.26 lower panels.

Since the spectral information of the relatively weak echo is totally different from the signal, the spectral information of that echo should be possible to identify in the spike pattern of the IC, while the directional output does not sense it. As can be seen in figure 5.26 both is true.

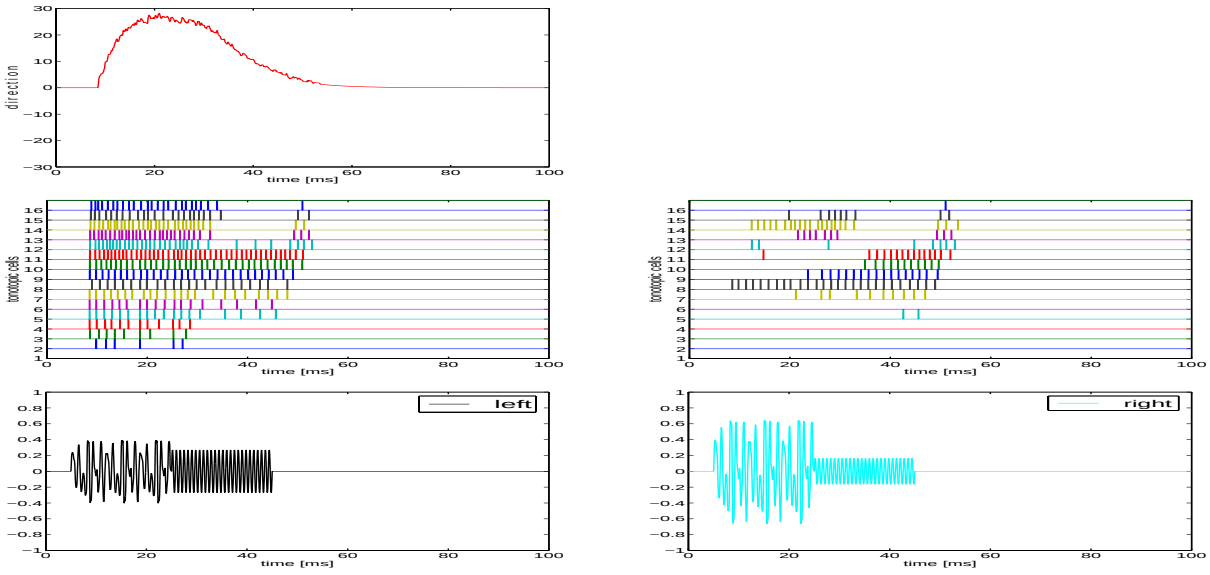

Figure 5.26: Preservation of spectral information during suppression of echo direction TOP PANEL: directional output - suppressing directional echo information MIDDLE PANELS: spike pattern of the left and right IC model - clearly showing the echo response BOTTOM PANELS: left and right stimuli

As observable the specific spike pattern, generated by the sinusoidal echo is clearly visible in both IC models during the second part of the signal, even if the directional output completely suppresses the directional information of that echo.

This effect has been tested with several other stimuli as well and clearly proved the capability of the model, to preserve spectral information of echoes during the suppression of their directional information.

Summary

The psycho-acoustical experiments presented in this section proved the hypothesis that the designed neural architecture for echo suppression during sound source localization is capable of modeling the 6 psycho-acoustical effects mentioned at the beginning of this section.

By modeling a specific subset of cell connectivity and dynamic behavior, the designed architecture cannot only duplicate these proven psycho-acoustical effects, but can also explain some of them based on hypothesis derived from the evaluation of intrinsic cell parameters during the processing of dynamic signals.

However, the current model will not at all cover the whole phenomena of echo perception. Important features like the evaluation of the reverberation radius or spectral modifications between signal and echo have not been taken into account yet. On the other hand, the promising results with regard to:

- Echo suppression under free field conditions
- Intensity dependent echo suppression
- Inter Stimulus Delay dependency of echo suppression and summing localization
- Duration dependency of echo suppression
- Echo suppression in response to midline sounds and
- Preservation of spectral information during echo suppression

suggest that further extension based on models of additional biological features and structures will be able to extend its ability into the sensation of spatial effects, as well as the usage of spectral cues.

For the purpose of an application to mobile robots these features might be of additional value, but have been out of scope in this thesis.

5.3 Real World Experiments

While the first and second section of this chapter were concerned with the duplication of physiological spike pattern and psycho-acoustical effects, this last section will test the ability of the model to cope with the real world. Therefore, the signals are no longer synthetic waveforms based on sinus waves or modified recordings. Instead here, we used purely real world signals, recorded under normal and especially reverberating conditions and presented them directly to the models input with no more pre-processing than a simple A/D conversion (by a Cirrus Logic Crystal Audio Codec Sound Card) and a linear scaling of amplitude to levels between 0 and 2 mV.

For the set of real world experiments three types of signals have been employed in two different environments. They where recorded up front and digitally stored as .wav files in order to become repeatable. The three types are displayed in figure 5.27 and contain increasingly complex sounds.

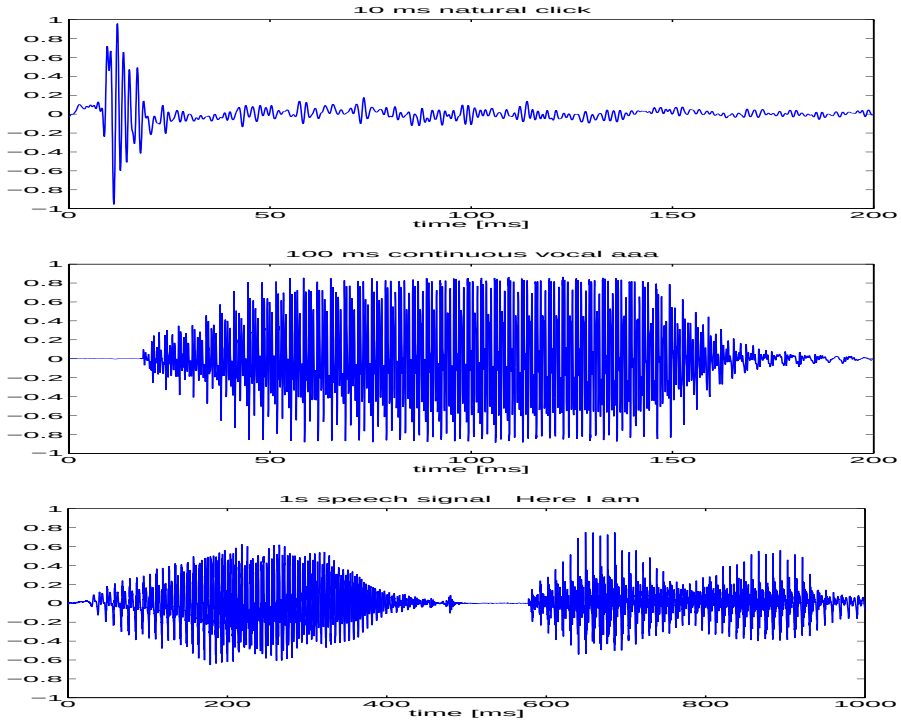

Figure 5.27: Pre-recorded signals used for real world experiments
upper panel *signal 1 - short natural click generated by a finger snip*
middle panel *signal 2 - continuing voice signal singing the vocal "aaa"*
lower panel *signal 3 - dynamic speech signal speaking the words "Here I am"*

The first signal (shown in the upper panel) is a natural click of only 10 ms length, generated by a finger snip and will be used to test the models ability to sense the direction of click like sonic events. The second signal (middle panel) contains a human male voice singing continuously the vocal "aaa" for about 100 ms. (Since this type of continuous signals has also been employed to track moving sound sources, another recording of the same type lasting for more than 1 s has been stored.) The third signal (lower panel) contained the words "Here I am" representing a highly dynamic speech signal of about 1 s duration.

These 3 types of signals have been send to a **single** Altec-Lensing speaker positioned about 1.5 m in front of a Teddy with a head base (ear distance) of nearly 20 cm - similar to the base of a human head. As already described, the head of the Teddy creates a fairly natural acoustic shadow (IID) and interaural time delay (ITD), as introduced in section 2.1 Interaural Disparities. The ears of the Teddy have been equipped with two Sennheiser K6 capacitor microphones, recording the real world sound while the signal was played by the speaker. By the help of an Toshiba SY-C12 stereo pre-amplifier, a Cirrus Logic Crystal Audio Codec Sound Card and the Cool Edit Pro stereo recording software, both channels have been digitalized and stored simultaneously. Later on, the recorded stereo .wav files have been converted into a MATLAB readable format and presented to the left and right input of the model.

Figure 5.28: Experimental setup for the recording of real world signals, using a Teddy with an ear base of 20 cm and two microphones in front of a single speaker at about 1.5 m distance.
left panel *office environment with normal echoic conditions*
right panel *tailed bathroom environment with highly reverberating - worst case - conditions*

This experimental setup was used in two different environments:

A normal environement - a small office room (lab) - something a robot would have to deal with under real world conditions

B worst case environment - a tiled bath room with strong echoes - a worst case condition for technical systems

To evaluate the performance of the model under normal real world and worst case conditions, the signal was always presented from 7 different directions with regard to the Teddy. As can bee seen in figure 5.28, the Teddy was turned from the left to the right in steps of approximately 30 degrees covering angles of right 90, right 60, right 30, 0 (midline), left 30, left 60 and left 90 degrees with regard to the stable sound source. Since the angel was determined by a simple mechanic measurement, it had a accuracy of about 10 degrees. The Teddy was repositioned during all experiments, as a mobile robot would do in the real world during the localization of sound sources.

To test the opposite, the Teddy was kept stable at the midline position and the sound source was moved, playing either signal 2 (the continuing tone) or signal 3 (the speech). This way the ability of the model to track a moving sound source, or a moving speaker was tested.

In order to achieve a stable output signal, capable to steer for example a motor, for all real world experiments the motor output of the model was used instead of the directional output in the previous section. As one might remember from section 4.7 Directional Sensor, it filters the directional output through a first order Butterworth Low Pass filter with a cut off frequency of 1 Hz, generating a slower reacting but smooth output of the model.

Furthermore the directional output element has been tuned to the specific conditions of the experimental setup by setting the integration time of the directional sensor to 4 ms, its weight to 0.1 and the calibration factor to 3.0. This must be re-calibrated if used with a different system, based on the specific shadowing effect between the two microphones (ears). Here, it has been calibrated to the shadowing of the Teddy-head and the employed Sennheisser microphones.

The following subsections will now display and discuss the obtained results for 5 different experiments in the two different environments. These experiments will test the following features:

1. Position of a click
2. Position of a continuing voice
3. Position of a speaker
4. Tracking of a moving continuing voice
5. Tracking of a moving speaker

5.3.1 Position of a click

The first set of experiments used the natural click signal to determine the capability of the model to detect click positions relative to the ear base. Therefore, the recorded finger snip was played by the speaker and the stimuli were recorded from the two ears (microphones). After normalization the recorded stimuli have been presented to the models input. This procedure was repeated under normal condition - in the office - and worst case conditions - in the bathroom. The recorded wave forms (stimuli) in response to the natural click are displayed in figure 5.29.

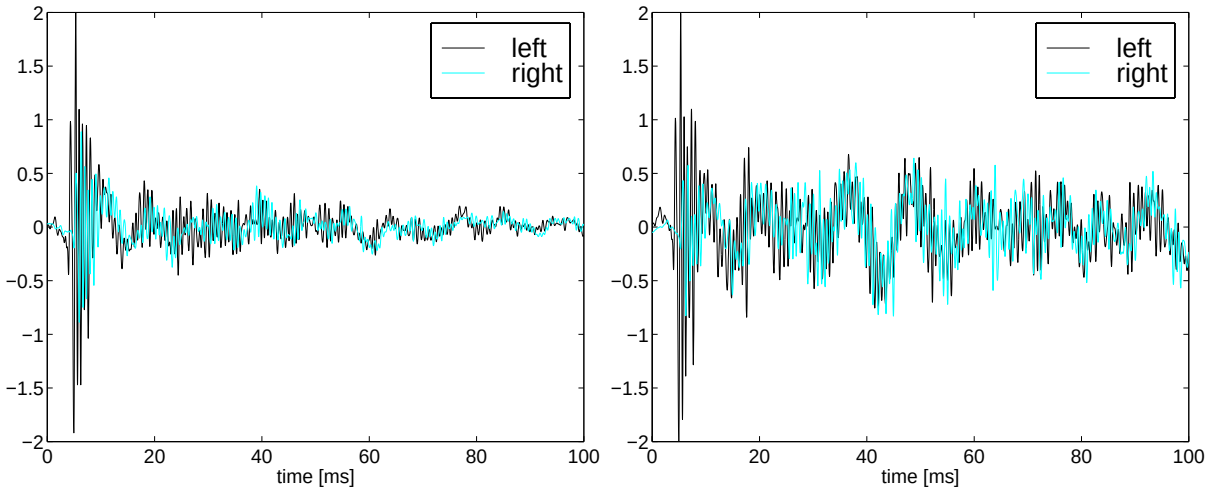

Figure 5.29: Recorded waveforms in response to a natural click under normal (left panel) and worst case (right panel) conditions

The left panel of figure 5.29 shows the recorded waveforms under real world conditions and displays the echoic "tail" caused by echoes, arriving up to 60 ms after the signal has ended. The recording was obtained in the office environment shown in figure 5.28 and displays the physical existence of echoes (during the tail of the click) which are not perceived by a human listener.

The specific feature of a small tailed bathroom (worst case environment) is that it generates strong echoes which return for several times from different directions with respect to the Teddyt's ear base. Since these echoes are of high intensity, they are clearly perceivable by humans (i.e during listening to the recordings), but do not disable the human ability to detect and hold the correct sound source location.

Recording the sound arriving at the two ears, under worst case conditions leads to the waveform displayed in the right panel of figure 5.29. It can be observed that the echo is much stronger, not only because the amplitude of the echoic tail is largely increased, but also because it lasts much longer, well beyond the 100 ms displayed here. These high amplitudes are obviously caused by many strong echoes arriving from the different walls overlapping in time and space, and generate a quite high degree of de-synchronization after the original signal has ended.

The question to be answered in this section is the following: Is the developed model of the auditory brainstem capable, to suppress the directional information of these echoes, as humans can do, and technical system need to do?

Looking at the firing pattern within the different model stages in case of a sound source position 60

degrees left from the midline provides the answer. The two recorded stimuli and the coincident firing pattern of both LSO, DNLL and IC models as well as the model outputs are therefore displayed in figure 5.30 under normal conditions (left panels) and worst case conditions (right panels).

To ease the interpretation of the intrinsic firing pattern, a specific display type has been chosen and was kept stable throughout this section. It shows the number of coincident spikes at each point in time (time resolution $10\mu s$) for each of the model stages (i.e LSO stage - blue panel, DNLL stage - green panel and IC stage - magenta panel). Herein, the right model (i.e the right LSO) is displayed in the upper part of each panel and the corresponding left model (i.e. the left LSO) can be directly compared by displaying it in the lower part of each panel. The left column of panels always represents the normal environment and the right column exhibits the model reactions in the worst case environment. Since this method of display has been kept stable throughout all of the real word experiments, the models reactions become comparable between the left and right hemisphere, the different environments and the different signal types.

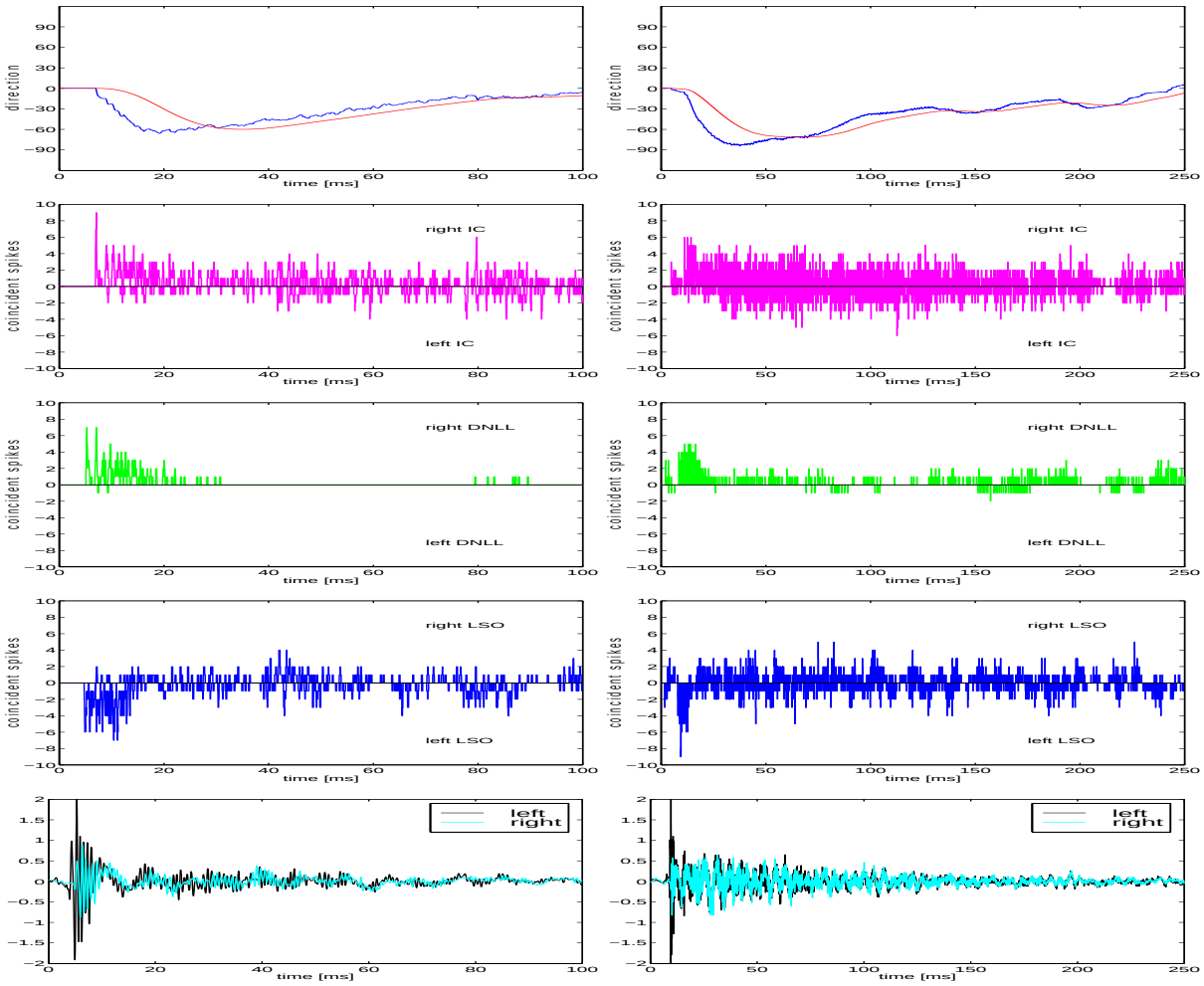

Figure 5.30: model response to a natural click from 60 degrees left recorded under real world conditions. **left panels** *normal conditions* **right panels** *worst case conditions* **bottom to top** *left (cyan) and right (black) stimuli; LSO response; DNLL response; IC response; blue-directional sensor, red-motor output*

The waveform of the two recorded ear stimuli at the bottom of figure 5.30 enhance that during the click under both conditions, the left (black) stimulus exceeds the right (cyan) stimulus intensity. However, during the tail caused by the echoes only, the right intensity sometimes exceeds the one of the left signal. Clearly, there is an echo arriving from the right. (Please note the different time scale of the left and right column, enhancing the far longer echoic tail in the worst case environment.)

Consequently, the firing intensity in the left LSO (lower part of the blue panels) initially exceeds the one of the right LSO (upper part of the blue panels), but as the echoes arrive, conditions sometimes become reversed. Without echo suppression, the output would move towards the positions of the different echoes as intensities switch back and forth.

Why this does not happen is displayed in the green DNLL panels. During the initial part, the right DNLL fires due to the contralateral excitation it receives from the left LSO. Without persistent inhibition, the left DNLL would do the same in case of an echo from the right. However, it does not react at all under normal conditions (left green panel) and only by a few late spikes under worst case conditions (right green panel). The reason is the persistent inhibition of the left DNLL, caused by the strong initial firing of the right DNLL and the left LSO - which have pushed it towards hyperpolarization and it has not yet recovered when the echo arrives.

The different firing behavior of the DNLL in turn influence the IC cells shown in the magenta panels of figure 5.30. Since they receive inhibitory inputs from the contralateral DNLL, only the right IC fires during the initial part of the signal. As can be well observed in the left magenta panel, the firing of the left IC is initially strongly suppressed and only returns along the echoic tail to normal firing behavior. Meanwhile, the right IC continues to fire all the way to the end without suppression. Caused by the DNLL, the difference in firing intensity between the left and right IC model remains clearly in favor of the right IC until the signals amplitude has vanished.

Consequently, the directional sensor as well as the motor signal (shown in the top panel) are not influenced by any echo direction and only sense the sound source location. It can be seen that in case of short clicks under normal conditions (upper left panel), the model can correctly sense the sound source location and ignore echoes arriving after the signal has ended.

Comparing the signals and model reactions between the left (normal) and right (worst case) sides of figure 5.30 it becomes clear that not only the echoic tail extends under worst case conditions but also the right DNLL fires much longer, causing the directional output of the model to return much more slowly to the zero position as in the normal environment (please note the different time scales).

The natural click in the worst case environment obviously generated a higher and longer lasting spike intensity (coincident spikes per time step - summed over all frequency channels) than the same signal under normal conditions. However, the principle remains the same. The strong excitation of the right DNLL (green panel) at the beginning and during the following sections of excitation causes the left DNLL to stay silent for nearly the entire course of the sonic event and the right IC never becomes inhibited.

For technical systems, based on pure wave form evaluations the correct position of the sound source in the worst case environment would only be detectable during the first 50 ms - later

on their output would move away from the source position towards the different echo locations. The right panels of figure 5.30 however, clarifies that echo suppression in this biological inspired model works even under worst case conditions and enables the system to sense the sound source location not only correctly during the first 50 ms, but it also keeps fairly stable during the tailing part of the signal. Fairly stable in this case means it does not change hemisphere at all and jitter does not exceed ±10 degrees.

Since figure 5.30 displayed the model behavior only in case of a sound source positioned 60 degrees left of the Teddy, figure 5.31 now shows the directional output for all 7 sounds source positions varying from right 90 to left 90 degrees. The left panel of that figure displays the motor output after presenting the click in the normal environment and the right panel does the same for the worst case environment.

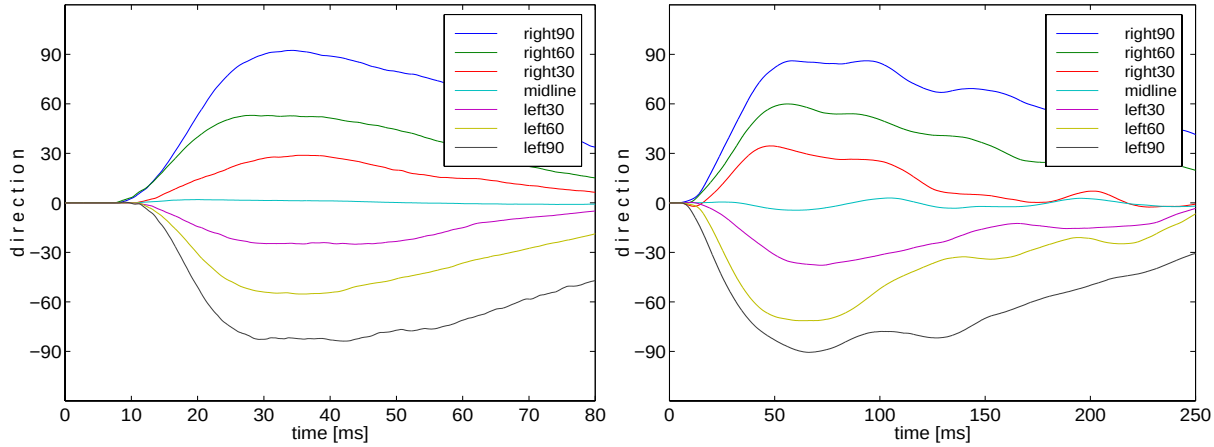

Figure 5.31: Model Output in response to a natural click of a single sound source recorded under normal and worst case conditions. **left panel** *normal environment* **right panel** *worst case environment*

As expected, under normal conditions, the short click is fairly well localized if the Teddy turned left (right position of the sound source - upper part of figure 5.31-left) as well as if he turned right (left position of the sound source - lower part of figure 5.31-left).

As can be observed in the right panel of that figure, this holds for the worst case conditions as well. Although, the tail is much longer and the information therein is only generated by echoes, the motor output returns to zero without changing hemispheres. Of course, it's course is not as smooth as in the left (normal) panel, but the accuracy of detection is not the task , when the signal is already gone. The task here is, not to get disturbed by the strong echoes and this clearly works well.

Overall, the model proves capable of detecting the horizontal direction of a short acoustic event under normal as well as worst case echoic conditions.

In case of such click like acoustic events, most of the echo arrives after the signal has ended. However, clicks are rather seldom signal types in real life. A more interesting question is therefore, whether the system can sense the correct location of continuing sound sources, when the echo directly overlaps with the signal. This has been tested in the next subsection.

5.3.2 Position of a continuing voice

During the second set of experiments, a 100 ms continuing voice signal has been played by the sound source and the resulting stimuli were recorded and presented to the model in the same way as before. Figure 5.32 displays the waveforms of the recorded stimuli.

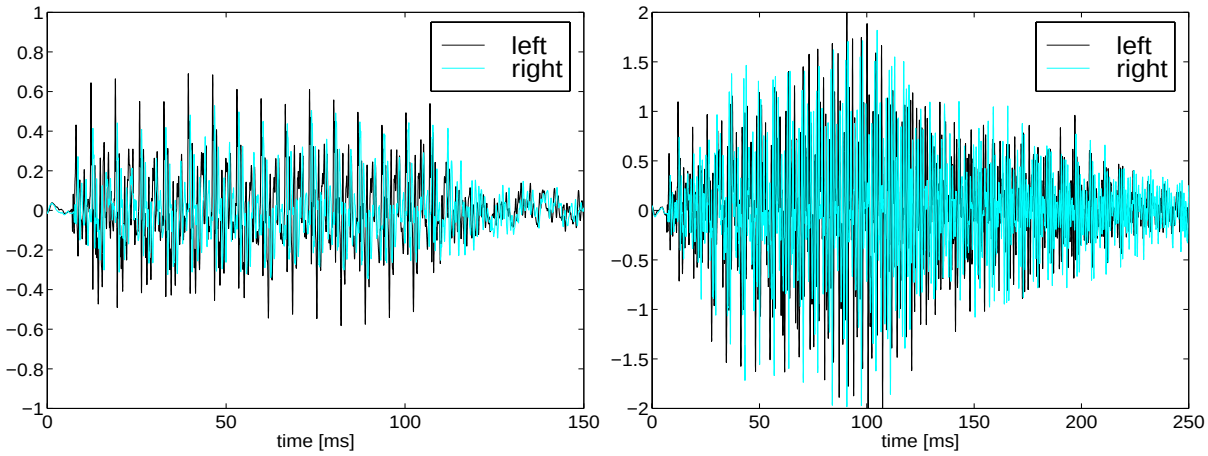

Figure 5.32: Recorded waveforms of a continuing voice arriving from 60 degrees left under normal (left panel) and worst case (right panel) conditions

The left panel shows that under normal conditions, the signal arrives fairly stable at the two ears with some high frequency components added. It is only after the signals end, that the overlapping echoes result in a change of direction and the intensity of the right (cyan) channel exceeds the one of the left (black) channel for about 20 ms.

This is different in the right panel, visualizing the waveforms recorded under worst case conditions. Here, the whole acoustic event not only lasts about 100 ms longer, it is also strongly amplified with increasing amplitude during the signal part (10-110 ms). At several points in time, during and after the signal (i.e. at 50, 120, 160, >200 ms), the echoes cause the intensity relationship between right (cyan) and left (black) channel to change in favor of the cyan one, indicating an echo arriving from the right hemisphere.

Due to the echo suppressing capabilities of the model, both signals are perceived in much the same way by the motor output of the model. It's reaction as well as the firing pattern within the model are shown in figure 5.33. Please note that the right panels display a time frame of 250 ms while the left panels contain only 150 ms.

From the firing pattern of the LSO (blue panels), it can been seen that the increased intensities under worst case conditions (right panel) not only cause higher activations in the LSO but also prevent the LSO from a correct sensation of sound source location, immediately after the signals onset. While the intensity difference under normal conditions (left blue panel) is clearly in favor of the left LSO, there is no clear message to be obtained from the LSO under the worst case conditions (right blue panel) - in fact the system would fail to respond correctly based on the LSO activation.

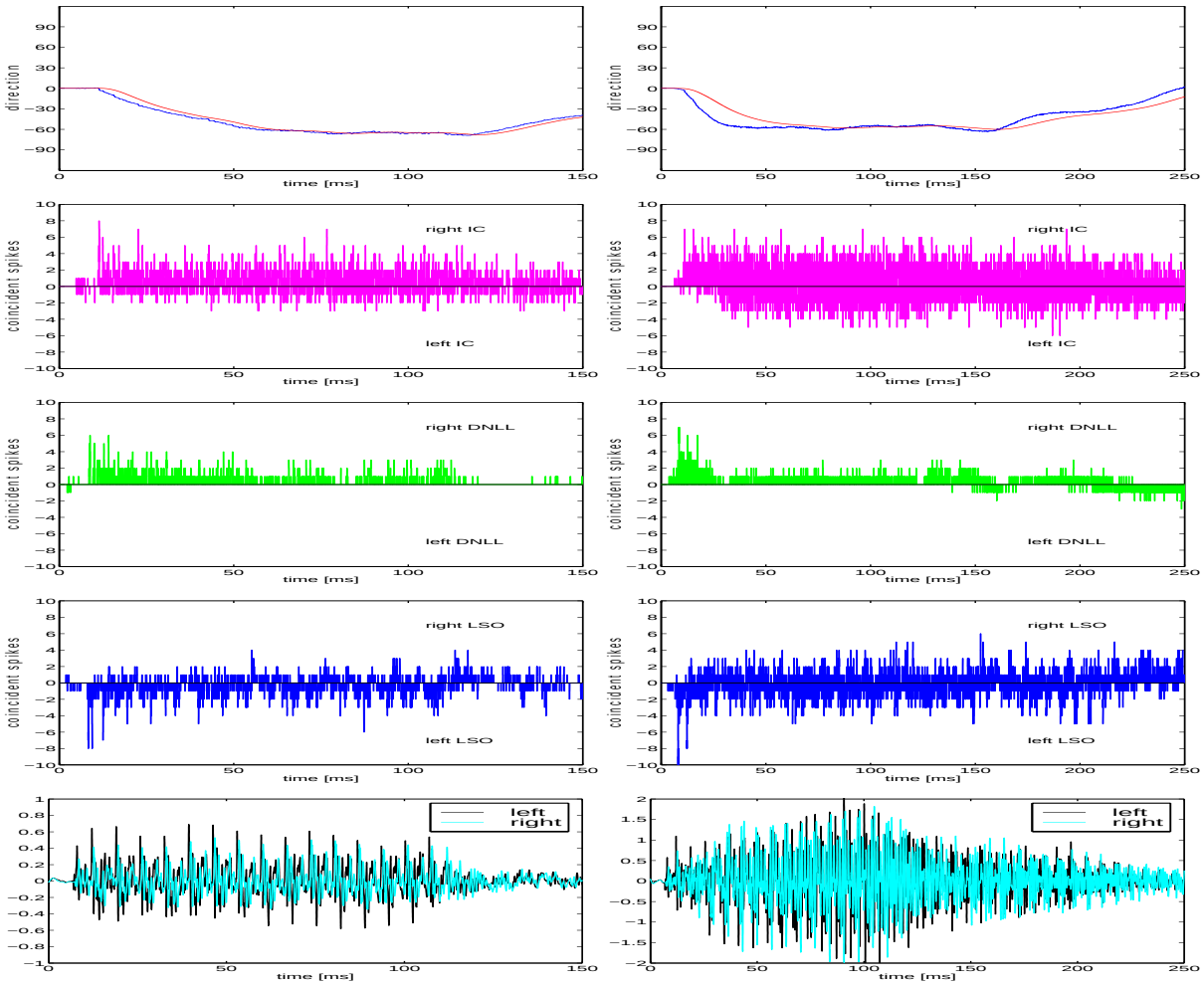

Figure 5.33: Stimuli, intrinsic spike pattern and model output in response to a continuous tone of a single sound source recorded under normal real world conditions. **left panels** *normal conditions* **right panels** *worst case conditions*

However, the behavior of the DNLL (green panels) during the first 150 ms is fairly similar under both conditions. This is due to the mentioned hyperpolarization effect, which prevents the left DNLL form firing after it has become initially inhibited by the signals onset. Only during the echoic tail (150-250 ms) the left DNLL can gain power under worst case conditions and starts to fire. However, since the signals amplitude during that part is low, the resulting movement of the motor output is only towards zero, as it would anyway since the signal has ended.

Based on the firing of the right DNLL, only the left IC becomes inhibited, as displayed at the magenta panels. And although the entire IC is activated under worst case conditions much higher than under normal conditions, the difference between left and right IC activation is quite similar and remains stable throughout the signal.

Consequently, the sensed sound source location, shown in the upper panels of figure 5.33, is identical under both conditions and does not change as long as the signal lasts. This is an important observation, since it proves that despite strong echoes, this architecture can not only sense the correct location of short clicks and during the signal onset, it also features a stable behavior during a continuing voice signal under normal and worst case conditions.

Since figure 5.33 contains only a single sound source location, at 60 degrees left of the Teddy, the capabilities of the model were also tested for all of the other directions. The obtained motor outputs are displayed in figure 5.34.

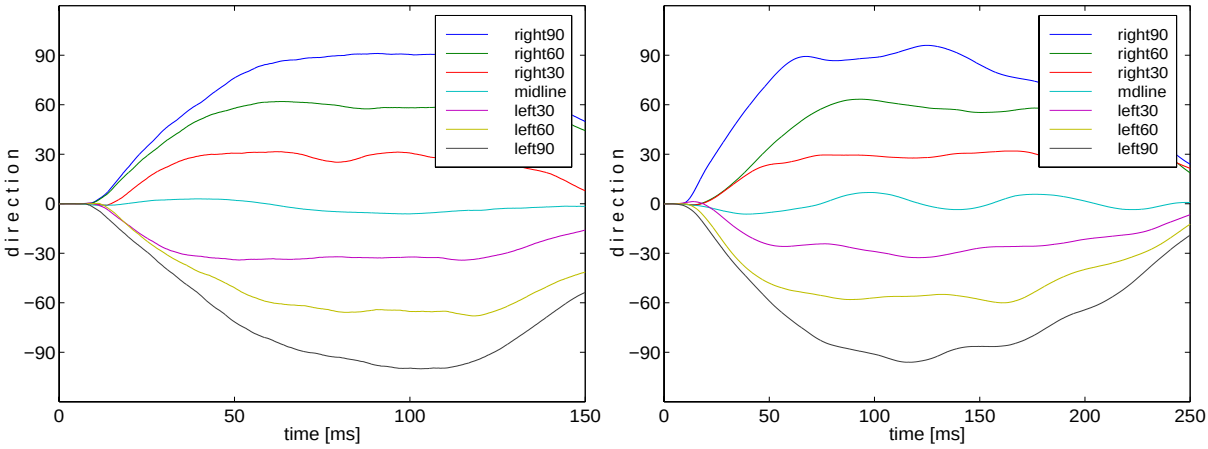

Figure 5.34: Motor Output in response to a continuous tone of a single sound source presented under normal (left panel) and worst case (right panel) conditions from 7 different directions with regard to the acoustic base

There are two facts to be observed from figure 5.34. First, the location of the sound source is sensed correctly under both conditions and not lost, even during the echoic tail of the signal. Second, it is sensed more slowly under the normal condition displayed in the left panel, than under worst case conditions in the right panel. The second observation is partly caused by the different time scales of both panels but is also real and originates from the slower and weaker onset of the perceived acoustic event under normal than under worst case conditions.

Furthermore, during the long signal tail in the right panel, there is some movement of the directional sensor indicating the fact that the stimuli is now entirely composed of echoes arriving from all kinds of directions and therefore slightly influencing the directional sensation.

However, this set of experiments clearly exhibits and explains the capability of the model, to correctly sense the position of continuing sound sources in the horizontal plane under normal as well as worst case real world conditions.

5.3.3 Position of a speaker

Finally, the third and most complex type of signals has been employed to test the ability of the model to localize dynamic speech signals. Here, the set of experiments was repeated using a male voice speaking the words “Here I am“ - displayed by the loudspeaker. The recorded waveforms under normal and worst case conditions are displayed in figure 5.35.

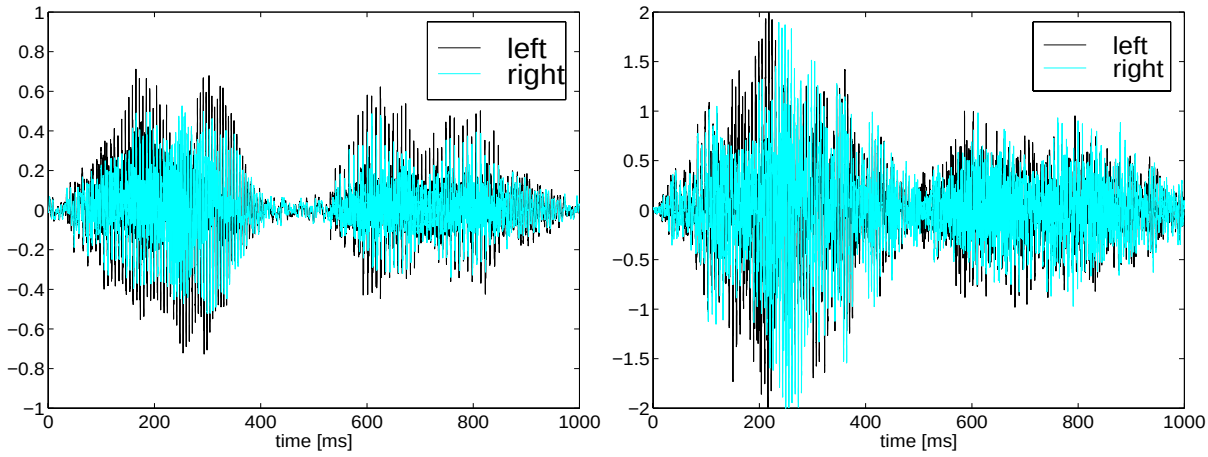

Figure 5.35: Recorded waveforms resulting from a continuing voice from 60 degrees left under normal (left panel) and worst case (right panel) conditions

These waveforms display the strong amplification and disturbance by the echoes in the worst case environment. It is also to observe that these strong echoes more or less fill the gap between the words “Here“ and “I am“ originally well to perceive in the signal at figure 5.27 and still observable under the normal conditions in the left panel of figure 5.35. The question is again, if the model is able to sense the correct sound source location under these much more complex conditions.

Figure 5.36 displays the internal reactions of the model in response to the speech signal presented under normal conditions (left panels) and worst case conditions (right panels).

Again, the firing pattern under normal conditions (left panels) clearly indicate that already the LSO senses a signal from the left during the first as well as the second part of this short sentence. The gap in between is clearly perceived and the output sensor (shown at the top) returns to zero position, since activation vanishes during that period. The left DNLL is completely inhibited, during both parts of the signal, and the right DNLL constantly fires. Consequently, only the left IC is inhibited, while the right one responds directly to the stimuli. The speakers position is sensed correctly and not much jitter is seen in the output element.

Under the worst case conditions (right panels), this task is much harder. As can be seen, the LSO models respond nearly equally during both parts of the signal, changing their preference several times between right and left positions. Furthermore, the gap between the words is nearly gone and filled with echoes from all over the place. However, the DNLL even under these extreme conditions stays fairly stable. As can been seen in the green (right) panel, it is the right DNLL who clearly succeeds its left counterpart. Even if the reaction is not as strong as under normal conditions it is by far sufficient to generate the expected intensity difference in the firing of the

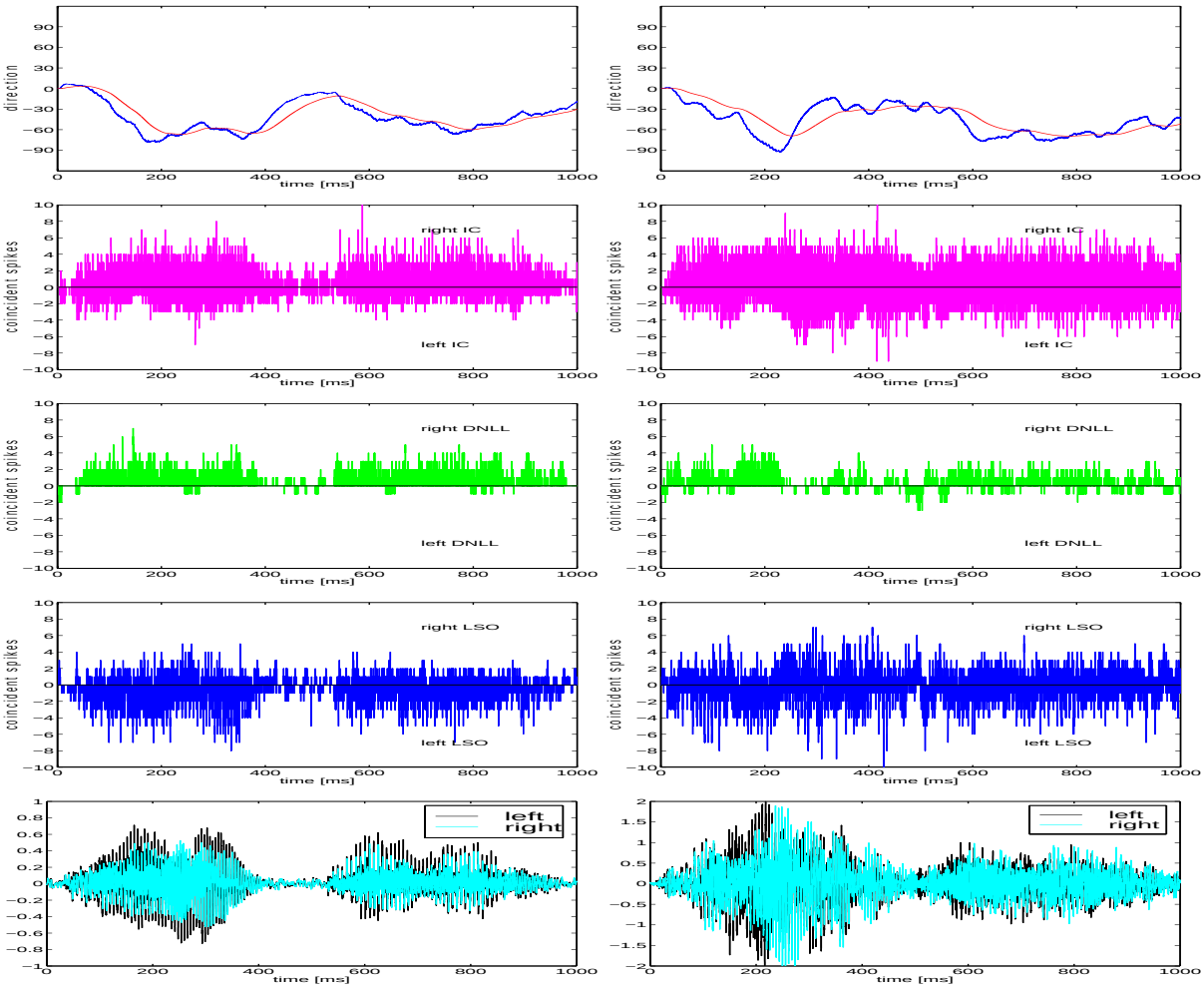

Figure 5.36: Stimuli, intrinsic spike pattern and model output in response to a speech signal from a single sound source 60 degrees left of the listener - recorded under normal (left panels) and worst case (right panels) conditions.

right and left IC and the appropriate output of the directional sensor. Although, the directional output raises very fast (and somewhat overshoots) the smoothed motor output moves to the correct position. Since there is no real gap in the recorded waveforms it does not return to zero in between, but it still indicates the existence of two words by moving significantly toward the zero line. Overall, the system can even cope with dynamic signals under worst case conditions.

If this holds for all directions was tested by turning the Teddy in the previously described manner and obtaining the motor signals displayed in figure 5.37.

The left panel of that figure displays a very convincing result with regard to the models capability to localize the position of a speaker in the horizontal plane under normal conditions. At this point one should remember that this is not an anechoic or free field condition and echoes are present here too. However the location of the speaker is sensed correctly during both parts of the sentence and the motor output could be used to turn towards a speaker as well as towards a continuing voice, as shown in the previous subsection.

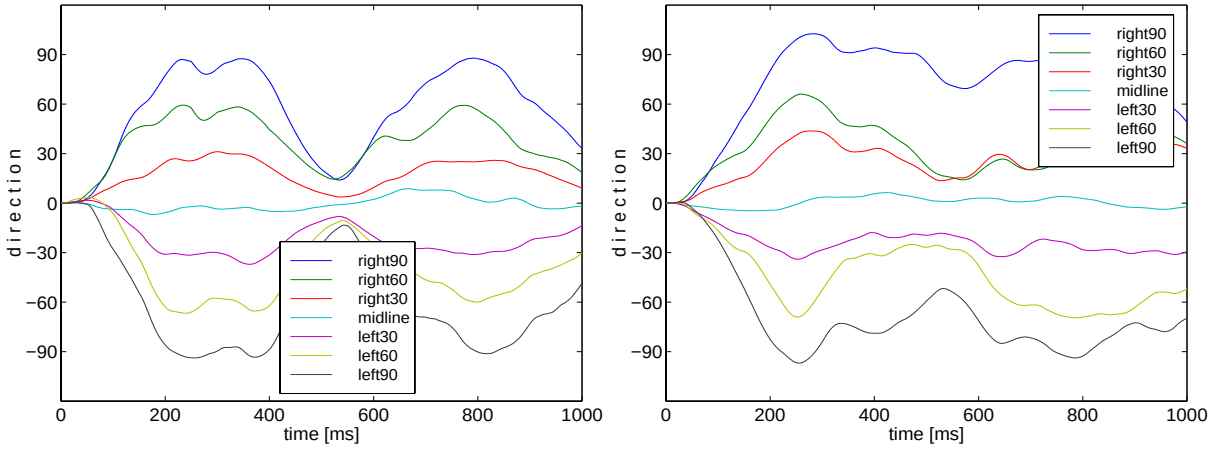

Figure 5.37: Motor Output in response to a speech signal from a single sound source presented under normal (left panel) and worst case (right panel) conditions from 7 different directions with regard to the acoustic base

Looking at the right panel of figure 5.37 this holds in principle for the worst case conditions as well. However, the accuracy now becomes limited as the sensor overshoots slightly in case of a signal arriving from right 90 degrees and failed to sense the second part correctly if the signal arrived from right 60 degrees. However, these limitations might be caused by principle problems and can well be due to incorrect conditions during recording (with regard to the 90 degrees) and/or a recording problem in case of the right 60 experiment.

However, being aware of these minor failures it seems fair to state that the model is capable of detecting the correct position of a speaker with acceptable accuracy even under worst case conditions and can be used to detect the position of a speaker with regard to a technical system.

5.3.4 Tracking of a moving continuing voice

The localization of sound sources, tested in the first three sets of real world experiments, is a common task in real live, but is not the only one. Despite the initial localization, the tracking of sound sources marks an important strength of the auditory system. The aim of this subsection, therefore, was to test the ability of the model architecture to track a continuing sound source as it moves its position in the horizontal plane.

In contrast to the localization experiments now the Teddy was kept stable but the speaker, playing a continuing voice, was moved manually in a nearly continuous manner from the left to the right and back again. The experiment was then repeated the other way around. The recorded waveforms for the left-right-left (LRL) experiment are shown in figure 5.38.

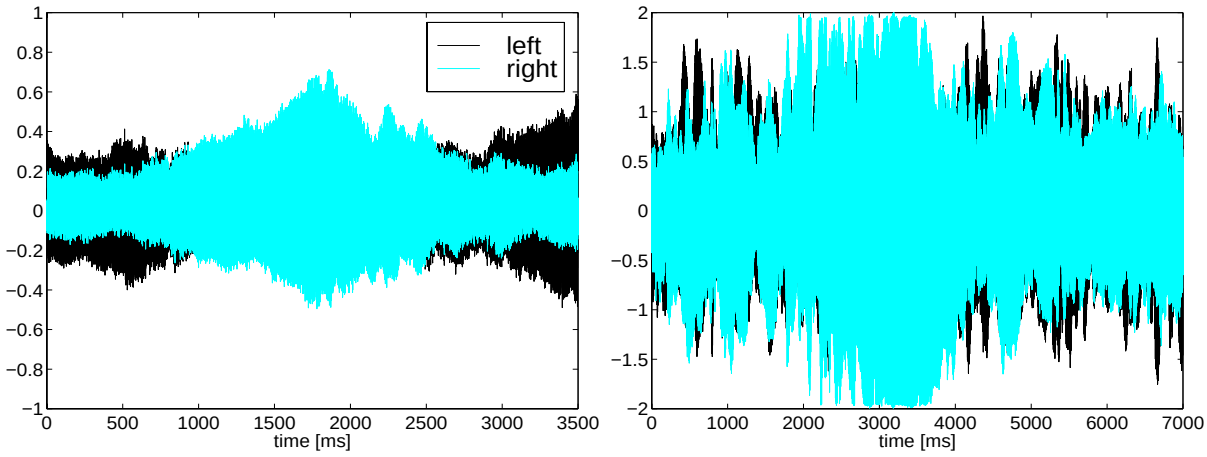

Figure 5.38: Recorded waveforms resulting from a continuing sound source moving from left to right and back to left position under normal (left panel) and worst case (right panel) conditions

As can been seen in the left panel, within only 3.5 seconds the sound source moved from the the left position (black left stimulus is louder) acorss a right position during the middle part of the waveform (cyan right stimulus more intense) back to the left side. Since this has been carried out by simply walking the speaker around the Teddy, the distance to the two ears changed, and obviously shortened - causing an increase in absolute intensity. Since interaural intensity differences are quite clearly to perceive, it is expected that the model will be able to track the sound source under these conditions.

The bathroom recording, displayed in the right panel, looks much less convincing. Obviously, the strong echoes triple the overall intensity and diminish the visually perceptible IID to a great extend. It should be noticed that the moving speed is slowed down here by half - so the experiments lasts 7 instead of 3.5 s. Overall, the detailed structure of the signal has become quite complex and disturbed (IID's switch between left and right preferences) and it is only the general time course, indicating a higher intensity of the black channel, at the beginning and the end. The question here is, whether the system, acting on a μs time scale, will be able to extract this general time course and track the moving sound source.

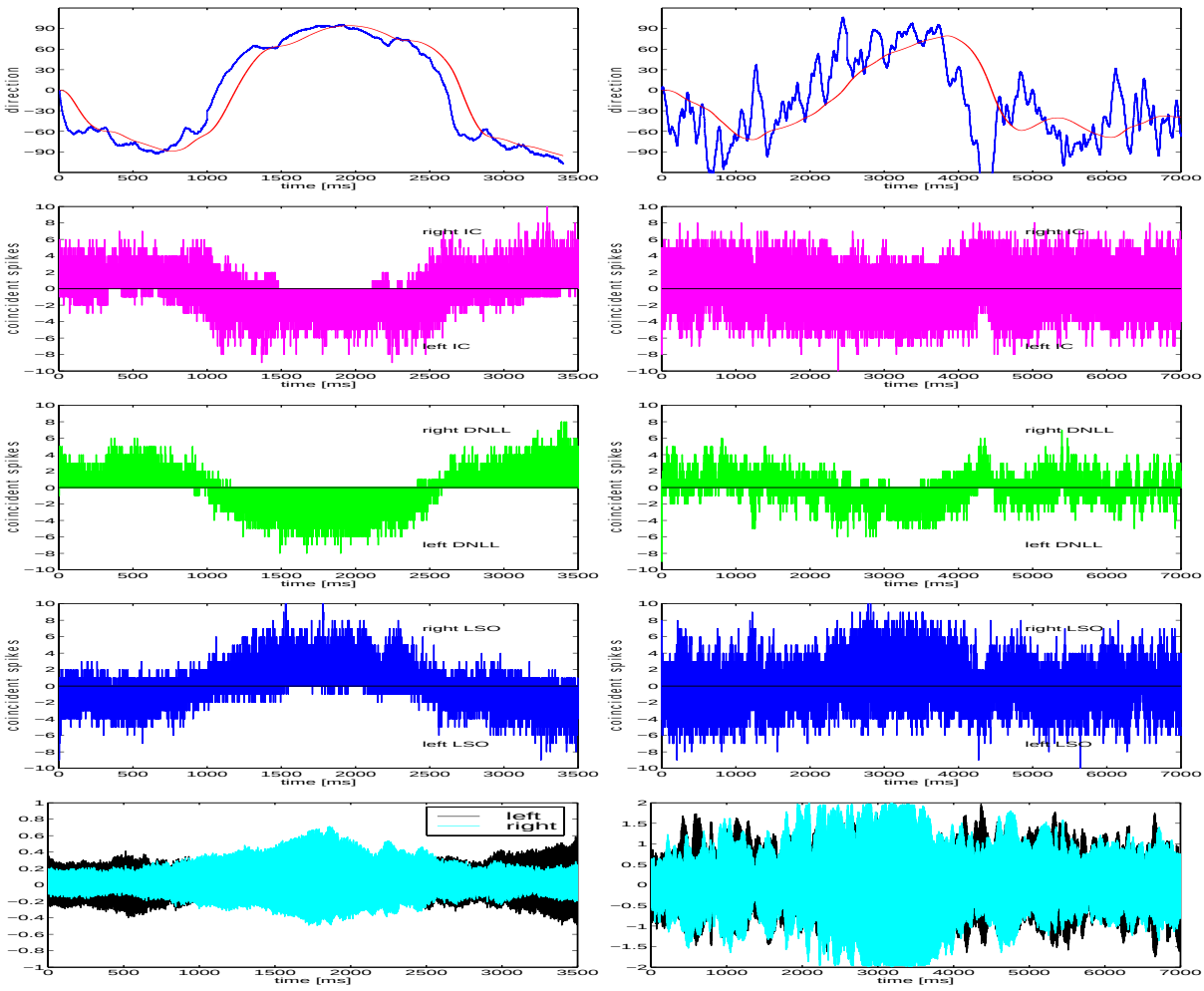

Figure 5.39: Stimuli, intrinsic spike pattern and model output in response to a moving continuous voice signal from a single sound source recorded under real world conditions. **left panels** *normal conditions;* **right panels** *worst case conditions*

The now well known display of the intrinsic model responses is very clear for the moving sound source under normal conditions, as displayed in the left panels of figure 5.39. Since the intensity difference is already sensed correctly at the LSO level, the DNLL exhibits a very clear response, causing a smooth and stable reaction at the IC and the motor output. Although, this result was expected, it proves an important feature of the developed architecture - the ability to track sound sources under normal real world conditions - and makes it quite suitable for use in technical applications.

And it clearly works under worst case conditions as well - right panels of figure 5.39 - even if the job is much harder. As expected from the waveform, the LSO model extracts the IID only during the first left and right positions (up to about 4000 ms). During the last part no specific directional information is perceived by the LSO. And the DNLL encounters problems as well, but manages it by an overall higher right than left firing intensity also during the last part of the signal. Let's have a more detailed look at the signal subset between 4000 and 5000 ms. At the 4200 ms point, the sound source position has turned back to the left hemisphere

first time, hence the left (black) intensity exceeds the right (cyan) intensity first time, after a period of opposite conditions. It takes up to the 4400 point untill the DNLL rewards this change of conditions and clearly indicates a left position. However, about 200 ms later a strong echo arrives from the right, turning physical conditions again upside down. But since the DNLL features persistent inhibition, it does not reward this with the strong change of intensities as the LSO does. Although, the right DNLL cannot fully suppress left response, this is at least diminished, resulting in a fairly equal firing intensity of both DNLL hemispheres and causes a nearly identical firing pattern at the IC level. This in turn creates only a dip in the delayed reaction of the motor output. It can bee seen that this very strong echo has not been suppressed completely but its influence has become diminished so that the overall sensation is not disrupted.

This example shows that the tracking of sound sources under these worst case condition is not a mission impossible for the model and although the directional output exhibits a high level of jitter, the motor output extracts the time course of the sound source movement fairly correctly and smooth. This fits very well with the human perception, when listening to the recorded signals over a head-set. The first impression of the author was: - there is a whole mess of sound from all over creating a voluminant but unspecific impression. However, after the signal ended, there was a intrinsic perception of some movement from left to right - obviously at a much higher level of abstraction as during listening to the signals recorded under normal conditions. It was not expected, but a positive surprise that the model generated the same type of sensational output - unspecific in detail, but correct on the general level of movement.

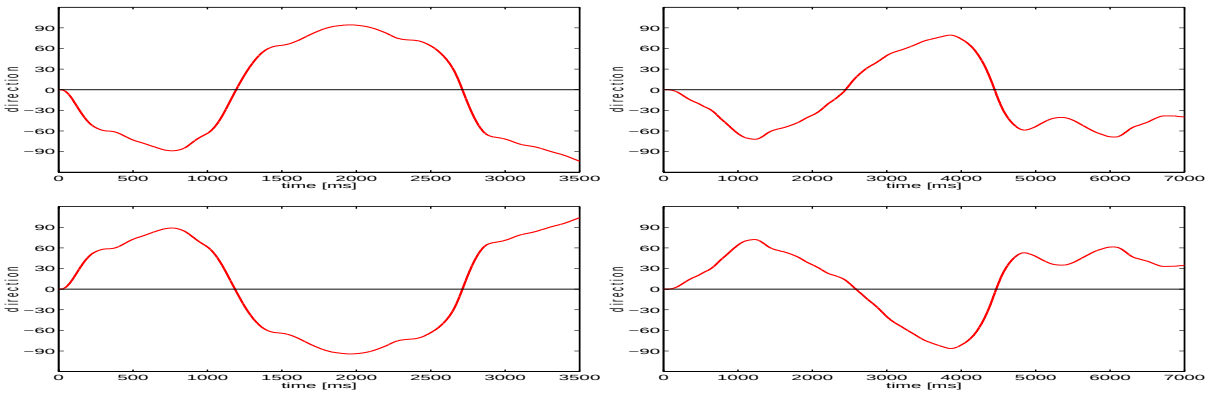

Figure 5.40: Motor output in response to a continuing voice moving left-right-left (upper panels) and right-left-right (lower panels) under normal (left panels) and worst case conditions (right panels)

Looking at the general level motor output only, the picture becomes more clear and proves the usability of the model to track moving sound sources under normal conditions (Left panels of figure 5.40) and worst case conditions (right panels of figure 5.40). The lower two panels are not just mirrored displays - they indeed exhibit the results of separate experiments, when the sound source moved along the right-left-right track. The fact, that the LRL sensation under worst case condition reaches only 75 to 80 degeeres instead of the expected 90 degrees can well be caused by the experimental conditions, since the lower panel shows that also under worst case conditions the 90 degrees can be sensed correctly.

Overall it can be stated that the bio-inspired model is capable of tracking moving continuing sound sources well under normal and fairly sufficient under worst case conditions.

5.3.5 Tracking of a moving speaker

The logical question arriving from the experiments discussed above is, whether the model can also track dynamic sound sources like speakers, as they move their positions in the horizontal plane. Figure 5.41 displays the recordings generated by such a moving speaker moving from the right to the left within only one second - obviously a running man.

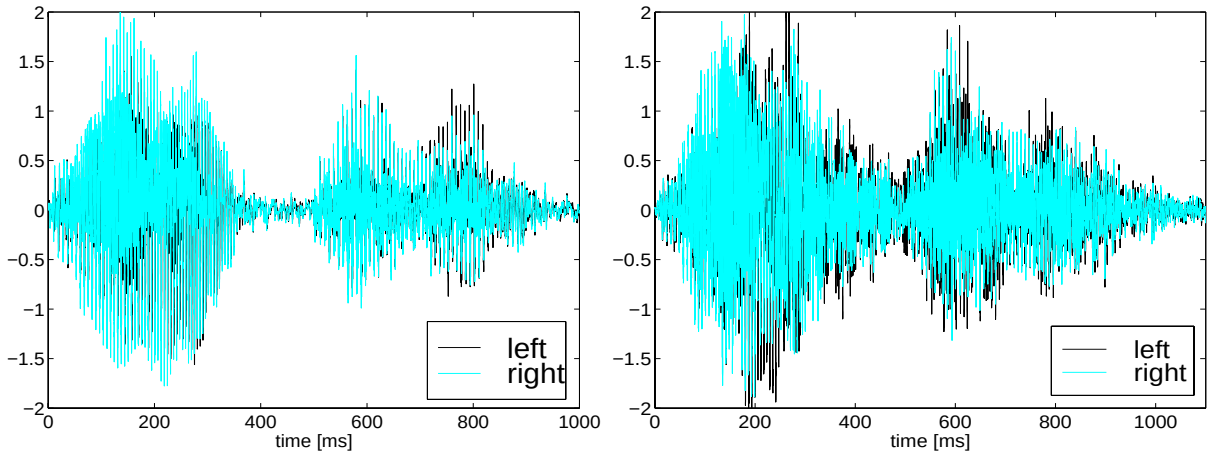

Figure 5.41: Recorded waveforms resulting from a speaker moving from the right to the left under normal (left panel) and worst case (right panel) conditions

Within the waveforms, displayed in figure 5.41 the three syllables of the spoken sentence “Here I am“ can be clearly distinguished even if the gap between the first part “Here“ and the second part “I am“ is much less clear under the worst case conditions, displayed in the right panel. In general the experimenter tried, to speak the first syllable “Here“ at a position right from the midline, the second “I“ about at midline position and the third one “am“ at a left position.

What happened within the model is again displayed in figure 5.42 for the normal conditions on the left and the worst case conditions on the right.

In both cases it is clear that LSO as well as DNLL sense the first Syllable from the right. There are some echoes visible between 200 and 300 ms at the spiking pattern of the left LSO, which become totally suppressed at the DNLL level under normal conditions and mainly suppressed under worst case conditions. During the signals gap (between 300 and 500 ms) the directional sensor moves towards zero under normal conditions, since there is only few activation at all. In the worst case environment the echoes cause higher amplitudes which already indicate the ongoing movement of the sound toward the left and cause the sensor to move towards the left. However, this might also be caused by a faster movement during the experiment, since the moving speed was not controlled and depended on the specific situation.

This might also have caused the fact that under normal conditions the second syllable is correctly sensed at the midline, while the left panels of figure 5.42 show that the intensity difference shifts towards the left side indicating a speaker position at the left hemisphere. In this case it remains unclear, if the observed effect is caused by echoes, or the faster movement and the experiment will be further investigated in the future.

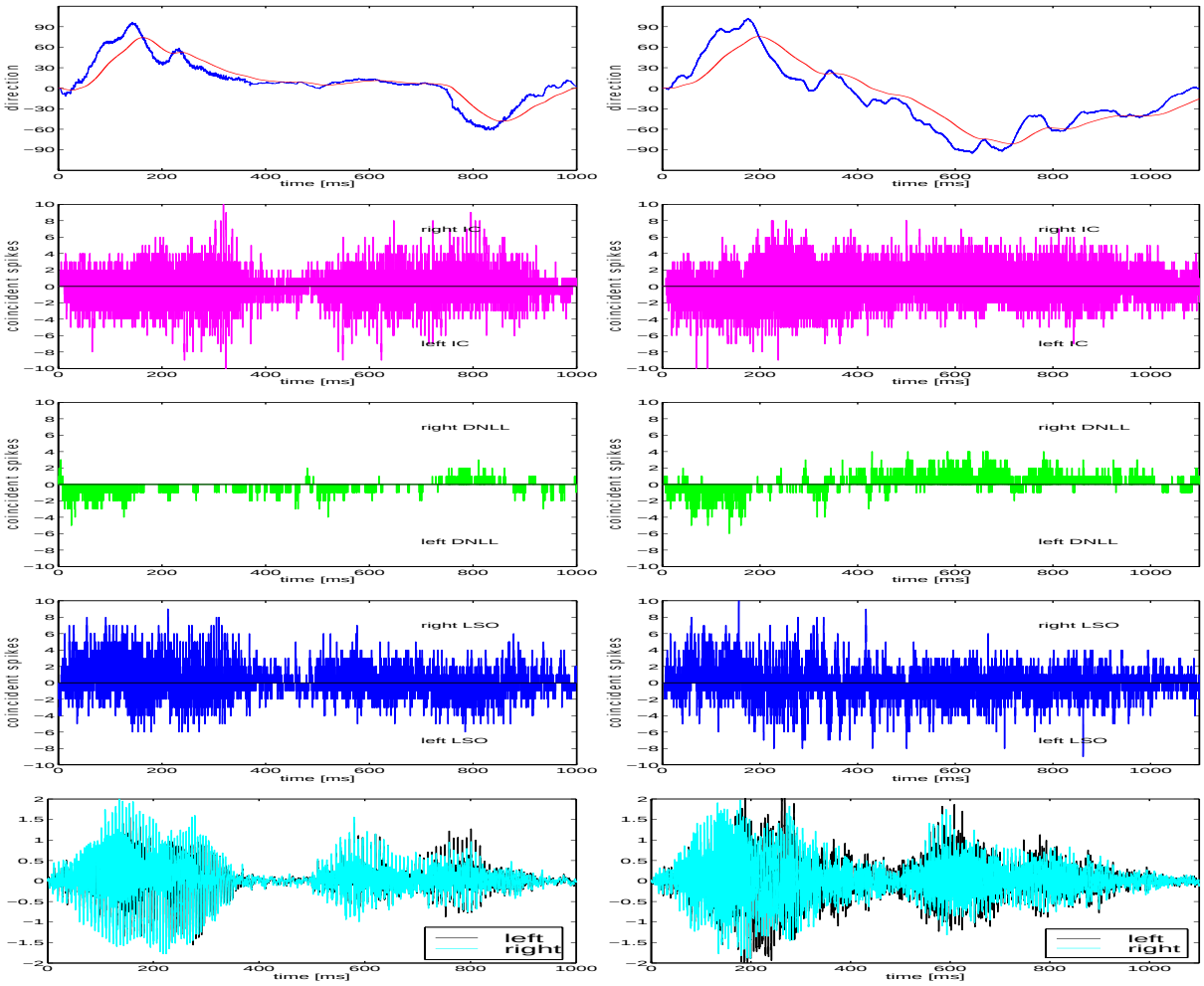

Figure 5.42: Stimuli, intrinsic spike pattern and model output in response to a moving speaker recorded at real world conditions. **left panels** *normal conditions;* **right panels** *worst case conditions*

However, the final part of the sentence is clearly sensed at the right hemisphere in both environments, as could be expected by visual evaluation of the stimuli displayed in the bottom panels of figure 5.42. The tracking of that very fast moving speaker has therefore be seen as successful, since the system under both conditions clearly indicates a sound source moving from the right to the left hemisphere.

Finally, this experiment was repeated for the opposite direction in both environments. Since this was done by separate experiments, moving speed may have been different again. An overview of the motor outputs obtained from all four experiments is displayed in figure 5.43.

Two statements can be derived from that overview. First, the tracking of moving speakers works correctly under normal real world conditions in both directions - as can be seen in the left panels of figure 5.42. Second, the tracking of moving speakers in a worst case environment works as well in both directions, but the sensation of the speakers position is influenced by echoes during speech gaps with no active signal.

Based on these results it can be stated that the model is capable of tracking moving dynamic and speech signals best, if they contain only small or no gaps. However, under the normal conditions a technical system would have to cope with tracking works very stable even during gaps.

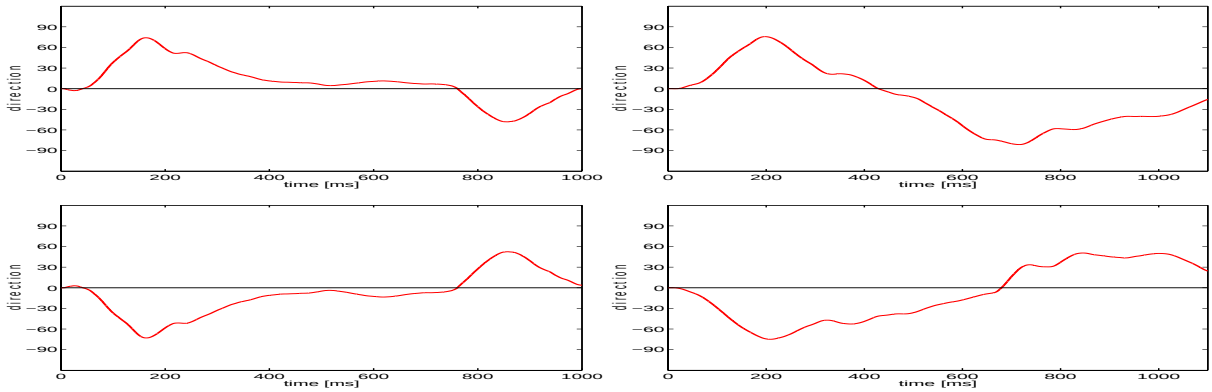

Figure 5.43: Motor output in response to a speaker moving from the right to the left (upper panels) and from the left to the right (lower panels) under normal (left panels) and worst case (right panels) conditions

In summary of that entire section on real world experiments, the bio-inspired model proved it's ability to localize and track sound sources in the horizontal plane, regardless of their acoustical structure. It delivers stable results under normal conditions, i.e. in office rooms, floors and large halls. Under the worst case conditions of small rooms with extremely intense echoes it looses some of it's accuracy, but is still functional.

Summary

More than 100 experiments have been performed and discussed to test the 3 general expectations to the developed neural architecture for echo suppression during sound source localization. Their results support the following statements:

- Persistent inhibition is present in the Gerbils DNLL and a possible general feature of mammals.
- The spike pattern of living single cells in the gerbils DNLL can be duplicated by the models of single cells in the developed architecture.
- The model is capable to duplicate 6 major psycho-acoustical features and help to explain their root causes.
- The model is capable to localize and track static and dynamic sound sources under real world conditions and delivers stable output signals, usable to control technical systems.

These general statements have been tested under specifically designed as well as selected real world conditions. However, even the more than hundred experiments are not sufficient to prove general compliance to the biological system. It is therefore intended to extend the experimental investigation of the model's behavior, as suggested at several points during this chapter.

On the other hand, the obtained results have been convincing and justify the claim made in the introduction of this thesis that detailed models of biological structures can contribute to the understanding of biology as well as fulfill practical tasks under real world conditions in a robust and stable manner.

Chapter 6

Discussion

As shown in the previous chapter, the proposed model architecture proved capable of duplicating physiologically recorded spike pattern as well as psychoacoustic phenomena and is able to suppress echoes of real world signals. However, since echo suppression is a common problem to technical systems, a variety of different models have been designed to make it accessible to man made machines. To prove the novelty and specific value of the proposed architecture, some of them will be introduced and comparably discussed during the first two sections of this chapter.

Here, the first section deals with *neural based echo suppression* - covering those models, attempting to duplicate the functions of auditory nuclei or employing artificial neural networks to suppress acoustical echoes. The second section introduces some modeling principles for *technical echo suppression* - employing algorithmic time or frequency domain methods, to separate and/or suppress acoustical echoes.

The final section of this chapter contains a brief discussion of the AVLSI implementation approach pursued during this study, aiming to provide building blocks for Spike Interaction Models (SIM) in mixed analog-digital silicon technology. It provides a short overview of the current state of AVLSI implementation.

6.1 Neural based Echo Suppression

As mentioned in the introduction of this chapter, this section attempts to compare the developed model for echo suppression during sound source localization to other neural models for echo suppression and sound source localization. Hereby, the first subsection deals with those models that attempt to duplicate the functions of biological auditory structures and nuclei. By contrast the second subsection focuses on models involving artificial neural networks to localize sound sources and to suppress acoustical echoes.

6.1.1 Models of auditory nuclei

Out of the huge number of models for sound source localization and echo suppression, only a few research groups approach the problem based on models of the auditory brainstem. Here, the modeled structures and nuclei vary from auditory nerve fibers up to the cells of the inferior colliculus. Since some of the models share essential features with the proposed architecture, they will be shortly introduced and compared at this point. However so far, no publication has been found to share the principle, level of detail and methods of the proposed model architecture, suggesting a truly novel approach.

Blum and Reed Model of DNLL cells
Obviously the most similar model to the one proposed in this thesis has been published by Jacob J. Blum and M.Reed at the Duke University, Durham USA [RB99][BR00]. The most prominent similarities are the following:

- it models the cell response of single DNLL cells to binaural stimuli, based on the hypothesis of Markovitz, Pollak and Yang [MP94][Pol97][YP97]
- it models the inhibitory and excitatory projections between AVCN, MSO, LSO driving the DNLL cell response
- it employs models of single synapses and neurons with internal time delays of 1 ms

This way Blum et al. could reproduce many of the perceived firing pattern of single cells in the DNLL like pauser, temporal inhibition and temporal dynamics of EI and EE cells. However, there are some essential differences making this model less suited to solve the technical task of echo suppression as well as the task of biological relevant modeling than the proposed architecture. These are:

1. It employs a fairly simple cell model, simulating only time continuous firing rates per second, no single spike events, and no dynamic postsynaptic responses.
2. Each auditory structure is represented by just one cell, representing only one frequency channel and therefore loosing the spectral coding of intensities.
3. The contralateral inhibition between left and right DNLL is divided between two different cells within the DNLL and, therefore, the entire model will act very differently under dynamic conditions.
4. The model lacks any ear model and it therefore cannot directly process sound waves - it rather needs to be fed with artificially designed spike rates, and finally,
5. the Blum model has not been designed and employed to perform echo suppression and it lacks any model of the inferior colliculus or higher structures.

Obviously, the models of Blum et al. can suit the needs of physiologists well up to a certain level of detail - although lacking the time course of internal cell potentials - but they cannot bridge the gap between biological relevant modeling and applicability to technical systems - since they were never confronted with real world tasks.

Cai Model of MSO-IC cells

Another model, which now includes the discharge properties of IC cells, was published by Cai and his colleagues at the Boston University in 1998 [CCC98]. This model attempts to simulate the different firing pattern of IC cells but does not include a DNLL model and it is rather concerned with the general localization task within the MSO - IC projection than the suppression of echoes. Similar to the architecture proposed here, it models the internal cell potential by Hodgkin - Huxley type equations and resembles the time course of excitation and inhibition with a high time resolution. For that reason it seems one of the very valuable biological modeling approaches for sound source localization even if it is probably not capable to realize effective echo suppression.

The most significant difference between Cai and the proposed model architecture lies within the fact that Cai exclusively employs interaural time differences (ITD's) and does not rely on interaural intensity differences (IID's) at all. The problem of a purely ITD and therefore low frequency driven attempt has been encountered by several groups, including the author [ZITP97] and arrives from the fact that under highly reverberating conditions, the timely relationship between the left and right channel becomes disturbed and fails to perform immediately after the onset period of natural sounds.

Overall, the models of Cai et al. are well suited to resemble physiological cell functions even in detail, but may fail to perform the entire task of echo suppression under real world conditions, due to the limited subset of auditory nuclei incorporated.

Leibold Model of MSO cells

Another localization model based on single cell responses and synaptic plasticity in the MSO has been published by Leibold [Lei02]. He uses a Spike Response Model (see page 16) to compare the traditional place coding model of Jeffress [Jef48] with a modern rate coding model of McAlpine and Grothe [MP02] [GN00]. As discussed in the first chapter, Spike Response Models (SRM) are very close to the applied Spike Interaction Models (SIM) and therefore, the level of detail in Leibolds model is close to the one of the proposed architecture.

However, there are three major differences between the two models. First, the focus of Leibolds model is on ITD evaluation in the MSO, and the focus in this study is on IID evaluation in the LSO. Second, Leibolds study concentrates on the comparison of principles than the technical applicability and is not concerned with computational or implementation tasks at all. And third, the Leibold model concentrates on the pure localization task and does not deal with echo suppression at all.

Horiuchi Model of LSO cells

By contrast, the model of Horiuchi and Hynna [Hor01] clearly attempts a technical solution by realizing the azimuthal localization mechanism of the Big Brown Bat as spike based VLSI model. Similar to the proposed architecture it uses interaural intensity differences as neural cues for sound source localization and models the responses of LSO cells at the level of cell potentials in order to guide the auditory attention of an autonomous robot. However, it does not include a DNLL or IC model and is also more focused on the general localization task than the suppression of echoes.

While the features of Horiuchi's silicon neuron will be discussed during a separate dedicated section, it shall be mentioned here that the entire model seeks to reassemble the localization

mechanism of bats and therefore operates in the ultrasonic frequency range around 40 kHz. This and the lack of auditory structures above the level of the LSO limit it's applicability to room echoes generated by human voice signals. The model of Horiuchi realizes just another biological motivated principle, exclusively used by bats, to localize objects rather than sound sources.

Although different in principle and not concerned with echo suppression in the sense of this study, this model is a very valuable step towards biological relevant technical modeling. Especially, since VLSI implementation assures a fully parallel and on-line performance and marks a significant step toward successful neuromorphic engineering.

Auditory Nerve Models of Colburn

Already in the 70'th of the last century Colburn et al. started to investigate the firing pattern in the auditory nerve (AN). Several publications concerning the theory of binaural hearing [Col70],[Col73],[CE76],[CL78] not only investigated the localization of sound sources based on timing differences between the left and right AN, but also employed a detailed computational model of the firing pattern within the left and right Auditory Nerve [CE76],[HCC01] to investigate the rate-place code important e.g. for source localization at higher frequencies.

Similar to the proposed model of this thesis, they found both, the spike timing as well as the number of frequency channels to carry significant information about the horizontal azimuth of sound sources. However, echo suppression, in their view, results from adaptive mechanism within the inner ear and auditory nerve and has been modeled, without being in special focus of their studies. While adaptive effects throughout the entire auditory brain stem cannot be denied, there seems to be no sufficient evidence to view them as the general mechanism for echo suppression. To the believe of the author, adaptive mechanisms play a mayor role for adaptive perceptual effects like the Clifton effect (see page 39), but thus cannot realize the entire task of echo suppression during sound source localization.

Although, Colburn et al. only model the very early stages of the auditory system (inner ear and acoustical nerve), their models are essential to understand the principles and importance of auditory coding and support the hypothesis of this thesis that spike time and place in the auditory nerve are essential for understanding and modeling higher auditory functions, like sound source localization and echo suppression.

Cochlear Model of Schwartz

The last model to be discussed during this section is the one of Schwartz et al. [SHP99]. Although, it does not model auditory cells it can be counted as a biological model, since it attempts to realize echo suppression by the help of the gamma filters modeling the inner ear.

The underlying assumption is that the rate of change during the onset part of a sound event determines the precedence effect. Therefore, the adaptive gamma filters, used in this study, enhance the onset part of a speech signals and suppress the ongoing components, this way increasing the initial rate of change during the onset compared to the ongoing part of acoustic signals. Since this study employs complex speech signals, instead of synthetic clicks, it's relevance to this discussion is given by the fact that it seeks to suppress echoes of natural speech signals under real world conditions based on a biological motivated adaptive model of the inner ear.

However, neither the sound source localization task nor the role of higher nuclei in the auditory brain stem are subject to that model and while it successfully suppresses speech components, after the onset period, it might still not be able to correctly sense the direction of speech signals under echoic conditions.

To summarize this subsection it can be stated that only few models use the biological auditory brainstem as guideline to model echo suppression during sound source localization. The only one concerned with the role of the DNLL during this task is the one of Blum et.al which does not show (and seek) technical applicability. Based on the available information it therefore seems valid, to see the proposed model of this study as the first to combine detailed modeling of auditory nuclei up to the DNLL-IC level with technical applicability of echo suppression to mobile robots and other technical systems. The significance of the sound localization task under reverberating conditions has just lately been confirmed by the european presentation of the HONDA humaniod robot ASIMO, which also depends on this ability to direct first attention towards his user.

6.1.2 Models using Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) in the sense of the statistical modeling approach, as discussed in section 1.2.1Neural Network Modeling Approaches, have proven their ability to handle statistical high dimensional problems successfully.

From an outside point of view the task of sound source localization can be seen as a mapping of the timing and intensity differences between the two sound channels toward an azimuthal -and possibly spherical- vector, pointing towards the sound source location. Hereby, the Head Related Transfer Functions (HRTF) and ear properties cause intensity and timing differences varying between frequency components and from system to system. If these variations are overlapped by echoes in a closed acoustic environments the mapping between input (binaural cues) and output (directional vector) becomes indeed a statistical, multi-dimensional problem.

Taking this perspective, some research groups attempt to solve the problem of echo suppression during sound source localization by the adaptive "learning" behavior of artificial neural networks. Two, more recent models of this kind shall be discussed within this subsection. At the end, another type of ANN applications, concerned with the solution of the "cocktail party effect" will be mentioned and introduced by a recent example.

MLP Model of Arslan
A recent ANN model for sound source localization has been published by Arslan et al. [AS00] at the University of Texas. The core of this model is a Multi Layer Perceptron (MLP), operating in the feed forward mode. It is feeded by feature vectors containing the maximal responding components of an estimated instantaneous cross power spectrum between adjacent pairs of sensors from a 4 microphone array.
Tested with narrow band sounds (tones) and broadband sounds (speech) the model reaches a localization accuracy in the horizontal plane of 3.5 degrees for near as well as distant sound sources, hence independent of the absolute sound pressure.

However, these results have been achieved after training the MLP for a defined position within a specific acoustical environment. With high probability, a new training session would be required if the experimental setup is moved to another location within the room or if the acoustic properties of the environment changes e.g., by opening or closing of window curtains. This limits its applicability to static systems in static environments and prevents its successful operation at mobile systems like robots. Although, the model claims to work in real time mode, the required up front training session, somehow undermines this statement.

In summary, the MLP approach might be well suited to suppress echoes during sound source localization under static conditions, but since the MLP, even after extensive training, represents only a specific solution to the high dimensional statistical problem, the achieved higher accuracy is abrogated by its limited mobility.

RBF-Fuzzy-Backpropagation Models of Nandy

The attempt of the neural network models published by Nandy [NA96],[NA01] is to extract interaural intensity differences (IID) from head-related transfer functions (HRTF) in the form of spectral cues to localize broadband high-frequency auditory stimuli, in azimuth and elevation. The spectral cues are assigned to specific locations using a novel discriminative matching measure (DMM), defined to characterize IID spectral matching.

The employed artificial neural network consists of two-dimensional Gaussians (Radial Bases Functions - RBF) which act as membership functions for the fuzzy set. The Error back-propagation algorithm is used to train the network to correlate input patterns (spectral IID's) to the desired output patterns (sound locations). The fuzzy outputs were used to estimate the location of the source by detecting Gaussians, using the max-energy paradigm. Hereby, the use of training data with additive noise provided good robustness to input errors.
The model shows that HRTF-based spectral IID pattern can provide sufficient information for extracting localization cues using a connectionist paradigm. Successful recognition in the presence of additive noise in the inputs indicates that the computational framework of this model is robust to errors made in estimating the IID patterns. The localization errors for such noisy patterns at various elevations and azimuths are compared and found to be within limits of localization blurs observed in humans.

In a later work [NA01], Nandy considers several solutions to the matching problem from a neural signal processing viewpoint. He compares correlation based approaches with DMM optimization approach and with a non-linear approach based on the error back-propagation algorithm. All three models have been implemented by neural networks and experiments showed that the back-propagation based neural network yields the best results in terms of DMM both for narrow-band and broad band excitation. The back-propagation neural network was also superior in matching noisy HRTF ratio vectors.

Similar to the proposed model of this thesis, Nandy uses frequency specific interaural intensity differences as binaural cues for sound source localization. However, his network also requires extensive training under specific static acoustical conditions and will not be able to perform under constantly changing real world conditions. Since echo responses will significantly change after dislocating the system or changing the environment, the fuzzy-adapted DMM - direction correlation, might no longer fit with the new situation and might lead to mis-judgements of

sound source location. While an ongoing adaptation of fuzzy sets might help to overcome this problem, its success is limited by the supervised learning principle of the error back propagation method. Therefore separate training for each location will be required and prevent the model from application to mobile systems.

Both, the model of Nandy and the one of Arslan are typical examples of the input-output mapping approach described above. Since both employ Artificial Neural Networks (ANN) to relate complex feature vectors, depending on the physics of the acoustical environment, to specific sound source locations, they suffer from the same problem. If the physical environment changes for whatever reason, the employed back-propagation algorithm requires new training, even if the generalization capabilities of ANN are successful under noisy and low SPL conditions. General changes in the input - output relationship cannot be adapted by those ANN without separate new training and therefore limit the performance of those models to acoustically stable environments and physically static positions.

In distinction, the proposed architecture described in chapter 4, is purely feed forward and exclusively relies on feature processing instead of an adaptation to specific relationships. Therefore, it is not confronted with the need to adapt to changing relationships caused by changing physical conditions and is not limited in mobility and applicability to mobile systems.

A completely different approach is pursued by the many researchers attempting to model the "cocktail party effect". Their major goal is the separation of a single sound source out of a complex mixture of sounds - as humans do when listening to a single speaker at a cocktail party. While this goal is not identical with the sound source localization task, many models use binaural cues and artificial neural networks to distinguish between the overlapping frequencies of the different sources.
The cocktail party problem is traditionally treated as blind source separation problem with many techniques offered to handle the separation. Some of the most prominent solutions include the information maximization approach of Bell and Sejnowski [BS95], and the Independent Component Analysis Networks (ICA) examined for example by Comon [Com94], Amari, Cichocki and Yang [ACY96] and Oja and Karhunen [OK95]. A typical attempt to solve the cocktail party problem, based on binaural cues, has been undertaken by Girolami [Gir98] employing an anti-hebbian Maximum Likelihood Estimator (MLE). His model shall represent this specific class of models at this point.

Ani-Hebbian MLE Model of Girolami
Based on comparative experiments under various echoic and unechoic conditions and the early findings of Kaiser and David [KD60] Girolami believes that interaural correlation might be the driving mechanism behind binaural unmasking. Similar to the proposed architecture of this thesis he therefore assumed that contralateral influence between the binaural channels are necessary to perform effective echo cancellation and source separation under echoic conditions. Like many other researchers, he employs a recurrent network structure (shown in figure 6.1) as base for an adaptive algorithm to separate blind sources of convolutive mixtures.

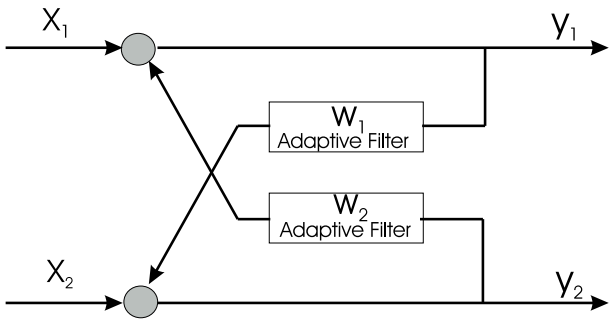

Figure 6.1: Recurrent binaural filter network

The adaptive coefficients of this artificial neural network are determined by an Maximum Likelihood Estimator (MLE), identifying the optimal non-linear term, required to fully utilize the information conveyed by the statistics of naturally occurring speech. The Anti-Hebbian adaptation mechanism used within the MLE reduces the redundancy within the model parameters.

While the experimental results achieved by Girolami under controlled (synthetic) conditions where quite impressive (SNR improvement of 34dB for the target speech), he described his results, under real world conditions (only 5 dB SNR improvement) as disappointing. One of the reasons he identifies is the fact that even the MLE approach assumes the convolution of the independent source signals with a linear model of the transmitting medium represented by a linear causal filter of infinite length. Since real world acoustic transfer functions do not fulfill this requirement, he suggests the use of Bayesian methods but also emphasizes the need for biologically and physiologically inspired models.

Different from the ANN models of Arslan and Nandy, Girolami does not attempt to "learn" a specific input-output relationship. Instead, similar to the proposed architecture, his recurrent network operates purely feed forward. However, the macroscopic adaptation of contralateral filter parameters, in order to model the complex interactions between the left and right perceptive channel, is obviously not sufficient to duplicate biological effects in the auditory brainstem. By selecting highly specific filters, modeling the inverse transfer function of the environment, it essentially lacks the ability to adapt to changing environments and acoustical conditions.
Representing just one example of the many ANN models developed to perform sound source separation based on binaural processing, this model displays the limitations of ANN quite clearly. While they are optimally suited to extract complex inverse models from statistical signals, they lack the natural ability to cope with dynamically changing environments, as present most of time in the real world.

Summarizing this subsection on artificial neural networks it can be stated that ANN are able to perform sound source localization as well as echo/noise suppression fairly well under static conditions. However, tracking a moving sound source or locating a source under changing positions of the perceptive system (as has been shown to work with the proposed biological motivated architecture) remains problematic due to the limited ability of ANN to represent continuously changing transfer functions, within their weights and elements.

6.2 Technical Echo Suppression

As mentioned in the introduction of this chapter, this section will provide a short overview on established technical methods for echo suppression and/or sound source localization and comparably discuss advantages and limitations of the proposed neural architecture.
Since the majority of current technical models follows the adaptive filter approach, this principle will be in focus of the first subsection. The second subsection introduces some other important modeling approaches and compares them to the one chosen in this thesis.

6.2.1 Adaptive Filtering models

During the last 15 years the most common technical application of echo suppression has been the transmission of voice signals via telephone lines. Therefore, the vast majority of echo cancellation model has been concerned with that topic.

The first problem to be solved in the context of voice transmission via telephone lines is the one of suppressing echoes between the near and the far end of the transmission line generated within the wire. Starting out with simple adaptive filtering in the frequency domain in the early 90'th (e.g.[AAS92],[PGT94]), techniques have been improved employing a variety of adaptive filtering techniques in the frequency and time domain to suppress echoes in the single channel case. By today, the well established standard techniques are based on LMS (Least Mean Square) adaptive filter algorithms proving successfully since the type of echoes in concern are not due to room acoustics and therefore do not depend on speaker position. Due to the static nature of the echo generating transfer function within telephone lines, fast adapting filters can easily extract the inverse transmission model and suppress this type of echoes. Although concerned with echo suppression, these models are not at all concerned with sound source localization in reverberating rooms (a single speaker directly talks into a microphone) and shall not be further investigated at this point. The interested reader might refer by example to Kwan [Kwa02], Chin [CFB01], Sugiyama [SJH01] or Haddad [HK00] for further insights.

A more complex echo suppression problem arrives with the duplex transmission of stereo signals as typically asked for in today's telephone-conferencing systems. Stereophonic acoustic echo cancellation (SAEC) is more difficult to solve, because of the necessity to uniquely identify two acoustic paths, which becomes problematic since the two excitation signals are highly correlated. The fundamental problems of SAEC are well described for example by Eneroth [EGGB01] and arise also from a coupling between loudspeakers and microphones of teleconferencing systems. But again, this type of decoupling and echo suppression is not concerned with the sound source localization, the model of this thesis is useful for, and therefore won't be subject of an intensive discussions. Recent publications covering stereophonic acoustic echo cancellation are found i.e. by Gansler [GB00][GB02], Tangsangiumvisai [TCC99] and Benesty [BG99].

A similar but even more complex problem marks the cancellation of echoes during the usage of hands-free telephony equipment. Here, beside the decoupling between loudspeakers and microphone the task is the suppression of echoes resulting from residual (room) acoustics. But the focus here is again on echo cancellation rather than on sound source localization.

One approach to solve the coupling between microphone and loudspeaker has been published by Hansler [HS00] and Bouquin-Jeannes [BJF01]. Both use the speech signal at the near end (opposite to the hands-free equipment) as a *shadow filter* to the signal received from the far end (hands-free equipment). While Hansler et al. perform a correlation analysis in order to suppress the microphone induced echo components, Bouquin-Jeannes and his colleagues employs a speech detector to stop the adaptation of the far end adaptive filter during speech submission. This way he assures that the adaptive filter only models the echo components generated by the room acoustics at the far end of the system which are assumed to be stable, when separated from the microphone induced dynamics. Basically, the far end system is assessed through the power of the residual echo.

Probably, the most modern filtering approach to deal with hands-free telephony problems has been undertaken by the research groups of Enzner [EMV02] and Gustavson [GMJV02]. They both employ *postfilters*, based on the residual echo power spectral density.
Enzner investigated several techniques to estimate the power spectral density of residual echoes, finally identifying a partitioned block-adaptive estimation technique as the most successful in strongly reverberant and noisy acoustic environments. Frequency-domain adaptive filter (FDAF) have been used simultaneously for residual echo power estimation and tracking of the echo path impulse response. In this way, the FDAF and the postfilter concept supplement each other in a true synergy with low complexity.
The postfilter of Gustafsson applies the spectral weighting technique and attenuates both the background noise and the residual echo. Two weighting rules for the postfilter are discussed. The first is a conventional one, known from noise reduction, which is extended to attenuate residual echo as well as noise. The second is a psychoacoustically motivated weighting rule. Both rules were evaluated and compared by instrumental and auditive tests. They succeed about equally well in attenuating the noise and the residual echo. In listening tests, however, the psychoacoustically motivated weighting rule is mostly preferred since it leads to more natural near end speech and to less annoying residual noise.

However both, the *shadow filters* of Hansler and Bouquin-Jeannes as well as the *postfilters* of Enzner and Gustafson rely on the underlying assumption of a stable residual echo transfer function - generating the echo at the site of the hands-free telephone - and being adapted by sub-band adaptive echo compensation filters. This assumption holds for a stable speaker-phone relationship fairly well but is lost during conditions, when speakers at different positions with regard to the microphones are present or the recording system itself moves its position within the room (as in case of mobile systems). In this case, sound source detection will become a separate task (as in today's video conference systems) which is not solvable by means of adaptive filtering.

To summarize, in sight of the author, none of the numerous adaptive filtering models attempts do deal with the localization problem as a prerequisite for speaker identification in video conferencing systems, and the presented architecture might add a useful additional tool to the established techniques for echo suppression during voice transmission based on adaptive filtering.

6.2.2 Other technical models

Beside the adaptive filtering, five other principle approaches have been identified in the field of technical echo suppression and shall be shortly mentioned during this subsection.

Caotic Coding models of Elmirghani and Muller
First of all, a purely technical but smart approach for echo cancellation during telephone transmission is used by Elmirghani [EMC94] and Muller [ME99]. They apply a chaotic-based modulation regime based on logistic mapping to the speech signal. This way the Power Spectral Density (PSD) of the speech signal is whitened and becomes much better distinguishable by dynamic deconvolution algorithms (DBD) as introduced by Muller or even classical echo canceling as employed by Elirghani. Applying the caotic modulation technique they achieved improvements of about 25 dB after 1000-2000 iterations. Although this is impressive results, the caotic coding method is not at all driven by physiological findings and while it proves helpful for the specific task of echo suppression it might prevent the system from further capabilities like sound source localization or speech interpretation.

Microphone array models of Kuo and Dahl
Another purely technical driven coding technique has been employed by Kuo [KP94],[KHP95] and Dahl [DC99]. The Kuo model relays on the fact that each acoustic echo cancellation microphone unit consists of two closely spaced directional microphones pointing in opposite directions and therefore a much lower order adaptive filter is required than in traditional echo cancellation systems. Dahl also uses more than two self calibrating microphone units but doesn't reveal the mechanism behind. Both models basically assign the echo suppression task to the receptor (microphone) using more than two of them and a subtraction of correlated signals as the cue for decomposition of echoes and noise from the target signal. While this is a technically useful method, it doesn't seem to overcome the limitations of sensor driven approaches, who cancel out rather than ignore echo information, resulting in a lack of ability to use the echo information for estimations of room acoustics (i.e. source distances) or the switch of attention from one source to another (i.e. as during the cocktail party effect).

Impulsive Transformation method of Suzuki
In 1999, Suzuki et al. published a paper [SOKM99] where they propose a coding of acoustical signals into impulsive signals. They claim that by conversion to an impulsive signal, the overlap in time between the interference signal pulse and the desired signal pulse is minimized, which helps to secure the interference signal suppression performance of the traditional adaptive filter. Since performance has so far only be demonstrated by computer simulations, it needs to be proven that this assumption also holds under real world conditions. Nevertheless, the discretisation in time, which is the underlying principle of Suzukis method, can be seen as a common feature between the spike based approach of the model designed in this thesis and Suzukis model. Since he still employes adaptive filters, as the main device for echo cancellation, timely discretisation in Suzukis model is rather seen as a signal enhancement method than as an underlying coding principle. Therefore, while the applied coding might be similar, the general principle remains different from the one proposed for the model of this thesis.

Binaural room modeling of Blauert

An important modeling approach to be mentioned during this section is the 3 phase inverse room modeling, proposed by Blauert et al. already in 1992 [BBL92]. It's three phases consists of a "sound"-specification phase, a design phase and a work-plan phase. Binaural recording, reproduction and room simulation techniques are used throughout the three phases allowing for subjective/objective specification and surveillance of the design goals. The binaural room simulation techniques include physical scale models and computer models of different complexity. The modules of the model are: outer-, middle- and inner-ear simulation, binaural processors, and a final evaluation stage. Using this model, various phenomena of sound localization and spatial hearing, such as lateralization, multiple-image phenomena, summing localization, the precedence effect, and auditory spaciousness can be simulated. Finally, an interesting application of Binaural Technology is presented, namely, a Cocktail-Party-Processor. This processor uses the predescribed binaural model to estimate signal parameters of a desired signal which may be distored by any type of interfering signals. In using this strategy the system is able to even separate the signals of competitive speakers.

Although not based on single cell models, the modules of the Blauert model are very similar to the proposed architecture. Based on binaural ear models he evaluates binaural cues expressed in time and intensity differences between corresponding frequency channels. But the general approach here is quite different. Since Blauert attempts to build a specific inverse model of a specific room - later employed to achieve specific design goals - his model will always fit to a specific acoustical environment and the designer will need to step trough the three phases again if the system needs to act in a new environment. While the Blauert model is perfect for the purpose it has been set up for, it cannot be employed for the goal of mobile sound source localization under echoic conditions, the proposed architecture of this thesis seeks a solution for.

Directional estimation models of Chiucchi and Yensen

Finally, the technical models getting closest to the proposed biologically motivated architecture are the ones of Chiucchi [CP01] and Yensen [YGL01]. Since both of them are designed to cope with multiple participant, full-duplex hands-free telephone and video conferencing tasks, they are confronted with microphone speaker coupling, residual echoes as well as cocktail party-like selection tasks. The idea behind their method is, that the desired sound component can be distinguished from echoes and noise by it's Direction Of Arrival (DOA). If this DOA is correctly estimated, only those components featuring the appropriate Interaural Time Delay are selected from the mixture at the far end, transmitted by a single channel and artificially redesigned as stereo signal by an active beamformer at the near end. The result is a "selective listening" to a desired sound source marked by a specific direction of arrival with regard to the stereophonic far end recording system.
While there is definitely no monaural transmission and artificial redesign of interaural features in the auditory system, the principle of sound source separation based on specific interaural cues might well fit to the principles of physiological sound source selection performed in the Thalamus and higher auditory centers.

However, the problem of this method will be exactly the one, being addressed by the proposed model - the secure determination of the direction of arrival. The algorithm employed for this task at the Chiucchi model is fairly basic and exclusively relies on a ITD (or Time Delay of

Arrival - TDOA) estimation, performed by an algorithmic cross phase spectrum estimation. In case of reverberating rooms this method of TDOA estimation will be confronted with all the limitations previously shown for purely ITD sensitive methods which are: Detection of wrong directions after the onset phase due to disturbances by echoes from other directions, non deterministic outputs at higher frequencies since several maxima occur due to wavelength smaller than interaural distances, and the inability to track moving sound sources due to on overlap in time of components arriving from different directions.

While the general principle of Chiucchi and Yensen is seen as the way to go forward in modeling sound source separation based on physiological models, the specific problem of echo suppression during sound source localization has not yet been solved by these models. Therefore the proposed principle of this thesis might add significant performance if applied to the DOA stages of the Chiucchi and Yensen model architecture.

Summarizing this section on technical modeling of echo suppression and sound source localization it can be said that the many echo cancellation models, based on adaptive filtering, mostly address electrical echoes within the wire and coupling echoes caused by the speaker - microphone interaction at the far end of the transmission system. The few filtering models dealing with residual echoes caused by the acoustical environment rely on the extraction of specific inverse acoustic response functions, which will not remain stable under mobile conditions.
The different technical solutions, briefly discussed in the last subsection apply computational efficient and partly physiologically plausible principles for echo suppression, but the only one concerned with the localization of sound sources employs a very basic mechanism, probably not capable of dealing with natural reverberating conditions.
Therefore, even with respect to technical modeling, the proposed architecture pursues a novel approach to an important problem and might enhance the performance of video conferencing systems as well as enable the sound source location in mobile technical systems.

6.3 Silicon Implementability

As mentioned in the introduction of this thesis, a major limitation of physiologically motivated Spike Interaction Models (SIM) arises from their computational expensivnes if simulated on purely digital and sequential von Neumann computing architectures.
This limitation can be turned into an advantage if the engineer attempts to build them into silicon technology. Due to the purely structural coding without any need for algorithmic calculations, as well as due to the very limited variety of necessary base elements, SIM are perfectly suited for analog-digital silicon implementation, this way opening the entire world of truly parallel and on-line processing regardless of the level of model complexity.

Since this fact was in focus of the DFG Graduiertenkolleg "Automatisierung des Entwurfs analog-digitaler Strukturen am Beispiel Neuronaler Netze" funding this thesis until 1998, some of the base elements introduced in chapter 3 have been designed, implemented and tested as paced analog circuitry. This work has been done and published in close cooperation with Richard Izak, Karten Trott and Uwe Markl at the department of Microelectronic Circuits and Systems at the Technical University of Ilmenau [IZTP97], [ITZM97], [ITZ97], [IZ97], [IZ97], [ZITP97], [ZIT97], [Zah96], [MZ96].

This chapter will briefly describe the design principles applied and the preliminary results achieved until the end of DFG support to this thesis in 1998. For further results one might refer to later publications by R.Izak and K.Trott.

According to the available technology to the Graduiertenkolleg at the late 90'th we used a 2.4 μm analog CMOS technology for implementation of single neurons, single synapses and a first test-chip containing an array of 34 neurons and 102 synapses. Turning toward analog implementations one has to be aware of the inaccuracy and stability problems of such systems. In addition, the general robust nature of the information processing in spike based systems, we used two asynchronous clocks of 1 MHz to prevent the system from oscillations and to ensure a time resolution of $1\mu s$ for spike emission and coincidence detection.

The following subsections will now describe the implemented silicon neuron, the implemented silicon synapse and the test chip, designed to perform a fully parallel sound source localization, based on spiking neurons and synapses, realizing the traditional Jefferson model of ITD evaluation in the MSO. Finally, the last section will mention some other neuromorphic approaches to VLSI implementations of spiking neurons and provide a short outlook with regard to the implementability of the proposed architecture for echo suppression during sound source localization.

6.3.1 Silicon Neuron

One of our goals was to find a neural cell model usable at different stages of the system just by tuning of a few electronic parameters. Therefore, we included only the essential functional blocks of Integrate and Fire neurons according to the Base Library element "extended IF neuron" as introduced in section 3.3.2. This was extended by a firing history block, generating the dendritic potential as a prerequisite for Hebbian learning at the site of learning synapses, and achieved the "learning extended IF Neuron" element of the Neural Base Library. The block structure of the silicon neuron is shown in figure 6.2 and contains five blocks, representing functional circuitry.

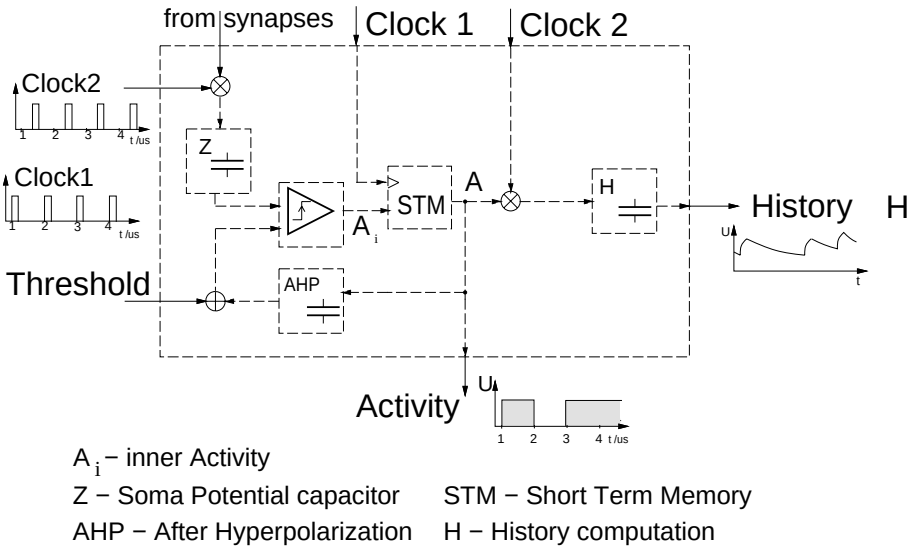

Figure 6.2: Block structure of the silicon neuron

Each neuron receives spatially added current pulses from affiliated synapses placed in the column above. The analog amplitude of each synaptic current depends on the stored weight and extends up to $\pm 15\,\mu A$. A spatio-temporal summation of all incoming current pulses is realized by charging the Soma Potential Capacitor (SPC) Element Z. A discharging resistor in parallel with the $5pF$ capacitor Z approximates the time course of postsynaptic potentials (PSP) as β-function with a repolarizing phase of $30\,\mu s$ and a resting potential of $0V$. The neural activity A at the output of the neuron is the result of the comparison between the soma potential and the threshold. The subsampling of the incoming current pulses by $Clock2$ (1 MHz, pulse width 500 ns) leads to a reduced charge period of the capacitor, extending its functional dynamic range and reducing stabilization errors at the beginning of the μs interval.

If the capacitor voltage Z exceeds threshold, the employed rail-to-rail comparator generates only a trigger point for the inner activity A_i. To prevent the immediate return due to the refractory process (AHP), A_i is stored in an edge triggered dynamic short term memory (STM), realized as gate-capacitance buffer. The binary activity A at the output of the neuron arises synchronous to the next L-H edge of $Clock1$ (see figure 6.3).

Each activity pulse is followed by a refractory period. When A turns to high, the AHP capacitor is charged and the threshold is lifted to V_{dd}. Therefore, the output of the comparator will return to low. This is functionally equivalent to the decrease of soma potential in biological neurons and the simulation block. The AHP potential returns to resting potential in a 2 stage process combining a defined absolute refractory period with an exponential decrease during the relative period.

The History (H) circuit is included to model the firing history of the neuron, which is propagated as dendritic potential H_d back to the site of those synapses, referring to the input of that neuron. During the activity pulse, the H capacitor is charged rapidly. Afterwards it is discharged with a time constant of $30\,\mu s$ similar to the Z capacitor. This decay time constant can be changed via modification of the employed discharge resistor. History potentials range from 2.5 to a maximum of $5V$.

Figure 6.3 shows the designed mixed analog-digital circuit at the left panel and the simulated circuit behavior after stimulation with a constant chain of presynaptic pulses at the right panel.

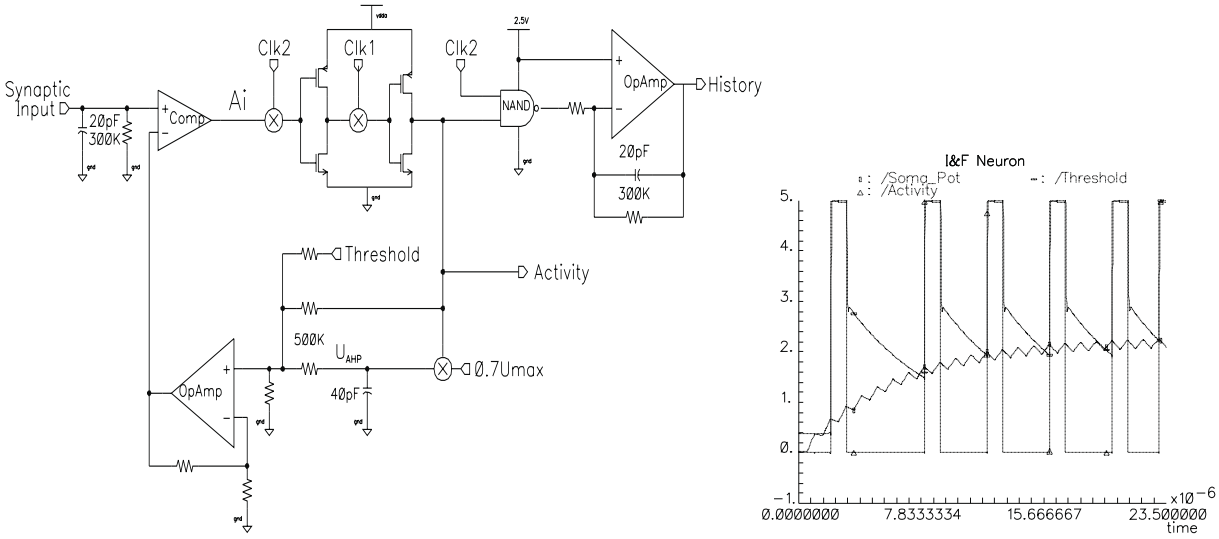

Figure 6.3: **left panel:** *analog-digital circuitry of the silicon neuron*
right panel: *circuitry simulation results after stimulation with presynaptic excitatory pulses*

As can be seen in the right panel of figure 6.3, summation over time is achieved by simply adding charge to the postsynaptic resistor-capacitor (RC) element, which results in a natural saturation, identical to the SIMULINK model.

The parametric values of the extended IF Neuron Base element: threshold, ahp(tau and max) as well as the postsynaptic potential parameters psp(tau-depolarization, tau-repolarization) can be tuned by simply changing the values of the associated RC elements without any need to modify the circuitry its self. This way, the designed extended IF Neuron can be employed at any stage and for any function of the proposed architecture even if the history output is not needed.

Finally, the technologically implementable layout, shown in figure 6.4, has been generated out of the the circuitry using a specially designed CADENCE-based analog-digital layout generator.

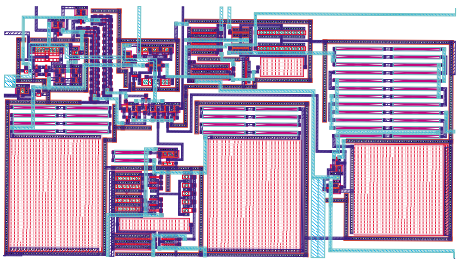

Figure 6.4: Implementable layout for the Silicon Neuron

6.3.2 Silicon Synapse

In general, the task of a silicon synapse is the weighted transmission of presynaptic voltage pulses into postsynaptic current levels within the dendritic tree. Since the Silicon Neuron has been implemented as extended IF Neuron, including the postsynaptic transfer function, it would have been sufficient for the implementation of the proposed architecture to implement the synaptic element as basic synapse, i.e. as a simple weighted transmission of the arriving voltage pulse into a dendritic current pulse. The necessary circuitry to perform this task is a simple Voltage-Current (U/I) converter as shown in figure 6.5.

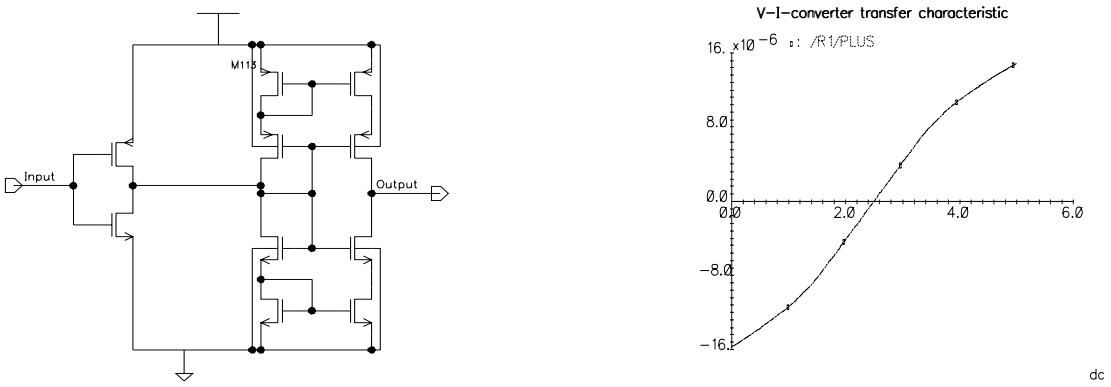

Figure 6.5: **left panel:** *analog circuitry of the voltage-current converter*
right panel: *transfer function of the voltage-current converter*

As can be seen, the nearly linear transfer function assures that positive (excitatory) voltage pulses become converted into inward flowing (charging) currents and negative (inhibitory) voltage pulses result in outward flowing (discharging) currents. This way, excitatory as well as inhibitory synapses become available, depending on the voltage of the arriving pulses. Note that, since neurons always emit positive voltage pulses (action potentials), a inhibitory synapse contains an inverted U/I converter enabling the discharge of the referring PSP- R/C element by an outward current flow.

Since the original purpose of the silicon synapse has been the implementation of adaptive (learning) synapses, referring to the NBL element "TOP-learning synapse", the design asked not only for the storage, but also for a continuous modification of the specific weight value following a purely local Hebbian learning rule.
In this case, the synaptic weight is stored locally, as a voltage across a $5\,pF$ poly capacitance and the weight modification is realized by charging or discharging currents depending on the potential difference between the local postsynaptic potential H and the back propagated dendritic potential H_d of the receiving neuron. Enhanced in Fig.6.6 the weight will increase if the receiving neuron spikes shortly after the arrival of an excitatory pulse and will decrease if the neuron spikes outside the shaded learning window just after the arrival of a postsynaptic spike. This principle was first proposed by Gerstner [GV94] for learning in the dynamic associative memory and successfully simulated by the SIMULINK base library element.

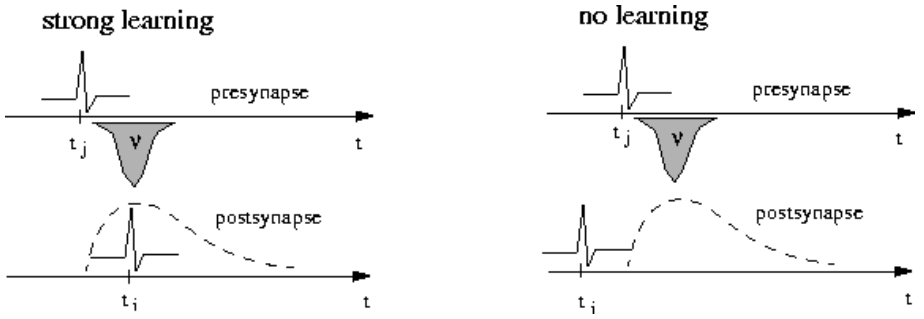

Figure 6.6: Principle of the Hebbian type Timing Of Potential (TOP) learning

The resulting more complex synapse is implemented by four compartments shown in figure 6.7. Besides the described U/I converter and the weight-storage capacitor it contains a weight-update element consisting of an analog multiplier, coupled to a charge pump responsible to modify the charge of the weight capacitor depending on the difference between the local postsynaptic potential H and the local presynaptic potential H_d. For a detailed description of the circuit's functionality the reader should refer to Izak et al.[ITZM97].

The long term storage, as well as the accurate modification of synaptic weights, mark a specific problem in analog hardware since leakage and discharge will cause the charge of the capacitor to disappear over time. To overcome these limitations, a refresh circuitry, as shown in figure 6.8, is introduced as the 4. compartment of the silicon synapse.

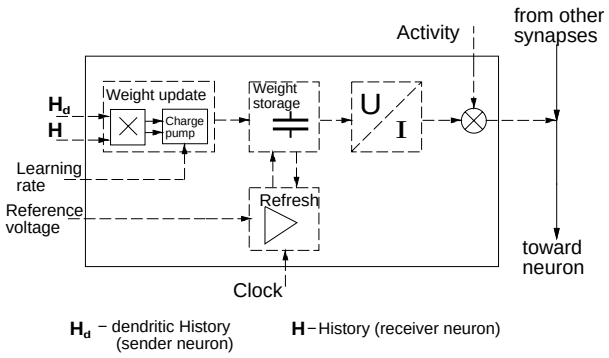

Figure 6.7: Building blocks of the TOP Learning silicon synapse

Observable in figure 6.7, the weight capacitor is charged or discharged by a charge pump depending on the multiplication of dendritic- and postsynaptic potential. Thereby, the applied Gilbert multiplier joins both history potentials with a linearity of 0.6% at $1 \times 1V$ inputs. The disadvantage of its dynamic range in the order of mV has been compensated with a higher comparator sensitivity at the subsequent charge pump, based on the idea of Morie [MA94].

The necessary accuracy level for weight storage and modification has been defined to 8 bit (20 mV accordingly). To achieve this high level of accuracy, the idea of Vittoz et al.[VOM91] to use a ramp reference voltage has been adapted. Here, a specific reference voltage increases continuously from 0 to 5 V during 256 μs and then starts over again. By continuously comparing the voltage across the weight capacitor with this reference, charge leakages as small as 10 mV can be detected and if the reference exceeds the actual capacitor voltage, the weight is carried along

with the reference until the next reset pulse occurs. Using a 1 MHz reset clock and a 256 μs refresh cycle, the achieved accuracy fulfills the 20 mV constraint. Principle and circuitry of the employed refresh method are displayed in figure 6.8.

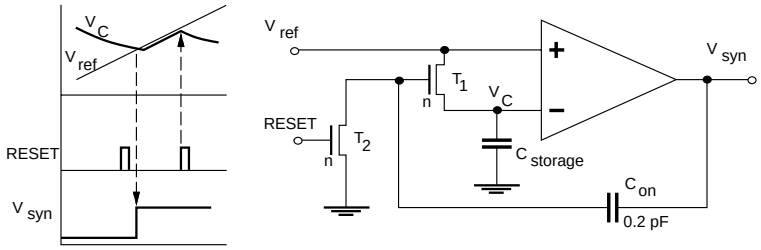

Figure 6.8: Implemented synaptic weight refresh principle as proposed by Vittoz et al.

Applying this principle and realizing all four compartments of the silicon synapse, the entire analog- digital circuitry of a TOP learning silicon synapse is displayed in figure 6.9 left panel. But since this circuitry has not only been simulated but also physically implemented via the Euro-chip program, the weight modification curve shown in the right panel of figure 6.9 contains both, results from circuitry simulations and measured results from physical silicon.

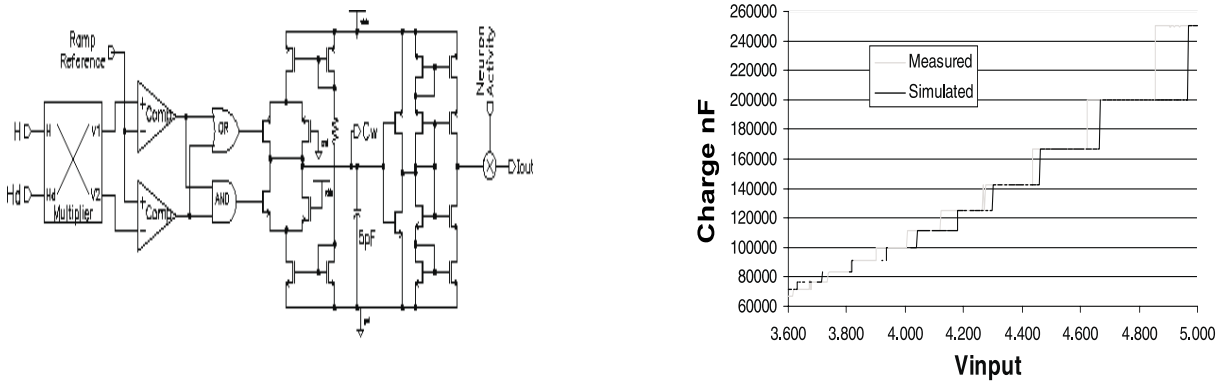

Figure 6.9: **left panel:** *analog-digital circuitry of the silicon synapse*
right panel: *simulated and measured synaptic weight modification during learning*

The implemented 2.4 μm layout generated by the developed CADENCE design generator can be seen in figure 6.10.

6.3.3 Chip design

Neural hardware design can be divided into all purpose and full custom design chips. To achieve the advantages of full custom design (more detailed modeling, optimal area utilization, etc.) by reducing the overwhelming design expense, the automation of layout generation has been the goal of the efforts in the Graduiertenkolleg. Based on a library of neural elements (cell types, synapses, delay units and others), the design generator developed by Trott and Izak is embedded in a CADENCE environment and produces layouts for different net sizes and connectivity

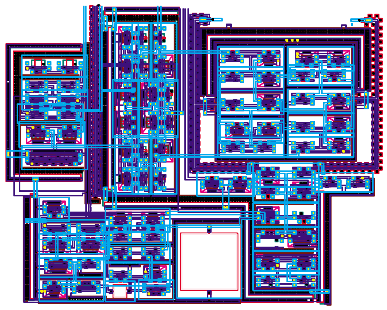

Figure 6.10: Implemented layout of the silicon synapse

structures. The general order of placement so far implemented is a matrix, placing a column of input-synapses above each neuron and enabling a full connectivity based on matrix wiring as shown in figure 6.11 left panel. If more synapses are needed in case of fully interconnected larger networks, an array of cascadable chips is generated, consisting of combined neuron-synapse chips and pure synaptic arrays as displayed in the right panel of figure 6.11. This enables the realization of a wide variety of network topologies, ranging from fully connected toward multi-layer or locally connected architectures.

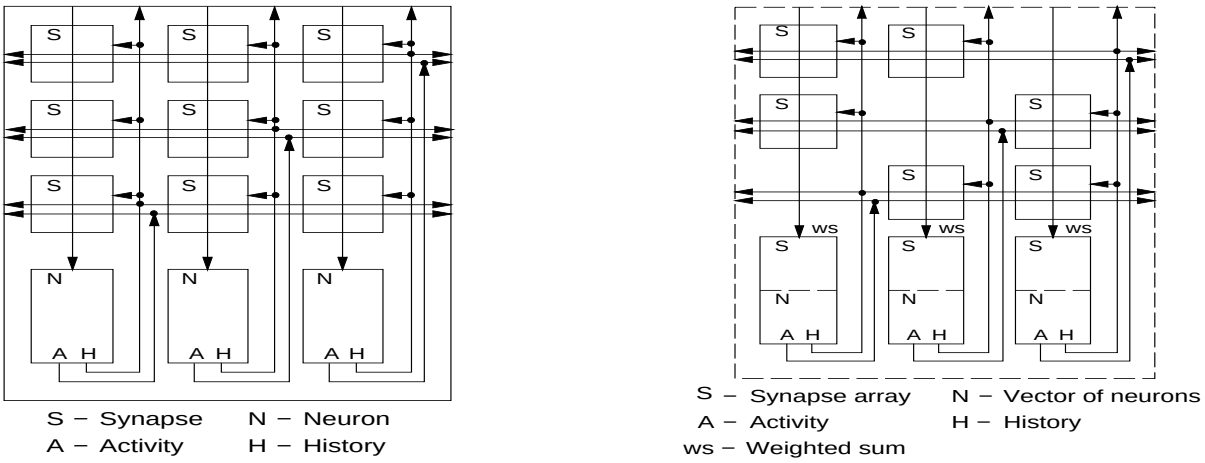

Figure 6.11: **left panel:** *Matrix of neurons with synaptic input colums placed on one chip*
right panel: *Chip cascade combining neurons and pure synaptic arrays*

Using the 2.4 μm CMOS technology at least 10 Neurons with 100 synapses can be placed on a single chip of 50 mm^2 area. However, today's technology in the range of nm allows more than 100 fully connected neurons to be placed onto a single waver. For even larger nets, the generator produces partitioned layouts for multi chip modules. Designed by the team of the Graduiertenkolleg, the generator (figure 6.12) includes optimization rules for wire transfer and routing. This way a wide range of networks and functionality becomes accessible to the inexperienced designer.

A typical example of the resulting neuromorphic chip design, realizing a coincidence based sound source localization by evaluation of interaural time differences, according to the traditional Jefferson model, is displayed in figure 6.13. Here, the left panels enhances the desired functionality and the right one displays the obtained implementable layout.

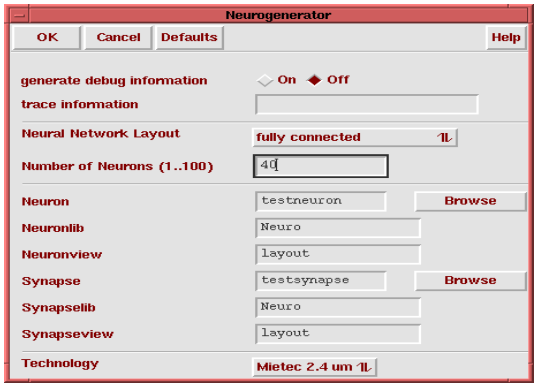

Figure 6.12: User Interface of the Design generator under development

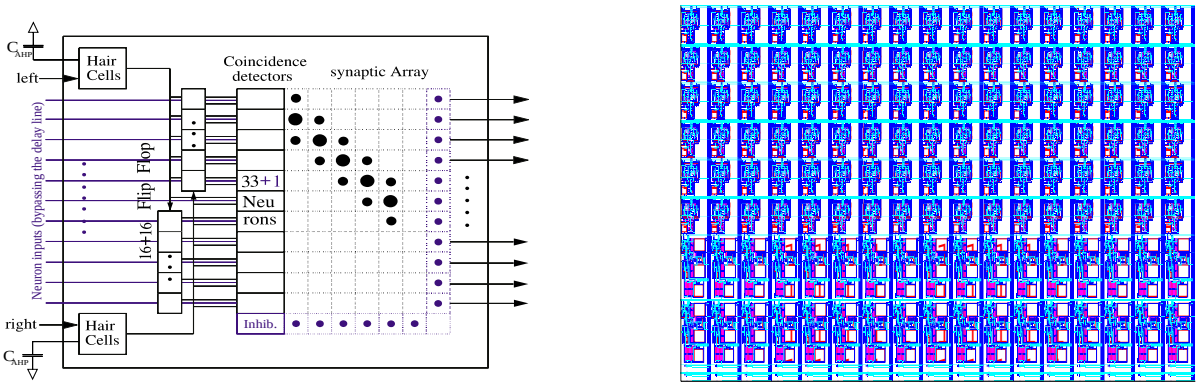

Figure 6.13: **left panel:** *Jefferson model sound source localization via spike coincidence detection* **right panel:** *Automatically generated chip layout to realize the functionality fully parallel*

These simulations and first implementations of silicon neural elements support our postulate that the combination of analog and digital circuit design is especially well suited to unfold the full power of neural networks. The fully parallel spatio-temporal nature of biological information processing and its simple but often repeated calculations demand a large number of distributed simple processors for high speed and accurate computing. At the same time, customized hardware design becomes crucial to the power of applications since neural functionality is mainly based on specific connectivity. So far, three critical elements, an IF neuron with refractory behavior and dendritic potentials, a locally learning synapse, and an optimized connection scheme have been developed. Further elements for axonal delays, non learning synapses and specific hair cells are under consideration.

6.3.4 Implementability of the proposed architecture

As shown by the two examples above it is obviously possible to implement the elements of the neural base library presented in chapter 3 with affordable effort into specialized silicon. Applying the idea of automated layout generation by a simple to use layout generator to translate the parameters of simulation models into specific electronics elements, the creation of special purpose neural processors becomes available to the designer, after successful simulation and evaluation of SIM models based on the SIMULINK environment.

Although a complete neural object library is not yet available, major elements as the IF neuron and a synaptic transmission have been designed and tested. Furthermore proven silicon models of the Lyon cochlear filters have meanwhile been designed and implemented at Bell Labs [SL98] and the swiss center of electronics and microtechnology S.A, [FVV97]. Other research groups like the one of Timmer Horiuchi [Hor01] at the University of Maryland or Andre von Schaik at the University of Sidney [vS01b], [vS01a] successfully developed single spiking neurons as well as different types of synapses and synaptic connections. Their recent work builds on the pioneering work of Carver Mead, John Lazzaro and Christoph Koch at the California Institute of Technology, Eric Vittoz at the EPFL and Terry Sejnowski at the Salk Institute. Horiuchi even implemented the already discussed model of sound source localization based on IID evaluation at the level of the LSO [Hor01].

Hence, specialized neuromorphic silicon is no longer a vision, it becomes reality and will be available to the scientific community and commercial applications before the end of this decade. As postulated by Watson in 1997 in his book "Neuromorphic engineering - Why can't a computer be more like a brain" [Wat97]. This might change our perception on the term "computer" from a standardized digital computing machine to specialized processors employing a completely different and much more brain like scheme of information procession, superior to digital computing in many instances.

With regard to the here proposed neural architecture for echo suppression during sound source localization, there is no doubt that today's technology is ready to implement it into silicon. Based on the currently available technology and the own designs, tests and implementations it should be possible, to implement all simulated structures into three chips representing the hair cell - AVCN level, the SOC level (with LSO and possible MSO models) and the DNLL-IC level. Since the neurons, synapses and connectivities are essentially the same at all levels, the chip design should be fairly straight forward with only some modification of cell parameters at each level. If coupled to a silicon cochlea, already donated to the Ilmenau Graduiertenkolleg by Andreas Andreou from the sensory communication and microsystems laboratory in Baltimore, all simulated building blocks of the proposed architecture seem to be implementable. However, this work has not yet been done and will require further investigations and practical problems will have to be solved. At this point, it can only be viewed as a promising outlook, worth to be pursued, not only for the implementation of the specific architecture proposed in this thesis, but also as a possible way forward to make biological principles of information processing available to man made machines, capable to interact with humans in the real world environment.

Chapter 7

Summary and Outlook

7.1 Summary

In the Introduction of this thesis, three sources of motivation were mentioned. First, the aim to provide a technically feasible and biological plausible model to suppress echoes during man machine interactions. Second, the aim to provide an easy to use modeling system to simulate the dynamics of neurons as spike interaction systems. And third, the aim to provide a physiological plausible explanation for persistent inhibition of neurons within the DNLL. In summary, it can be claimed, that all three aims have been achieved.

As discussed in Section 4, 5.3 and 6 of this thesis, the developed model proves capable of suppressing strong echoes during localization tasks in normal and highly reverberating environments. At the same time it is able to deal with real world signals and fist steps have been taken to enable real time performance based on fully parallel silicon implementation. It will, therefore, help technical systems, like robots, to reliably localize their users. This way, it will be able to apply biological principles to technical systems during man-machine interactions.

As described in chapter 3, a fairly detailed multi purpose simulation system has been designed, satisfying the five reasons to use spike interaction models given in chapter 1. The designed neural base library proved capable of modeling neural structures without demanding extended programming skills but with sufficient detail to assure comparability with physiological experiments. The dynamic elements of the developed neural base library realize a new type of IF neurons containing dynamic transfer functions and enabling the design of spike interaction models as logical extension of the current spike response models.

And finally, the neurophysiological hypothesis of chapter 2 declaring an internal push-pull effect as a reason for persistent inhibition in the DNLL has been supported by the physiological experiments of section 5.1. This was achieved by confirming the existence of persistent inhibitory effects in the Gerbils DNLL. Detailed compliance between the simulated model cells, and the animal experiments suggest, that strong hyperpolarization within the DNLL cells is indeed one of the reasons for their persistent inhibition.

Since the aim of this thesis was to provide a comprehensive view from neurophysiology via psychoacoustic and computational modeling toward neuromorphic implementation it has been limited to the narrow field of echo suppression during sound source localization. However, the applied principles of research and modeling might enable further studies of adjacent fields in auditory processing and help tommorrows engineers and biologists to understand and implement the principles of biological information processing.

7.2 Outlook

This thesis has been an interdisciplinary study and does not claim comprehensiveness nor completeness of the performed investigations. It rather contains proof of concept to five essential areas:

1. the investigation of single neuron dynamic response pattern in the DNLL of the Mongolian Gerbil
2. the development of a multi purpose simulation system for spike interaction models
3. the realization of a biologically realistic computational model of specific nuclei in the auditory brainstem
4. the application of a biological motivated system for echo suppression during sound source localization to technical systems
5. the implementation of spiking neural networks by specialized analog-digital silicon hardware

In all of these areas further investigations are necessary and planed for the upcoming years. A general outlook on the planned activities shall be given within this last chapter. It will address the five areas of investigation separately and conclude with a general outlook towards the future of neuromorphic modeling.

1. Investigation of single neurons within the DNLL

The neurophysiological investigations performed during this study should be continued to further provide the major source for generating ideas and hypothesis for biologically realistic modeling of neural networks usable for verification, forecasting and technical application of biological principles.

For the specific task of investigating the biological correlates of the psycho-acoustical precedence effect, it is first of all planed to continue the electro-physiological experiments within the gerbils DNLL. Here, more cells exhibiting long lasting inhibitory effects need to be found in order to increase the statistical significance of this study and to obtain quantitative measures.

Furthermore, different types of experiments are planed on these cells, including the investigation of influences to the inhibitory effect by the absolute sound pressure level, the duration of the lead and the lag signal as well as the spectrum and the intrinsic dynamic of the stimuli. Finally, this study needs to be extended to the low frequency range, asking whether the perceived persistent inhibition effects are limited to a specific type of high frequency EI cells or if they represent a common feature of the DNLL.

Secondly, a new type of experiments within the gerbils DNLL is planed to prove the push-pull hypothesis of this thesis. Using patch clamp techniques to directly access the time course of soma potentials within the neuron it should be possible to gain direct access to the hypothetical concept of persistent hyperpolarization.

And finally it will be necessary to further investigate the way of neural information towards and from the DNLL, by performing anatomical studies based on antero- and/or retrograde markers and by selectively blocking of GABA and/or glycine.

Since the inferior colliculus is probably the major site, echo suppression is carried out, it should also be subject to further investigations as already started by the groups of Pollack, Litovsky and Kelly among others. Here, a closer cooperation is intended in order verify one of the numerous current hypotheses on echo suppression in the auditory brainstem.

In summary, the physiological experiments need to be continued to further support and prove the hypothesis and findings of this thesis and to extend the model beyond the specific high frequency cells within the DNLL.

2. Further development of the simulation system

The neural base library introduced in chapter 3 is already capable of building more complex and adaptive models of neural systems not only within the auditory brain stem. However, it is still in it's infancy and needs to be grown into an professional and generally applicable simulation system.

To approach this goal, three major steps are planed for the near future. First, to add more detailed elements, capable to model ionic currents and cell potentials in a more realistic way (somewhat similar to the GENESIS level of detail). Second, to further develope the dynamic and adaptive properties of the current elements and to test them by inclusion into biological motivated models (i.e. for the Cliffton effect). And third, to document and publish the NBL in a professional way to ensure easy accessibility and usage by the community of biologists and biological modelers.

Some of these tasks have already been started by documenting the library elements and making it publicly available within the Intranet of the Max Planck Insitute of Neurobiology. Further plans include public accessibility via the Internet and further collaborations with biological scientists and engineers. First steps have already been made by establishing a loose cooperation with David Mc Alpines Lab at the University College in London and some researches using the alpha version of this simulation system at the Max Plack Institute of Neurobiology in Munich.

In summary, the task is to establish the NBL as a common modeling tool among biologists and neuromorphic engineers.

3. Realization of a biologically realistic model of the auditory brainstem

Another plan to further develop the current model has also been started. This is to extend the model by the missing nuclei and connections of the MSO (and the LNTB) in order to achieve a more complete model of the auditory brainstem.

Early simulations of the MSO suggest that the Grothe model of fast inhibition as a major principle to achieve a rate coding of ITD's rather than the traditional place code in the MSO is well possible to be included and leads to a similar representation of azimuthal direction in

the LSO and the MSO. The modeling of their interaction and combined influence on the DNLL promises to lead to very interesting insights and hypotheses on the general representation of azimuthal sound source position as well as on their influence on the dynamic firing pattern in the DNLL and the IC.

A more long term plan is to include adaptive features into the brainstem model in order to investigate the principles of early sound suppression and the Cliffton effect.

Finally, it should be possible to use this model for the evaluation of intrinsic echo parameters like intensity, delay, duration and reverberation radius. With only a few modifications it should then be capable of adding the sensation of sound source distances to the current azimuthal information combining them to a system, able to detect real sound source positions.

In summary, it seems realistic to extend the current model of the auditory brainstem by further essential features and enhance its validation and prediction capabilities beyond the level of IID processing.

4. Application to technical systems

One of the major motivations for this study has been derived from the inability of the mobile robot system PERSES to consistently locate a user in reverberating environments. Since the real world experiments carried out during this study suggest a robust capability of the model to solve this task, it is a major goal for the near future to apply this model to the robot system and test it in the even more complex environment of a department store.

To achieve this goal, real time performance is necessary and will be strived for by using the C-code generation options of SIMULINK. Additionally, the available dSPACE DSP will be employed, to take on a major part of the computing power in demand of the current model. And finally, an optimization and pruning of the intrinsic structures and parameters will be performed in order to limit the number of calculations, necessary to locate a sound source under echoic conditions.

However, besides the application on mobile robots, further areas of application have been identified. One is the application on video conference camera systems, in order to make them direct their objectives towards the current speaker within reverberating conference rooms. Another is the pre-processing of speech signals in order to enhance the correct identification rate of speech identifiers.

In summary, the model shall be computationally speeded up in order to take on specific tasks during man-machine interactions in real-world environments.

5. Silicon Implementation

The sample implementations shown in this thesis are still far from commercial usage, but they prove the concept of silicon implementability for IF-neurons and even for adaptive dynamic features of living cells and synapses.

The direction of future research in the field of analog-digital implementation of cell models is clear and promises a truly parallel processing with a tremendous power to solve man-machine

interaction problems as well as adaptation problems in real-world environments.

The next step will be the revised implementation of the IF-Neuron, of the static synapse and of small networks based on these two elements. As soon as the basic elements become electronically available, today's advanced micro- and nano-technology combined with the developed layout generator will enable even the un-experienced scientist to design their own circuits and to include them into larger technical environments.

Existing contacts with the groups of van Schaik and Andreou [vS01b] [vS01a] are planed to be reactivated in order to speed up the progress in this area and achieve the final goal of truly neuromorphic engineering.

The silicon implementation of the developed neural elements still demands a significant amount of work but promises powerful solutions.

Conclusion

The number and variety of plans and activities mentioned above clearly shows that neuromorphic engineering depends on true interdisciplinary team work which is rather a vision then a reality by today. However, this thesis want's to add a small piece of evidence that this way is a possible and promising approach of computational neuroscience. Once the existing gulces are bridged, it is the believe of the author that computational neuroscience will not only extend the capabilities of today's computers but also alter their general paradigm of information processing during man-machine interactions.

List of Figures

Bibliography

[AAB70] L. M. Aitkin, D. M. Anderson, and J. F. Brugge. Tonotopical organization and discharge characteristics of single neurons in the nuclei of the lateral lemniscus of the cat. *Journal of Physiology*, 33, 1970.

[AAS92] A. ABOUSAADA, T. ABOULNASR, and W. STEENAART. An echo tail canceler based on adaptive interpolated fir filtering. *Ieee Transactions on Circuits and Systems Ii-Analog and Digital Signal Processing*, 39(7):409–416, 1992.

[Abb98] L. Abbott. Temporally asymetric hebbian plasticity: spike synchrony and response variability. *Winterschool on Networks with Spiking Neurons and Synaptic Plasticity*, 1998.

[ACY96] S. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In D. Touretzky, M Mozer, and M. Hasselmo, editors, *Advances in neural information processing systems*, book chapter 8, pages 757–763. MIT Press, 1996.

[AIW84] L. M. Aitkin, D. R. F. Irvine, and W. R. Webster. Central neural mechanisms of hearing. In *Handbook of Physiology, The Nervous System Vol.III Sensory Processes Part 2*, pages 675–737. American Physiological Society, 1984.

[AS00] G. Arslan and E. A. Sakarya. A unified neural-network-based speaker localization technique. *Ieee Transactions on Neural Networks*, 11(4):997–1002, 2000.

[BBL92] J. Blauert, M. BODDEN, and H. LEHNERT. Binaural signal-processing and room acoustics planning. *Ieice Transactions on Fundamentals of Electronics Communications and Computer Sciences*, E75A(11):1454–1459, 1992.

[BBM+02] A. Brand, O. Behrend, T. Marquardt, D. McAlpine, and B. Grothe. Precise inhibition is essential for microsecond interaural time difference coding. *Nature*, 417:543–547, 2002.

[BCV89] J. Blauert, G. Canevet, and T. Voinier. The precedence effect: no evidence for an "active" release process found. *J. Acoust. Soc. Am.*, 85(6):2581–2586, 1989.

[BD88] J. Blauert and P. Divenyi. Spectral selectivity in binaural contralateral inhibition. *Acoustica*, 66:267–274, 1988.

[BG99] J. Benesty and A. Gilloire. A frequency domain stereophonic acoustic echo canceler exploiting the coherence between the channels. *Journal of the Acoustical Society of America*, 106(3):L30–L35, 1999.

[Bis96] C. M. Bishop. *Neural Networks for Pattern Recognition.* Oxford University Press, 1996.

[BJF01] R. Bouquin-Jeannes and G. Faucon. Control of an adaptive echo canceller using a near-end speech detector. *Signal Processing*, 81(3):483–489, 2001.

[BKP00] E. E. Bauer, A. Klug, and G. D. Pollak. Features of contralaterally evoked inhibition in the inferior colliculus. *Hearing Research*, 141(1-2):80–96, 2000.

[Bla74] J. Blauert. *R"aumliches H"oren.* S. Hirzel Verlag Stuttgart, 1974.

[Bla96] J. Blauert. An introduction to binaural technology. In R. Gilkey and T. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments.* Lawrence Erlbaum USA-Hilldale NJ, 1996.

[Bla01] J. Blauert. *Spatial Hearing.* MIT Press, 3'rd edition, 2001.

[Blu31] A. D. Blumlein. Improvements in and relating to sound-transmission, sound-recording and sound-reproducing system. *Brit. Pat.*, (394 325), 1931.

[BMM+99] V. M. Bajo, M. A. Merchan, M. S. Malmierca, F. R. Nodal, and J. G. Bjaalie. Topographic organization of the dorsal nucleus of the lateral lemniscus in the cat. *Journal of Comparative Neurology*, 407(3):349–366, 1999.

[Bos02] M. Bosque. *Understanding 99Tricks.* Writers Club Press, 2002.

[BP01] R. M. Burger and G. D. Pollak. Reversible inactivation of the dorsal nucleus of the lateral lemniscus reveals its role in the processing of multiple sound sources in the inferior colliculus of bats. *Journal of Neuroscience*, 21(13):4830–4843, 2001.

[BR00] J. J. Blum and M. C. Reed. Model calculations of time dependent responses to binaural stimuli in the dorsal nucleus of the lateral lemniscus. *Hearing Research*, 149(1-2):77–90, 2000.

[BS66] H. Babkoff and S. Sutton. End point of lateralizationof dichotic clicks. *J. Acoust. Soc. Am.*, 39:87–102, 1966.

[BS95] A Bell and T Sejnowski. An information maximisation approach to blind separationand blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[BT68] J. C. Boudreau and C Tsuchitani. Binaural interaction in the cat superior olive s segment. *Journal of Neurophysiology*, 31(3):442–454, 1968.

[Bul02] J. A. Bullinaria. *Connectionist Models of Cognition and Perception (Progress in Neural Processing).* Word Scientific, 2002.

[BVdRR98] V. M. Bajo, A. E. P. Villa, F. de Ribaupierre, and E. M. Rouiller. Discharge properties of single neurons in the dorsal nucleus of the lateral lemniscus of the rat. *Brain Research Bulletin*, 47(6):595–610, 1998.

[BWH+02] H.-J. Boehme, T. Wilhelm, T. Hempel, Chr. Schroeter, J. Key, C. Schauer, and H.-M Gross. An approach to multimodal human-machine interaction for intelligent service robots. Elsevier Science, 2002.

[CCC98] H. M. Cai, L. H. Carney, and H. S. Colburn. Model for binaural response properties of inferior colliculus neurons. i. a model with interaural time difference-sensitive excitatory and inhibitory inputs. *Journal of the Acoustical Society of America*, 103(1):475–493, 1998.

[CE76] H. S. Colburn and P. Esquissaud. Auditory-nerve model for interaural time discrimination of high-frequency complex stimuli. *Journal of the Acoustical Society of America*, 59:S23–S23, 1976.

[CF89] R. K. Clifton and R. L. Freyman. Effect of click rate and delay on breakdown of the precedence effect. *Percept. Psychophys.*, 46(2):139–145, 1989.

[CFB01] W. H. Chin and B. Farhang-Boroujeny. Subband adaptive filtering with real-valued subband signals for acoustic echo cancellation. *Iee Proceedings-Vision Image and Signal Processing*, 148(4):283–288, 2001.

[CFLM94] R. K. Clifton, R. L. Freyman, R. Y. Litovsky, and D. MCCALL. Listeners expectations about echoes can raise or lower echo threshold. *Journal of the Acoustical Society of America*, 95(3):1525–1533, 1994.

[CK87] L. E. Cornelisse and J. B. Kelly. The effect of cerebrovascular accident on the ability to localize sounds under conditions of the precedence effect. *Neuropsychologia*, 25(2):449–452, 1987.

[CKW99] L. Chen, J. B. Kelly, and S. H. Wu. The commissure of probst as a source of gabaergic inhibition. *Hearing Research*, 138(1-2):106–114, 1999.

[CL78] H. S. Colburn and J. S. Latimer. Theory of binaural interaction based on auditory-nerve data .3. joint dependence on inter-aural time and amplitude differences in discrimination and detection. *Journal of the Acoustical Society of America*, 64(1):95–106, 1978.

[Cli87] R. K. Clifton. Breakdown of echo suppression in the precedence effect. *J. Acoust. Soc. Am.*, 82(5):1834–1835, 1987.

[CM76] L. Cremer and H. A. Mueller. *Die wissenschaftlichen Grundlagen der Raumakustik*, volume I and II. S.Hirtzel Verlag Stuttgart, 2 edition, 1976.

[Col70] H. S. Colburn. Model of binaural hearing based on auditory nerve patterns. *Journal of the Acoustical Society of America*, 47(1P1):130–134, 1970.

[Col73] H. S. Colburn. Theory of binaural interaction based on auditory-nerve data .1. general strategy and preliminary results on interaural discrimination. *Journal of the Acoustical Society of America*, 54(6):1458–1470, 1973.

[Com94] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36:287–314, 1994.

[CP01] S. Chiucchi and F. Piazza. A virtual stereo approach to stereophonic acoustic echo cancellation. *Signal Processing*, 81(3):491–503, 2001.

[Dam71] Damaske. Head-related two-channel stereophonywith loudspeaker reproduction. *Journal of the Acoustical Society of America*, 50:1109–1115, 1971.

[DC99] M. Dahl and I. Claesson. Acoustic noise and echo canceling with microphone array. *Ieee Transactions on Vehicular Technology*, 48(5):1518–1526, 1999.

[DG87] M. S. Dekin and P. A. Getting. In vitro characterization of neurons in the vertical part of the nucleus tractus solitarius. ii. ionic basis for repetitive firing patterns. *J. Neurophysiology.*, 58, 1987.

[DH62] E. E. David and R. L. Hanson. Binaural hearing and free field effects. In *4'th international Congress on Acoustices*, volume H4, 1962.

[Div92] P. L. Divenyi. Binaural supression of non-echoes. *Journal of the Acoustical Society of America*, 91:1078–1084, 1992.

[DMS01] J Dudel, R Menzel, and R. F. Schmidt. *Neurowissenschaft.* Springer Verlag, Heidelberg, 2 edition, 2001.

[DV39] K. DeBoer and R. Vermeulen. On improving of defect hearing. *Philips Technical Review*, 9:8–13, 1939.

[eaH88] et al. Hafter. The neurobiological bases for hearing. In et al. Hafter, editor, *Auditory Function.* Wiley, New York„ 1988.

[EGGB01] P. Eneroth, S. L. Gay, T. Gansler, and J. Benesty. A real-time implementation of a stereophonic acoustic echo canceler. *Ieee Transactions on Speech and Audio Processing*, 9(5):513–523, 2001.

[EMC94] J. M. H. ELMIRGHANI, S. H. MILNER, and R. A. CRYAN. Echo cancellation strategy using chaotic modulated speech. *Electronics Letters*, 30(18):1467–1468, 1994.

[EMV02] G. Enzner, R. Martin, and P. Vary. Partitioned residual echo power estimation for frequency-domain acoustic echo cancellation and postfiltering. *European Transactions on Telecommunications*, 13(2):103–114, 2002.

[ER97] G. Ehret and R. Rom. *The Central Auditory System.* Oxford University Press, 1997.

[Fau94] L. V. Fausett. *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications.* Prentice Hall, 1994.

[FBAR93] C. L. Faingold, C. A. Boersma-Anderson, and M. E. Randall. Stimulation or blockade of dorsal nucleus of the lateral lemniscusalters binaural and tonic inhibition in contralateral inferior colliculus neurons. *Hearing Research*, 69:98–106, 1993.

[FCL91] R. L. Freyman, R. K. Clifton, and R. Y. Litovsky. Dynamic processes in the precedence effect. *J. Acoust. Soc. Am.*, 90(2 Pt 1):874–884, 1991.

[FKBT95] D. C. Fitzpatrick, S. Kuwada, R. Batra, and C. Trahiotis. Neural responses to simple simulated echoes in the auditory brain stem of the unanesthetized rabbit. *J. Neurophysiol.*, 74(6):2469–2486, 1995.

[FKK+99] D. C. Fitzpatrick, S. Kuwada, D. O. Kim, K. Parham, and R. Batra. Responses of neurons to click-pairs as simulated echoes: auditory nerve to auditory cortex. *J. Acoust. Soc. Am.*, 106(6):3460–3472, 1999.

[Fla60] J. L. Flanagan. Models for aproximating basilar membrane displacement. *Bell Sytems Technology*, 39:1163–1191, 1960.

[FMC98] R. L. Freyman, D. D. McCall, and R. K. Clifton. Intensity discrimination for precedence effect stimuli. *J. Acoust. Soc. Am.*, 103(4):2031–2041, 1998.

[FVV97] E. Fragniere, A. VanSchaik, and E. A. Vittoz. Design of an analogue vlsi model of an active cochlea. *Analog Integrated Circuits and Signal Processing*, 13(1-2):19–35, 1997.

[GB00] T. Gansler and J. Benesty. Stereophonic acoustic echo cancellation and two-channel adaptive filtering: an overview. *International Journal of Adaptive Control and Signal Processing*, 14(6):565–586, 2000.

[GB02] T. Gansler and J. Benesty. New insights into the stereophonic acoustic echo cancellation problem and an adaptive nonlinearity solution. *Ieee Transactions on Speech and Audio Processing*, 10(5):257–267, 2002.

[GBHM81] K. K. Glendering, B. N. Baker, K. A. Hutson, and R. Masterton. Ascending auditory afferents to the nuclei of the lateral lemniscus. *Comparative Neurology*, 197:673–703, 1981.

[Gir98] M. Girolami. A nonlinear model of the binaural cocktail party effect. *Neurocomputing*, 22(1-3):201–215, 1998.

[GK02] W. Gerstner and W. Kistler. *Spiking Neuron Models: An Introduction.* Cambridge University Press, Cambridge, gerstner,w.; kistler,w. edition, 2002.

[GMJV02] S. Gustafsson, R. Martin, P. Jax, and P. Vary. A psychoacoustic approach to combined acoustic echo cancellation and noise reduction. *Ieee Transactions on Speech and Audio Processing*, 10(5):245–256, 2002.

[GN00] B. Grothe and G. Neuweiler. The function of the medial superior olive in small mammals: temporal receptive fields in auditory analysis. *Journal of Comparative Physiology A-Sensory Neural and Behavioral Physiology*, 186(5):413–423, 2000.

[GS93] B. Grothe and D. H. Sanes. Bilateral inhibition by glycinergic afferents in the medial superior olive. *Journal of Neurophysiology*, 69:1192–1196, 1993.

[GV94] W. Gerstner and J. L. VanHemmen. Coding and information processing in neural networks. In E. Domany, J. L. VanHemmen, and K. Schulten, editors, *Models of Neural Networks II*, pages 1–93. Springer Verlag, 1994.

[GZ96] A. W. Gummer and P. Zenner. Central processing of auditory information. In *Comprehensive Human Physiology*, book chapter 36, pages 729–756. Springer Verlag Heidelberg, 1996.

[HA97] R. H. Helfert and A. Aschoff. Superior olivary complex and nuclei of the lateral lemniscus. In *The Central Auditory System*, book chapter 3, pages 193–259. Oxford University Press, 1997.

[Haa51] H. Haas. Ueber den einfluss des einfachechos auf die hörsamkeit von sprache. *Acoustica*, 1:49–58, 1951.

[HCC01] M. G. Heinz, H. S. Colburn, and L. H. Carney. Evaluating auditory performance limits: I. one-parameter discrimination using a computational model for the auditory nerve. *Neural Computation*, 13(10):2273–2316, 2001.

[Heb49] D. O. Hebb. *The Organization of Behaviour, A neurophysiological Theory.* John Wiley, 1949.

[Hen49] J. Henry. Presentation for the american association for the advancement of science. In *Scientific Writings of Joseph Henry*, volume Part II, pages 295–296. Smithsonian Institution, 1849.

[HH52] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology London*, 117, 1952.

[HK49] A. L. Hodgkin and B. Katz. The effect of sodium ions on the electrical activity of the giant axon of the squid. *Journal Physiology London*, 108, 1949.

[HK00] T. F. Haddad and M. A. Khasawneh. A new forced lms-based adaptive algorithm utilizing the principle of potential energy. *Journal of the Franklin Institute-Engineering and Applied Mathematics*, 337(5):515–542, 2000.

[HK01] A. Hyvarinen and J. Karhunen. *Independent Component Analysis.* John Wiley and Sons, 2001.

[Hop82] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci.*, 79, 1982.

[Hor01] T. Horiuchi. A vlsi-based model of azimuthal echolocation in the big brown bat. *Autonomous Robots*, 11(3):241–247, 2001.

[HS86] R. H. Helfert and I. R. Schwartz. Morphological evidence for the existence of multiple neuronal classes in the cat lateral superior olivary nucleus. *Journal of Computational Neurology*, 244, 1986.

[HS00] E. Hansler and G. U. Schmidt. Hands-free telephones - joint control of echo cancellation and postfiltering. *Signal Processing*, 80(11):2295–2305, 2000.

[ITZ97] R. Izak, K. Trott, and T. Zahn. Hardware implementation of a mixed analog-digital neural network. *Proceedings of the Dortmunder Fuzzy Tage*, 1997.

[ITZM97] R. Izak, K. Trott, T. Zahn, and U. Markl. Analoge synapse f"ur ein burstpropagierendes neuronales netz. *Proceedings of the international E. I. S Workshop at the University of Hamburg*, 1997.

[IvK96] M. Ito, B. vanAdel, and J. B. Kelly. Sound localization after transection of the commissure of probst in the albino rat. *Journal of Neurophysiology*, 76(5):3493–3502, 1996.

[IZ97] R. Izak and T. Zahn. Vlsi implementation of a neural network for acoustic processing. *Journal of Analog Integrated Circuits and Signal Processing*, 2:1–8, 1997.

[IZTP97] R. Izak, T. Zahn, K. Trott, and P. Paschke. A mixed-signal neural network for auditory attention. *Proceesings of the Micro Neuro Conference Dresden*, 1997.

[Jef48] L. A. Jeffress. A place theory of sound localization. *Journal of Comperative Physiological Physiology*, 61, 1948.

[KBK98] J. B. Kelly, A. D. Buckthought, and S. A. Kidd. Monaural and binaural response properties of single neurons in the rat's dorsal nucleus of the lateral lemniscus. *Hearing Research*, 122(1-2):25–40, 1998.

[KBP99] A. Klug, E. E. Bauer, and G. D. Pollak. Multiple components of ipsilaterally evoked inhibition in the inferior colliculus. *Journal of Neurophysiology*, 82(2):593–610, 1999.

[KD60] J. F. Kaiser and E. E. David. Reproducing the cocktail party effect. *Journal of the Acoustical Society of America*, 32(7):918–918, 1960.

[KHP95] S. M. Kuo, Y. C. Huang, and Z. B. Pan. Acoustic noise and echo cancellation microphone system for videoconferencing. *Ieee Transactions on Consumer Electronics*, 41(4):1150–1158, 1995.

[KHRT94] N. Kunstmann, C. Hillermeier, B. Rabus, and P. Tavan. An associative memory that can form hypotheses: A phase-coded neural network. *Biological Cybernetics*, 72:118–132, 1994.

[KK96] S. A. Kidd and J. B. Kelly. Contribution of the dorsal nucleus of the lateral lemniscus to binaural responses in the inferior colliculus of the rat: Interaural time delays. *Journal of Neuroscience*, 16(22):7390–7397, 1996.

[KK00] J. B. Kelly and S. A. Kidd. Nmda and ampa receptors in the dorsal nucleus of the lateral lemniscus shape binaural responses in rat inferior colliculus. *Journal of Neurophysiology*, 83(3):1403–1414, 2000.

[KL97] J. B. Kelly and L. Li. Two sources of inhibition affecting binaural evoked responses in the rat's inferior colliculus: The dorsal nucleus of the lateral lemniscus and the superior olivary complex. *Hearing Research*, 104(1-2):112–126, 1997.

[KLv96] J. B. Kelly, L. Li, and B. vanAdel. Sound localization after kainic acid lesions of the dorsal nucleus of the lateral lemniscus in the albino rat. *Behavioral Neuroscience*, 110(6):1445–1455, 1996.

[KLvAI98] J. B. Kelly, A. Liscum, B. van Adel, and M. Ito. Projections from the superior olive and lateral lemniscus to tonotopic regions of the rat's inferior colliculus. *Hearing Research*, 116(1-2):43–54, 1998.

[KMF+93] J. A. KALTENBACH, R. J. MELECA, P. R. FALZARANO, S. F. MYERS, and T. H. SIMPSON. Forward masking properties of neurons in the dorsal cochlear nucleus - possible role in the process of echo suppression. *Hearing Research*, 67(1-2):35–44, 1993.

[Koe98] G. Koehler. *Implementation eines Cochlea Modells auf der Basis verkoppelter Resonanzfilter*. Studienarbeit, Technical University of Ilmenau, Dep.of Neuroinformatics, 1998.

[Koh77] T. Kohonen. *Associative Memory: A System-Theoretical Approach*. Springer Verlag, 1977.

[KP94] S. M. Kuo and Z. B. Pan. Development and analysis of distributed acoustic echo cancellation microphone system. *Signal Processing*, 37(3):333–344, 1994.

[KPP95] A. Klug, T. J. Park, and G. D. Pollak. Glycine and gaba influence binaural processing in the inferior colliculus of the moustache bat. *Journal of Neurophysiology*, 74(4):1701–1713, 1995.

[KSJ91] E. Kandel, J. H. Schwartz, and T. M. Jessell. *Principles of Neuroscience*. Appleton & Lange, third) edition, 1991.

[KSSG02] C. Kapfer, A. H. Seidl, H. SCHWEIZER, and B. Grothe. Experience-dependent refinement of inhibitory inputs to auditory coincidence-detector neurons. *nature neuroscience*, 5(3):247–253, 2002.

[KT96] C. H. Keller and T. T. Takahashi. Binaural cross-correlation predicts the responses of neurons in the owl's auditory space map under conditions simulating summing localization. *Journal of Neuroscience*, 16(13):4300–4309, 1996.

[Kwa02] H. K. Kwan. Adaptive iir digital filters with saturation outputs for noise and echo cancellation. *Electronics Letters*, 38(13):661–663, 2002.

[LB58] J. P. A. Lochner and J. F. Burger. The subjective masking of short time delayed echoes their primary sounds and their contributionto the intellegibility of speech. *Acoustica*, 8:1–10, 1958.

[LCYG99] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman. The precedence effect. *J. Acoust. Soc. Am.*, 106(4 Pt 1):1633–1654, 1999.

[LD02] R. Y. Litovsky and B. Delgutte. Neural correlates of the precedence effect in the inferior colliculus: effect of localization cues. *J. Neurophysiol.*, 87(2):976–994, 2002.

[Lei02] C. Leibold. *Die Rolle synaptischer Plastizität bei der zeitcodierten Schallokalisation*. Phd, dr. rer. nat., Technical University of Munich, Department of Physics, 2002.

[LFT02] R. Y. Litovsky, B. J. Fligor, and M. J. Traino. Functional role of the human inferior colliculus in binaural hearing. *Hearing Research*, 165(1-2):177–188, 2002.

[LHFZ00] R. Y. Litovsky, M. L. Hawley, B. J. Fligor, and P. M. Zurek. Failure to unlearn the precedence effect. *J. Acoust. Soc. Am.*, 108(5 Pt 1):2345–2352, 2000.

[Lit97] R. Y. Litovsky. Developmental changes in the precedence effect: estimates of minimum audible angle. *J. Acoust. Soc. Am.*, 102(3):1739–1745, 1997.

[LK92] L. Li and J. B. Kelly. Inhibitory influences of the dorsal nucleus of the lateral lemniscus on binaural responses in the rat's inferrior colliculus. *J. Neurosci.*, 12:4530–4539, 1992.

[LLV74] W. J. Loskota, P. Lomax, and A. M. Verity. *A Stereotaxic Atlas of the Mongolian Gerbil Barin.* Ann Arbor Science Publishers, Inc., Ann Arbor, Michigan 48106, 1974.

[LM88] C. F. Lyon and C. Mead. An analog electronic cochlea. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36:1119–1134, 1988.

[LP97] E. Liebenthal and H. Pratt. Evidence for primary auditory cortex involvement in the echo suppression precedence effect: a 3clt study. *J. Basic Clin. Physiol Pharmacol.*, 8(3):181–201, 1997.

[LP99] E. Liebenthal and H. Pratt. Human auditory cortex electrophysiological correlates of the precedence effect: Binaural echo lateralization suppression. *Journal of the Acoustical Society of America*, 106(1):291–303, 1999.

[LRYH97] R. Y. Litovsky, B. Rakerd, T. C. Yin, and W. M. Hartmann. Psychophysical and physiological evidence for a precedence effect in the median sagittal plane. *J. Neurophysiol.*, 77(4):2223–2226, 1997.

[LSC01] R. Y. Litovsky and B. G. Shinn-Cunningham. Investigation of the relationship among three common measures of precedence: fusion, localization dominance, and discrimination suppression. *J. Acoust. Soc. Am.*, 109(1):346–358, 2001.

[Lyo97] R. F. Lyon. All-pole models of auditory filtering. In *Diversity in Auditory Mechanics*, pages 205–211. World Scientific Publishing, 1997.

[MA94] T. Morie and Y. Amemiya. An all-analog expendable neural network lsi with on-chip backpropagation learning. *Journal of SCC*, 29(6):1086–1093, 1994.

[Mar91] T. M. Martinetz. A 'neural gas' network learns topologies. In *Proccedings of of the International Conference on Artificial Neural Networks ICANN91 Espo Finland.* Elsevier Science Publishers), 1991.

[Max33] J. P. Maxfield. Some physical factors affecting the illusion in sound motion pictures. *J. Acoust. Soc. Am.*, 4:69–80, 1933.

[MCR90] C. A. Moore, J. L. Cranford, and A. E. Rahn. Tracking of a "moving" fused auditory image under conditions that elicit the precedence effect. *J. Speech Hear. Res.*, 33(1):141–148, 1990.

[ME99] A. Muller and J. M. H. ELMIRGHANI. Blind channel estimation and echo cancellation using chaotic coded signals. *Ieee Communications Letters*, 3(3):72–74, 1999.

[MLB+98] M. S. Malmierca, T. B. Leergaard, V. M. Bajo, J. G. Bjaalie, and M. A. Merchan. Anatomic evidence of a three-dimensional mosaic pattern of tonotopic organization in the ventral complex of the lateral lemniscus in cat. *Journal of Neuroscience*, 18(24):10603–10618, 1998.

[MLFS97] H. Markram, J. Luebke, M. Frotscher, and B. Sakmann. Regulation of synaptic efficency by coincidence of postsynaptic ap's and epsp's. *Science*, 275 (5297):213–215, 1997.

[MM01] B. J. Mickey and J. C. Middlebrooks. Responses of auditory cortical neurons to pairs of sounds: correlates of fusion and localization. *J. Neurophysiol.*, 86(3):1333–1350, 2001.

[Moo87] J. K. Moore. The human auditory brain stem: A comparative view. *Hearing Research*, 29:1–32, 1987.

[MP43] W. S. McCulloch and W. H. Pitts. A logical calculus of the ideas imminent in nervous activity. *Bulletin Mathematical Biophysics.*, 5, 1943.

[MP93] N. S. MARKOVITZ and G. D. Pollak. The dorsal nucleus of the lateral lemniscus in the moustache bat - monaural properties. *Hearing Research*, 71(1-2):51–63, 1993.

[MP94] N. S. MARKOVITZ and G. D. Pollak. Binaural processing in the dorsal nucleus of the lateral lemniscus. *Hearing Research*, 73(1):121–140, 1994.

[MP02] D. McAlpine and A. R. Palmer. Blocking gabaergic inhibition increases sensitivity to sound motion cues in the inferior colliculus. *J. Neurosci.*, 22(4):1443–1453, 2002.

[MRLB95] M. S. Malmierca, A. Rees, F. E. N. LeBeau, and J. G. Bjaalie. Laminar organization of frequency-defined local axons within and between the inferior colliculi of the guinea pig. *Journal of Comparative Neurology*, 357:124–144, 1995.

[MS52] E. Meyer and G. R. Schodder. Ueber den einfluss von schallrueckwuerfen auf richtungslokalisation und lautstaerke bei sprache. In *Nachrichten der Akademie der Wissenschaften in Goettingen*, volume H6, pages 31–42. Vandenhoeck und Ruprecht, 1952.

[MZ96] U. Markl and T. P. Zahn. Hardwareimplementation pulscodierter neuronaler netze. *Proceedings of the DOCINN Conference in Stuttgart*, 1996.

[NA96] D. Nandy and J. Ben Arie. An auditory localization model based on high-frequency spectral cues. *Ann. Biomed. Eng*, 24(6):621–638, 1996.

[NA01] D. Nandy and J. Ben Arie. Neural models for auditory localization based on spectral cues. *Neurol. Res.*, 23(5):489–500, 2001.

[Ner88] W. Nernst. On the kinetics of substances in solution. *Z. physik. Chemie*, 2:613–622, 1888.

[OK95] E Oja and J Karhunen. Signal separation by nonlinear hebbian learning. In M. Palaniswami, Y. Attikiouzel, R. Marks, D Fogel, and T Fukada, editors, *Computational Intelligence-a dynamic system perspective*, pages 83–97. IEEE Press, 1995.

[PBK03] G. D. Pollak, R. M. Burger, and A. Klug. Dissecting the circuitry of the auditory system. *Trends in Neuroscience*, 26:33–39, 2003.

[PBP+02] G. D. Pollak, R. M. Burger, T. J. Park, A. Klug, and E. E. Bauer. Roles of inhibition for transforming binaural properties in the brainstem auditory system. *Hearing Research*, 168(1-2):60–78, 2002.

[PGT94] T. PETILLON, A. Gilloire, and S. THEODORIDIS. The fast newton transversal filter - an efficient scheme for acoustic echo cancellation in mobile radio. *Ieee Transactions on Signal Processing*, 42(3):509–518, 1994.

[Pol97] G. D. Pollak. Roles of gabaergic inhibition for the binaural processing of multiple sound sources in the inferior colliculus. *Annals of Otology Rhinology and Laryngology*, 106(5):44–54, 1997.

[Pol02] G. D. Pollak. Neurobiology - model hearing. *Nature*, 417(6888):502–503, 2002.

[PP93a] T. J. Park and G. D. Pollak. Gaba shapes a topographic organization of response latency in the moustache bats inferior colliculus. *Journal of Neuroscience*, 13(12):5172–5187, 1993.

[PP93b] T. J. Park and G. D. Pollak. Gaba shapes sensitivity to interaural intensity disparities in the moustache bats inferior colliculus - implications for encoding sound location. *Journal of Neuroscience*, 13(5):2050–2067, 1993.

[PP93c] G. D. Pollak and T. J. Park. The effects of gabaergic inhibition on monaural response properties of neurons in the moustache bats inferior colliculus. *Hearing Research*, 65(1-2):99–117, 1993.

[PP94] T. J. Park and G. D. Pollak. Azimuthal receptive-fields are shaped by gabaergic inhibition in the inferior colliculus of the moustache bat. *Journal of Neurophysiology*, 72(3):1080–1102, 1994.

[PPWL92] T. J. Park, G. D. Pollak, J. A. Winer, and D. T. LARUE. The circuitry and functional-role of inhibitory projections to neurons sensitive to interaural intensity difference in the inferior colliculus. *Journal de Physique Iv*, 2(C1):185–188, 1992.

[RA97] R Romand and P. Avan. The central auditory system. book chapter 2 Anatomical and Functional Aspects of the Cochlear Nucleus. Oxford University Press, 1997.

[RB99] M. C. Reed and J. J. Blum. Model calculations of steady state responses to binaural stimuli in the dorsal nucleus of the lateral lemniscus. *Hearing Research*, 136(1-2):13–28, 1999.

[RE01] S. Roberts and R. Everson. *Independent Component Analysis: Principles and Practice*. Cambridge University Press, 2001.

[Ros58] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 1958.

[Rou97] E. M. Rouiller. The central auditory pathway. book chapter 1 Functional Organization of the Auditory Pathways. Oxford University Press, 1997.

[Rug92] M. A. Ruggero. Coding of sound in the auditory nerve. In A. Popper and R. Fay, editors, *The Mammalian Auditory Pathway*, book chapter 2. Springer Verlag NewYork, 1992.

[SGB02] D. Surmeli, H.-M. Gross, and H.-J. Boehme. Behavior coordination for a mobile visuo-motor system in an augmented real-world environment. In *From Animals to Animats 7 -Intern. Conf. on Simulation of Adaptiv Behavior*, pages 103–104. MIT Press, 2002.

[SHP99] O. Schwartz, J. G. Harris, and J. C. Principe. Modeling the precedence effect for speech using the gamma filter. *Neural Networks*, 12(3):409–417, 1999.

[SJH01] A. Sugiyama, Y. Joncour, and A. Hirano. A stereo echo canceler with correct echo-path identification based on an input-sliding technique. *Ieee Transactions on Signal Processing*, 49(11):2577–2587, 2001.

[SL98] R. Sarpeshkar and R. F. Lyon. A low-power wide-dynamic-range analog vlsi cochlea. *Analog Integrated Circuits and Signal Processing*, 16(3):245–274, 1998.

[Sla93] M Slaney. Auditory toolbox. Aplle Technical Report 45, 1993.

[SN80] F. J. Sigworth and E. Neher. Single na channel currents observed in cultured rat muscle cells. *Nature*, 287, 1980.

[SOH88] A. Shneiderman, D. L. Oliver, and C. K. Henkel. The connections of the dorsal nucleus of the lateral lemniscus. an inhibitory parallel pathway in the ascending auditory system. *Journal of Comparative Neurology*, 276:188–208, 1988.

[SOKM99] N. Suzuki, A. Okamura, T. Kirimoto, and S. Mano. An echo canceler to suppress interference pulses closely correlated to desired pulses. *Electronics and Communications in Japan Part I-Communications*, 82(8):22–31, 1999.

[SP90] K. Saberi and L. Perrott. *J. Acoust. Soc. Am.*, 87:1732–1737, 1990.

[TCC99] N. Tangsangiumvisai, J. A. Chambers, and A. G. Constantinides. Higher-order time-varying allpass filters for signal decorrelation in stereophonic acoustic echo cancellation. *Electronics Letters*, 35(1):88–90, 1999.

[vAKK99] B. A. van Adel, S. A. Kidd, and J. B. Kelly. Contribution of the commissure of probst to binaural evoked responses in the rat's inferior colliculus: interaural time differences. *Hearing Research*, 130(1-2):115–130, 1999.

[Vap95] V. Vapnik. *The Nature of Statistical Learning Theory.* Springer, 1995.

[VC71] V. N. Vapnik and A. Y. Chervonenkis. On the uniform convergenceof relative frequencies of events to their probabilities. *Theory Probab. Appl.*, 16, 1971.

[VDC02] J. L. VanHemmen, E. Domany, and G. D. Cowan. *Models of Neural Networks 4. Early Vision and Attention.* Springer Verlag Telos, 2002.

[vEMH26] v. E. M. Hornbostel. Das r"aumliche h"oren. In A. Bethe, editor, *Handbuch der normalen und pathologischen Physiologie*, volume 11, pages 601–618. Springer Verlag Berlin, 1926.

[VGH+90] J. L. VanHemmen, W. Gerstner, A. Herz, R. Kühn, B. Sulzer, and M. Vaas. Encoding and decoding of patterns, which are correlated in space and time. In G. Dorffner, editor, *Konnektionismus in Artificial Intelligence and Kognitionsforschung*, pages 153–162. Springer Verlag, 1990.

[VHKG92] M. Vater, H. Habbicht, M. Koessl, and B. Grothe. The functional role of gaba and glycine in monaural and binaural processing in the inferrior colliculus of horseshoe bats. *Journal of Comparative Physiology A-Sensory Neural and Behavioral Physiology*, 171:541–553, 1992.

[VOM91] E Vittoz, H. Oguey, and H Maher. Analog storage of adjustable synaptic weights. In O. Nys, E. Dijkstra, and M Chevroulet, editors, *VLSI design of Neural Networks*, pages 47–63. Kluwer Academic Publishers, 1991.

[vS01a] A. van Schaik. An analog vlsi model of periodicity extraction in the human auditory system. *Analog Integrated Circuits and Signal Processing*, 26(2):157–177, 2001.

[vS01b] A. van Schaik. Building blocks for electronic spiking neural networks. *Neural Networks*, 14(6-7):617–628, 2001.

[War41] H. Warncke. Die grundlagen der raumbezueglichen stereophonischen uebertragung im tonfilm. *Akust. Z.*, (6):174–188, 1941.

[Wat97] A. Watson. Neuromorphic engineering - why can't a computer be more like a brain. *Science*, 277(5334):1934–1936, 1997.

[WDCW78] I. C. Whitfield, I. T. Diamond, K. Chiveralls, and T. G. Williamson. Some further observations on the effects of unilateral cortical ablation on sound localization in the cat. *Exp. Brain Res.*, 31(2):221–234, 1978.

[Wen63] K. Wendt. *Das Richtungshoeren bei der Uebertragung zweier Schallfelder bei Intensitaets und Laufzeitstereophonie.* PhD thesis, TH Aachen, 1963.

[Whi74] I. C. Whitfield. A possible neurophysiological basis for the precedence effect. *Fed. Proc.*, 33(8):1915–1916, 1974.

[WHL+89] A. Waibel, T. Hanazawa, K. J. Lang, G. Hinton, and K. Shikano. Phoneme recognition using time-delay neural networks. *EEE Transactions on Acoustics, Speech, and Signal Processing,*, 37:328–339, 1989.

[WK95] S. H. Wu and J. B. Kelly. In vitro brain slice studies of the rats dorsal nucleus of the lateral lemniscus. *Journal of Neurophysiology*, 73:794–809, 1995.

[WLP95] J. A. Winer, D. T. LARUE, and G. D. Pollak. Gaba and glycine in the central auditory-system of the moustache bat - structural substrates for inhibitory neuronal organization. *Journal of Comparative Neurology*, 355(3):317–353, 1995.

[WNR49] H. WALLACH, E. B. NEWMAN, and M. R. ROSENZWEIG. The precedence effect in sound localization. *American Journal of Psychology*, 62(3):315–336, 1949.

[WNR73] H. WALLACH, E. B. NEWMAN, and ROSENZWE.MR. Precedence effect in sound localization. *Journal of the Audio Engineering Society*, 21(10):817–826, 1973.

[Woo73] A. M. Woodhull. Ionic blockage of sodium channels in nerve. *J. Gen. Physiol.*, 61, 1973.

[Wu98] S. H. Wu. Synaptic excitation in the dorsal nucleus of the lateral lemniscus. *Progress in Neurobiology*, 57:375–375, 1998.

[YG97a] X. Yang and D. W. Grantham. Cross-spectral and temporal factors in the precedence effect: discrimination suppression of the lag sound in free-field. *J. Acoust. Soc. Am.*, 102(5 Pt 1):2973–2983, 1997.

[YG97b] X. Yang and D. W. Grantham. Echo suppression and discrimination suppression aspects of the precedence effect. *Percept. Psychophys.*, 59(7):1108–1117, 1997.

[YGL01] T. N. Yensen, R. A. Goubran, and I. Lambadaris. Synthetic stereo acoustic echo cancellation structure for multiple participant voip conferences. *Ieee Transactions on Speech and Audio Processing*, 9(2):168–174, 2001.

[Yin94] T. C. Yin. Physiological correlates of the precedence effect and summing localization in the inferior colliculus of the cat. *J. Neurosci.*, 14(9):5170–5186, 1994.

[YLP96] L. C. Yang, Q. Liu, and G. D. Pollak. Afferent connections to the dorsal nucleus of the lateral lemniscus of the mustache bat: Evidence for two functional subdivisions. *Journal of Comparative Neurology*, 373(4):575–592, 1996.

[YP94a] L. C. Yang and G. D. Pollak. Binaural inhibition in the dorsal nucleus of the lateral lemniscus of the mustache bat affects responses for multiple sounds. *Auditory Neuroscience*, 1:1–17, 1994.

[YP94b] L. C. Yang and G. D. Pollak. Gaba and glycine have different effects on monaural response properties in the dorsal nucleus of the lateral lemniscus of the moustache bat. *Journal of Neurophysiology*, 71(6):2014–2024, 1994.

[YP94c] L. C. Yang and G. D. Pollak. The roles of gabaergic and glycinergic inhibition on binaural processing in the dorsal nucleus of the lateral lemniscus of the moustache bat. *Journal of Neurophysiology*, 71(6):1999–2013, 1994.

[YP97] L. C. Yang and G. D. Pollak. Differential response properties to amplitude modulated signals in the dorsal nucleus of the lateral lemniscus of the mustache bat and the roles of gabaergic inhibition. *Journal of Neurophysiology*, 77(1):324–340, 1997.

[YP98] L. C. Yang and G. D. Pollak. Features of ipsilaterally evoked inhibition in the dorsal nucleus of the lateral lemniscus. *Hearing Research*, 122(1-2):125–141, 1998.

[YPR92] L. C. Yang, G. D. Pollak, and C. RESLER. Gabaergic circuits sharpen tuning curves and modify response properties in the moustache bat inferior colliculus. *Journal of Neurophysiology*, 68(5):1760–1774, 1992.

[Zah96] T. P. Zahn. *Neuronale Architektur zur on line Spektralanalyse akustischer Signale.* Dipl. ing., Technical University of Ilmenau, 1996.

[Zen94] P. Zenner. *Hoeren - Physiologie,Biochemie,Zell- und Neurobiologie.* Georg Thieme Verlag Stuttgart * New York, 1994.

[ZF99] E. Zwicker and H. Fastl. *Psycoacoustics. Facts and Models.* Springer, Heidelberg, New York, 2'nd edition, 1999.

[ZIT97] T. P. Zahn, R. Izak, and K. Trott. Mixed analog-digital neurochip for acoustical attention. In *Jahrestagung.* Gesellschaft für Informatik, 1997.

[ZITP97] T. P. Zahn, R. Izak, K. Trott, and P. Paschke. A paced analog silicon model of auditory attention. In L. S. Smith and A. Hamilton, editors, *Neuromorphic Systems - Engineering Silicon from Neurobiology*, pages 99–112. World Scientific, 1997.

[ZLKW98] D. X. Zhang, L. Li, J. B. Kelly, and S. H. Wu. Gabaergic projections from the lateral lemniscus to the inferior colliculus of the rat. *Hearing Research*, 117(1-2):1–12, 1998.

[Zur92] J. M. Zurada. *Artificial Neural Systems.* West Publishing Company, 1992.

Index

Theses

on the Dissertation of Thomas P. Zahn

Neural Architecture for Echo Suppression during Sound Source Localization based on Spiking Neural Cell Models

1. Biological principles of echo suppression can be applied to technical systems in order to achieve stable localization and tracking of static and dynamic sound sources in reverberating rooms.

2. The specifically developed model of the auditory brain stem is capable of suppressing the direction of echoes during sound source localization tasks in real world and especially reverberating environments.

3. Spatio-temporal interaction of single excitatory and inhibitory spikes is the main information parameter in the auditory system

4. Spike Interaction Models (SIM), modeling the time course of intrinsic cell parameters and communication exclusively via spatio-temporal spike pattern, are able duplicate the precise firing pattern of auditory neurons.

5. Persistent inhibition in specific high frequency neurons of the Dorsal Nucleus of the Lateral Lemniscus (DNLL), is caused by strong and fast hyperpolarization combined with weaker and slower depolarization due the push-pull effect in the auditory brainstem.

6. Persistent inhibition in the DNLL contributes to the generation of the psycho-acoustic precedence effect in the Inferior Colliculus (IC) and higher auditory centers.

7. Persistent Inhibition exists in the DNLL of the Mongolian Gerbil and is a common feature of vertebrates.

8. The elements of the developed Neural Base Library (NBL) based on the dynamic simulation system MATLAB/SIMULINK are capable to sufficiently model the firing behavior of living neurons including the persistent inhibition effect.

9. The Elements of the NBL can be implemented as analog-digital circuits in silicon hardware and realize large neural networks, operating in real time mode fully in parallel

10. Biological information processing demands a new pardigm of computation using the dimensions of space, time and the dynamics of analog potentials as major parameters of data processing.