Development and Objective Evaluation of an EAS Cochlear Implant Model

Attila Fráter

June 3, 2014

AALBORG UNIVERSITET

Institute of Electronic Systems
Electronics and Information Technology
Fredrik Bajers vej 7
9220 Aalborg East
Phone 99 40 86 00
http://es.aau.dk

Title:
Development and Objective Evaluation of an EAS Cochlear Implant Model

Theme:
Auditory system modeling
Cochler Implants

Projectperiod:
P10 - Springsemester 2014

Projectgroup:
1061

Participants:
Attila Fráter

Supervisor:
Christian Sejer Pedersen
Søren Kamaric Riis
Jonathan Laudanski
Patrick Maas

Number of printed copies: 5
Number of pages: 81
Appendices: 4
Finished on: June 3, 2014

Abstract:

The project at hand attempts to implement a model capable of simulating combined electric and acoustic stimulation (EAS) in cochlear implant (CI) users. To model acoustic stimulation of human auditory system, the Matlab Auditory Periphery (MAP) model, developed by the Hearing Research Lab at the University of Essex, is used by adjustments to mimic a hearing loss. For modeling electric stimulation a CI model provided by Neurelec/Oticon Medical is used. The Goldwyn point process framework serves as a common platform for combination of the acoustic and electric part. The combination is based on the assumption of superposition of neural firing intensities evoked by the separate parts. The complete model is evaluated by the help of the neurogram similarity index measure (NSIM) and problems occurring during EAS (e.g. temporal desynchronization and place-mismatch) are investigated.

Abstract

The project at hand attempts to implement a model capable of simulating combined electric and acoustic stimulation (EAS) in cochlear implant (CI) users. To model acoustic stimulation of human auditory system, the Matlab Auditory Periphery (MAP) model, developed by the Hearing Research Lab at the University of Essex, is used by adjustments to mimic a hearing loss. For modeling electric stimulation a CI model provided by Neurelec/Oticon Medical is used. The Goldwyn point process framework serves as a common platform for combination of the acoustic and electric part. The combination is based on the assumption of superposition of neural firing intensities evoked by the separate parts. The complete model is evaluated by the help of the neurogram similarity index measure (NSIM) and problems occurring during EAS (e.g. temporal desynchronization and place-mismatch) are investigated.

Contents

1 Introduction

The first experiments on electrical excitation of the human auditory system were conducted by Alessandro Volta in the 1800s [1] when he concluded that hearing sensation can be elicited by electric stimulation. After a long break, in the 1970s the first cochlear implants (CI) using single electrode were implanted [2]. Despite their simplicity some patient showed benefit of using these devices in their daily life. In the 1980s the first multielectrode arrays were accepted for implantation [2]. These electrode arrays were capable of using the tonotopic organization of the cochlea to provide a better hearing sensation. Since then many speech processing strategies were developed to improve CI performance.

The most recent surgical techniques make it possible to preserve hearing in the low frequency range [1]. This residual hearing helps to improve speech understanding [1][3]. Electroacoustic stimulation (EAS) based CIs tends to have a less strict candidation criteria compared to regular CIs. This results in an increase of patients who can be candidates for CIs [4]. As an example, elder people not benefiting from casual hearing aids, but not classified for regular CIs use, can be potential users in the future.

Until recently CIs and hearing aids (used for amplifying sound for the residual hearing) were developed separately. This results in the suboptimal usage of their combination [5]. To develop devices optimized for combined electric and acoustic stimulation, the nature of EAS has to be investigated.

This project aims to develop a model simulating EAS in CI users. Additionally the possibilities of an objective method for fitting the acoustic and electric part can be investigated. Moreover an optimal stimulation, by means of frequency allocation and temporal synchronization of CI and hearing aid, can be proposed.

2 | Human hearing

This part first gives a theoretical overview of the human auditory system. It is followed by a section discussing human sound perception for electrical stimulation, hearing aids and cochlear implants and thoughts about the combination of electric and acoustic hearing seals the chapter.

2.1 Human auditory system

It is a complex way until the pressure variation generated by a sound source transforms to sensation at the human brain. The sound travels through a medium before it reaches the outer ear. The waves are collected by the pinna and directed through the ear canal, finally arriving at the tympanic membrane. In the middle ear, vibrations of the membrane are conducted by ossicles to the oval window located at the cochlea in the inner ear. This causes motion of fluid and the basilar membrane inside the cochlea. These mechanical changes are transferred to chemical changes by the inner hair cells and eventually stimulate the connecting auditory nerves. These stimuli are lead through various brain levels (cochlear nucleus, superior olivatory nucleus, inferior colliculus and medial geniculate nucleus) to finally reach the central auditory cortex where they are perceived as sound.

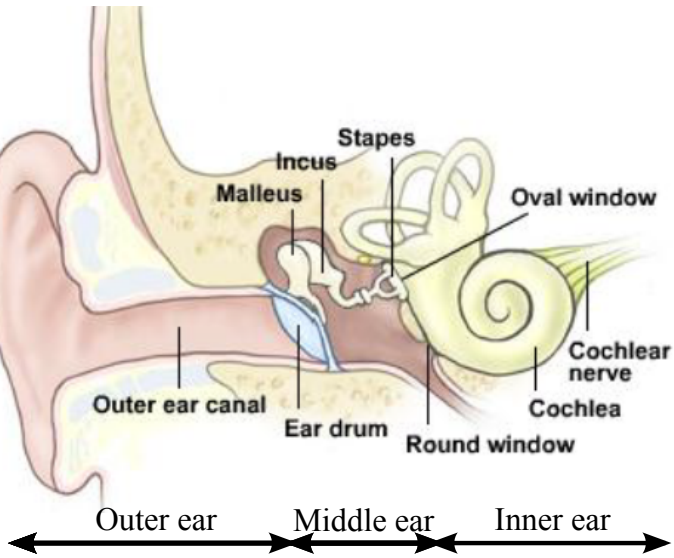

Figure 2.1: Physiology of human ear [6]

Each of the aforementioned parts of the auditory system contributes to the way in which the actual sound is perceived. The main parts are depicted in Figure 2.1 and these are described in details in the following sections, relying mainly on the work of Abbas[7].

2.1.1 Outer Ear

By the time sound arrives at the observer, it has already gone through several changes. These can be related to the sound source itself, surroundings around the source and properties of the medium in which it is situated in. Environmental properties, such as surfaces and obstacles causing reflections and absorptions as well as temperature, humidity, wind, etc. have a great influence on the characteristics of the sound that reaches the ear. The listed phenomenons are important for sound perception, but are not part of the auditory system, so they are out of the scope for the present project thus not discussed further.

Individual's body properties contribute to spectral and temporal alternations of the incoming sound. These mostly support sound localization and due to their direct effect on the input to the ear canal it is necessary to include them in any auditory model. The shape and absorption properties of the

torso and the head have significant influence on the frequency characteristics of the sound reaching the ear canal. The effect of these body parts can be expressed by the head related transfer functions (HRTFs). HRTFs are the transfer functions from the sound source to each ears with reference to the sound to the position in the middle of the head without the listener being present.

The head is also responsible for the interaural time and intensity differences (ITD and IID) being the main cues for sound localization besides HRTFs.[8]. ITD is the difference in time needed for the sound to reach the observers ears and IID is the difference in sound intensity measured at the ears.

Before the sound arrives at the tympanic membrane, it has to travel through the ear canal that introduces further resonances in the spectrum of the signal. Finally the pressure fluctuation at the end of the ear canal causes the tympanic membrane to vibrate. From this point the sound signal continues its way in the middle ear as vibration.

2.1.2 Middle Ear

As it is depicted in Figure 2.2 the middle ear confines three ossicles - malleus, incus, stapes - and two muscles - tensor tympani muscle and stapedius muscle.

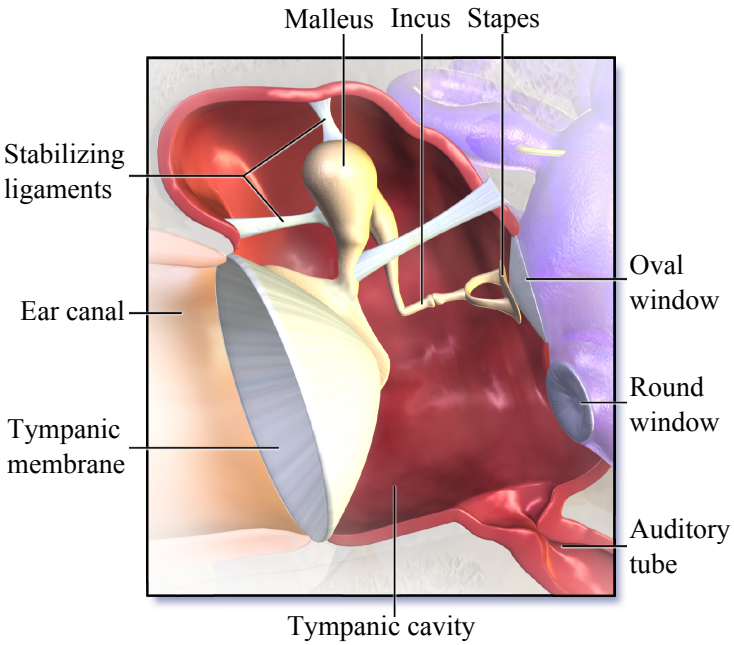

Figure 2.2: Physiology of the middle ear
http://en.wikipedia.org/wiki/File:Blausen_0330_EarAnatomy_MiddleEar.png

This mechanical system is responsible for transforming energy from the air to fluid in the inner ear, i.e impedance matching.[9] The transformation is necessary due to the difference in characteristic impedance between air and fluid. Since the impedance of the fluid inside the cochlea is much higher than the impedance of the air at the external ear, a direct transition of energy would be inefficient. This means that the pressure has to be increased at the oval window. According to Abbas [7], the middle ear achieves this increase in three different ways. First, there is a difference in the effective vibrating area of the tympanic membrane and the area of the stapes footplate. Secondly, as the malleus and incus acts as a unit, this structure can be considered as a single lever, meaning that the movement of the incus is less than the movement of the malleus and thus resulting in a higher force, hence more pressure. The third phenomena, called the curved membrane effect is also a leverage to the system, introduced by the fact that some parts of the tympanic membrane vibrates more compared to the connection to the malleus. Due to these properties of the middle ear a pressure increase is occurring at the oval window, but at the same time it results in lower vibration velocity too. This means that middle ear system behaves as a passive system, that is like a transformer.

The tensor tympani and stapedius muscle are responsible for active changes in the middle ear's transfer function. A sound event can result in the contraction of the stapedius muscle, increasing the stiffness of the system. It is known as the acoustic reflex, and is most dominant in the low frequency range and has a relatively high threshold of 70dBSPL for pure tones [10]. Many suggested theories are available for describing the role of the acoustic reflex. One hypothesis is a protecting role, but due to the long (35 to 150ms [7]) latency this can not be applied for impulsive noises. An other role can be a defensive mechanism against self-generated noises (e.g. muscle contraction, breathing)[9].

2.1.3 Inner Ear

The inner ear consists of two organs, the semi circular canals and the cochlea. The first is responsible for balancing, while the latter one functions as microphone by converting vibrations to electrical pulses [9].

The cochlea can be divided into three parts by three main canals, the scala vestibuli, scala media (cochlear duct) and the scala tympani. The first two canals are separated by the vestibular membrane and the basilar membrane separates the latter two. All three parts are filled with different fluids, endolymph can be found in the scala media and perilymph fills the other two canals.

The organ of corti as it is shown in Figure 2.3 is located within the scala media and comprises the basilar membrane, tectorial membrane, inner and outer hair cells, besides this the place of auditory nerve contacts.

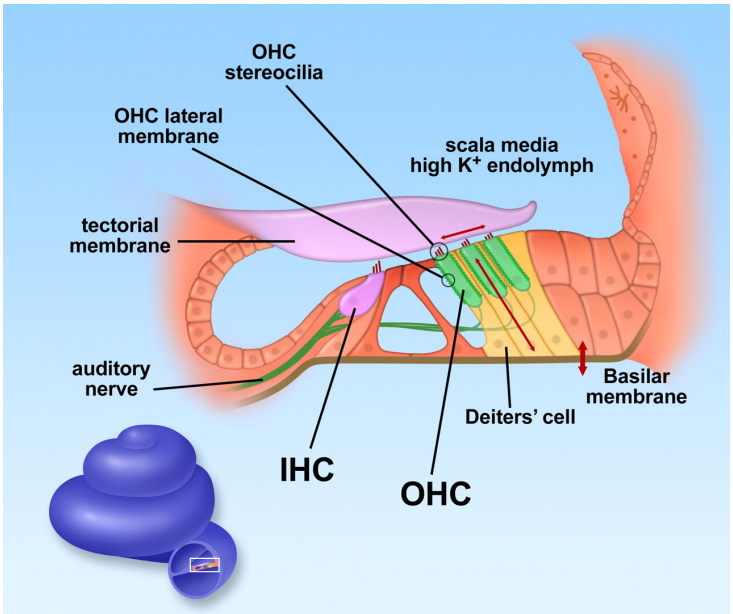

Figure 2.3: Organ of Corti
http://physrev.physiology.org/content/physrev/88/1/173/F1.large.jpg

Basilar Membrane

The stapes directly contacts the cochlea through the oval window that is an opening to the scala vestibuli. Its vibration travels further in the form of pressure changes in the endolymph and perilymph resulting in displacements along the basilar membrane (BM). The pressure in the fluids enters the scala tympani throuh the helicoterma and it is finally released at the round window.

The membrane has a shape that narrows from the apex to the base (i.e. narrow at the oval window and wide at the other end) and has a varying compliance by means of being stiff at the base and loose at the apex [9].

The basic understanding of BM movements were grounded by Békésy György in the 1960s. His stroboscopic illumination technique based experiments to observe the motion of the membrane were made on cadaver ears and at high stimulus levels (130 dB SPL [7]). These experiments described a traveling wave behavior i.e. a vibration according to the input signal's frequency travels along the membrane with a base to apex direction and with different amplitudes at different positions. The point with the highest response is

characterized by the input frequency and thus it is called the characteristic frequency of that particular point. Due to this property the BM can be considered as the Fourier analyzer of the auditory system. The relation between the characteristic frequency and the location is described by a species dependent function published by Greenwood [11]:

$$F = A(10^{ax} - k) \tag{2.1}$$

Where F is the characteristic frequency of the position defined by x in mm (x=0 corresponds to the most apical position). A is a constant of 165.4 for human and it is used for obtaining F in Hz. The constant a arises from the critical-band function and has a value of 0.06 when x is expressed in millimeters. The integration constant k is 1 for humans.

The envelope of the traveling wave shows asymmetry, it grows gradually until the peak and decreases rapidly afterward [7]. This phenomena results in the fact that while a low frequency input stimulates a wide range of the membrane, a high frequency input result in excitation concentrated more at the basal part.

The wavelength of the traveling wave decreases towards the apex as it is depicted in Figure 2.4. Consequently basal parts are vibrating in phase while at the apical section it changes over distance, yielding a diminishing traveling wave speed in the direction of the apex [7].

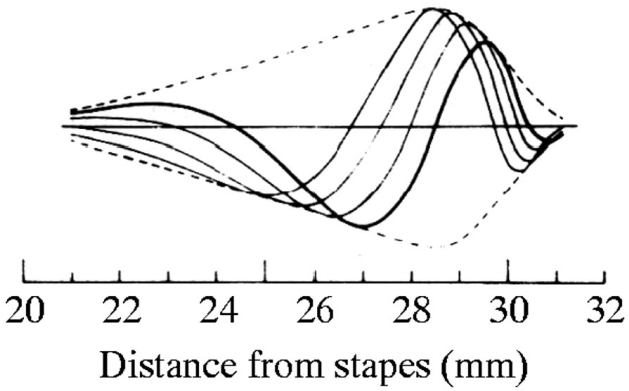

Figure 2.4: Traveling wave motion of the basilar membrane
http://jeb.biologists.org/content/208/1/157/F1.large.jpg

The frequency selectivity of the BM and the asymmetry of the traveling wave's envelope are consistent with the shape and compliance of the membrane. Indeed the direction of the traveling wave is determined by the gradient of the compliance. These properties would suggest the modeling of this organ as a simple resonance system [7], but due to the interaction of the BM and the present fluid (i.e. fluid pressure results in membrane displacement and membrane displacement causes fluid flow), this simple representation is not possible.

To obtain information about the dynamic behavior of the BM more sophisticated experimental methods had to be evolved compared to the ones used by Békésy. The Mossbauer and the capacitive probe or interferometer techniques are capable of recording much samller movements thus suitable for experiments at low input levels as well [7].

Studies relying on the aforementioned techniques showed a nonlinear characteristics of the BM's input-output function. A significant response is apparent even at very low stimulus levels e.g. several nm-s at 20dBSPL input [7] and the characteristic is linear in this region (approximately until 20-30dBSPL [9]). In the following region the function is compressed and above 90dBSPL [9] it becomes linear again. The latter phenomena is the result of the active mechanisms not being able to contribute an attenuation anymore [9]. In addition to the non-linear input-output function, the possibility of conducting measurements at different input stimulus levels also revealed a level dependent frequency tuning of the membrane. The frequency response is more highly tuned for low level stimuli compared to responses for high input levels [7].

Hair Cells

The hair cells are located at the organ of corti, between the basilar and the tectorial membrane as it can be seen in Figure 2.3. This part of the auditory system is responsible for an active amplification process within the cochlea. Despite the similar general strucure (all hair cells consist of the cell body and stereocilias on its top), two main types, the inner hair cells (IHC) and the outer hair cells (OHC) can be distinguished due to their differences in morphology and function.

Approximately 3500 [9] IHCs are organized in a row close to the base of

the tectorial membrane, but without their stereocilia being connected to it. These sensory cells are innervated by the afferent fibers, thus most of the information is conducted to the brain by the IHCs.

12000 OHCs [9] construct three to five rows further away from the tectorial membrane base clearly separated from the IHCs and with their tallest stereocilia connecting the membrane. Despite the lack of contact to the afferent fibers [7] they also contribute to the fiber response, meaning that the two hair cell systems are in interaction. Furthermore, OHCs are also responsible for an active feedback mechanism increasing the BM response for low level input stimuli. This active mechanism is caused by the ability of OHCs to produce motile response to electric potential and while it enhances low level signals, it shows saturation at high levels.

IHCs have an intracellular potential of approximately -35mV [7] and the surrounding endolymph has a potential of 80-90mV [7] resulting in a resting membrane potential of 120-150mV [7] that causes a steady state current in the stereocilia. Motion of the BM results in the deflection of stereocilia through a viscous coupling. When shorter stereocilias bend towards the taller ones, current increases through the cell membrane yielding an increase in receptor potential. Similarly, current and receptor potential decreases when the direction of deflection is the opposite. Hair cell's sensitivity to stereocilia displacement is in the range of 100nm [7]. These changes around the steady state potential are caused by the change of conductance that is the result of K^+ ion flow. Individual hair cells show saturation in potential change for high level input stimulus and a low-pass filter characteristic also describes their behavior i.e. a lower response potential is produced for frequencies higher than the best frequency[1] of the cell. Besides the nonlinear amplitude characteristics, a high degree of frequency selectivity is also an important property of the hair cells. The tuning of the cells mainly reflects the ones observed at the BM, but some inherent properties such as height and mechanical properties of the sterocilia varying along the cochlea also contributes to the tuning of the sensory cells.

[1] the frequency generating the greatest BM displacement at hearing threshold at the position

Auditory Nerve

The signal path from hair cells to the brain is established by the auditory nerves (AN) connecting to the cells through a synapse. An AN fiber consist of neurons having three main parts; the dendrite, that collects the input signal, the soma being the main part of the neuron and the axon, that conducts the information to other neurons.

Change in a hair cell's receptor potential results in change of calcium concentration within the cell. As the potential increases, the calcium concentration near the hair cell-auditory nerve synapse also increases. When it is high enough, vesicles of transmitter are released into the synapse, that elicits an action potential in the AN. Action potentials can be interpreted as spikes and are the basic elements of information taken to the brain. The process of an action potential traveling through an AN fiber is usually referred as the firing of the fiber.

Fibers transferring information towards the brain are called afferent fibers and can be categorized due to their property of showing firing activity without the presence of a sound. Fibers firing less than a half spike per second are low spontaneous rate (LSR) fibers. Medium and high spontaneous rate (MSR, HSR) fibers have firing rates of 0.5-18 and above 18 spikes/second respectively. These types also show differences in threshold, dynamic range and population. LSR fibers represent 16% [9]of ANs and possess the highest threshold up to 80dBSPL [9] with the largest dynamic range starting from 50dBSPL [12]. 61% [9] of the fibers are HSR, being the most sensitive with a threshold as low as 0dBSPL [9] and having the narrowest dynamic range between 10-20dBSPL [12]. The rest of the fibers (23%) are MSR with intermediate properties.

The frequency threshold or tuning curve shows the firing threshold of a nerve with respect to frequency. The frequency with the lowest threshold is the characteristic frequency (CF) of the fiber and it is determined by the place of innervation [7]. Frequency selectivity of AN fibers are similar to the BM's and hair cells', that is the tuning curves are asymmetric and nerves with a higher CF have a longer tail in the curve towards the low frequencies. The level of firing rate increases with sound intensity until a saturation is reached[2] and so does its frequency range (the tuning curve is getting wider) that is

[2]Some fiber show decrease in firing rate in spite of saturetion [7].

consistent with the behavior described for the BM and the hair cells.

AN fibers have a phase locking ability i.e. action potentials tends to occur at a certain phase of the input signal. Similarly to frequency selectivity, phase locking is also level dependent and becomes stronger with higher levels. Regarding input signal frequency, the strength of phase locking starts to decrease above 6-800Hz and ceases around 4000Hz [7].

Due to synaptic transmission properties adaptation behavior in firing rates of a neuron can be observed. A typical time course of firing can be seen in th lowest panel of Figure 3.8. An overshoot is present at the onset of the sound that is followed by a slow approximately exponential decrease to a steady state level. A short period, called refractory period, of firing level below spontaneous activity follows the signal's offset, before it again reaches the spontaneous level. This adaptation process is missing from hair cell responses, thus it can be explained by chemical changes within the synapse.

Fibers transferring information from the brain to the hair cells are called efferent fibers. Cells with synapse at IHCs have bodies at the lateral part of superior olivary complex, while cells with synapse at OHCs have bodies at the medial superior olive [7]. The latter effect is referred as Medial Olivo Cochlear (MOC) attenuation by Meddis [10]. Efferent fibers are responsible for reducing the response of the afferent ones and show similar tuning properties with the same frequency to position map. Their effect has a long latency of 5-50ms[7].

Brain stages

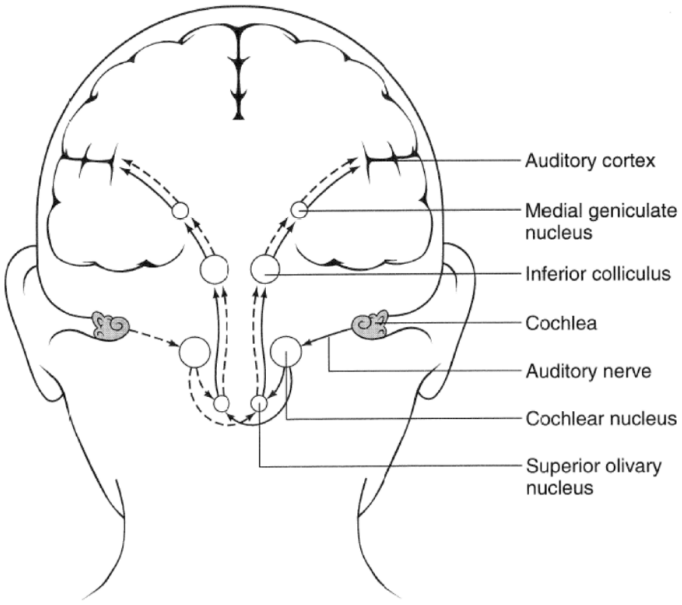

Figure 2.5: Brain stages of the auditory system [13]

The signal goes through several brain stages before it arrives to the auditory cortex. Figure 2.5 illustares the signal path in the brain. All stages show some form of tonotopic organization and tuning similar to the AN. ANs terminate at the cochlear nucleus from where information is directed towards the superior olivary nucleus. From this point signals from both ears are processed that is binaural processing takes place also at the inferior colliculus, the medial geniculate nucleus and the auditory cortex. These parts include cells that are excited from both ears (EE cells) and cells that being excited from the input of one ear and inhibited by the input from the other (EI cells). Due to differences in EE and EI cells the superior olivary nucleus is sensitive to interaural intensity and time differences so it is considered as an important part for sound localization [7]. Besides EE and EI cells, the auditory cortex also includes cells responding only to binaural atimulation [7].

2.2 Electric hearing

Sensation of hearing can be elicited by direct electrical stimulation of ANs. In this situation the parts of auditory system before the ANs are bypassed and this comes with several consequences.

The most significant difference compared to normal hearing appears in the highly reduced dynamic range. Usually the dynamic range of electrical hearing is defined by the barriers of barely audible and uncomfortable or extremely loud levels [2]. Acoustic hearing possess a wide dynamic range of approximately 120 dBSPL in contrast to the electric, that has a range as narrow as 10-20 dB [1]. This reduction is also present at speech levels, that varies in a 30dBSPL range corresponding to a 5 dB electric range [2]. This huge reduction is originated in the loss of cochlear compression.

An other deviation is the loss of tuning caused by the active (OHCs) and passive (BM) mechanisms present during acoustic stimulation. The tuning of electrical hearing is completely determined by the electrical field inside the cochlea [1].

At last the loss of stochastic nature of AN firing have to be mentioned. Due to bypassing the hair cells and the AN synapse, being responsible for the stochastic behavior of neuron firing, electric stimulation elicits a highly synchronized firing pattern [1].

2.3 Hearing aids and cochlear implants

Hearing disorders can be categorized by time course, location in the ear, cause and severity[14]. Regarding time course, the disorder can be hereditary, congenital, acquired pre- or post-lingual. According to location in the ear it can be conductive or sensorineural that can be categorized further into cochlear and retrocochlear hearing losses (HL). A conductive HL can be originated in the outer or middle ear while sensorineural HL are related to the inner ear (cochlear HL is in connection with cochlea and the origin of retrocochlear HL is beyond it in the auditory system). HL among others

can be caused by high noise levels, acoustic trauma or simply appears due to aging. The severity of HL can be ranked by the raise in hearing threshold in decibels with the base being the hearing threshold in free field of a normal hearing 18-year-old, that is the definition of dB hearing level (dB_{HL})[14]. From 25 to 40 dB_{HL} the loss is considered mild, from 40 to 70 dB_{HL} it is moderate, severe HL is between 70 and 90 dB_{HL} and a hearing loss is marked as profound when it lies between 90 and 110 dB_{HL} [14].

Within this section the various hearing aid solutions are summarized with a greater emphasis and a more thorough discussion on cochlear implants.

2.3.1 Hearing aid solutions

Hearing aids

In general a hearing aid consists of a receiver microphone an amplifier with a gain control and a loudspeaker. Recent devices are built around a digital signal processor, that provides a better adjustment that is needed during the fitting the aid to the users HL. The basis of the fitting process is usually a loudness model, but as a first approximation the so called half gain rule can be applied. The half gain rule defines the applicable gains in dB_{HL} for the input sound as the half of the HL values in each frequency band of the client's audiogram.

The three main types of hearing aids are the behind the ear (BTE), in the ear (ITE) and completely in the ear canal (CIC) types. The main purpose of all types is the restoration of speech intelligibility for the user that can have a sensorineural HL in the mild to profound range.

Bone anchored hearing aids

Bone anchored hearing devices bypass the outer and middle ear by conducting vibrations to the cochlea through the skull. The instrument is attached directly to the skull behind the ear and uses a high power vibrator for transmitting the collected sound waves.

These devices also aim for reproducing normal levels of speech intelligibility and can be applied for mild to moderate conductive HL e.g. congenital atresia (lack of outer ear/ear canal), otosclerosis (fixation of middle ear ossicles) or any middle ear malfunction.

Middle ear implants

Similarly to bone anchored hearing devices, middle ear implants also elicit vibrations to excite the cochlea. In comparison to the latter device, in this case the transducer is implanted and the vibrators directly excite the middle ear ossicles or the round window.

Middle ear implants are applicable to patients with moderate to severe sensorineural or conductive HL.

Cochlear implants

Cochlear implants are used in cases when the user has profound sensorineural HL or a complete deafness. The device directly excites the auditory nerve by electric stimulation that is going to be explained more in details in the following section.

2.3.2 Cochlear implants

As it is mentioned in the previous section cochlear implants (CI) are used when a sensorineural HL is in the profound range or the patient is completely deaf. These cases can occur when hair cells are partly or completely missing from the auditory system. As it is described in subsection 2.1.3, hair cell are responsible for the active amplification within the inner ear and beyond this stage the signal is conducted in the form of an electrical pulse by the ANs. CIs rely on the latter phenomena and stimulates the nerves directly with electrical pulses thus replacing the IHCs.

Parts of the cochlear implant

In Figure 2.6 a general CI setup can be observed. The external parts consist of a behind the ear device that usually contains a microphone and a speech processor and its signal is lead to a transmitter coil. This coil is attached to the head by a magnet that holds it in front of the receiver thus a transcutaneous connection is realized. The receiver confines a coil and a decoder that is responsible for transforming the signal to electrical current that is conducted to the electrodes. The electrodes enter the scala tympani [1] through the round window or the cochlestomy that is an artificially drilled hole near the round window.

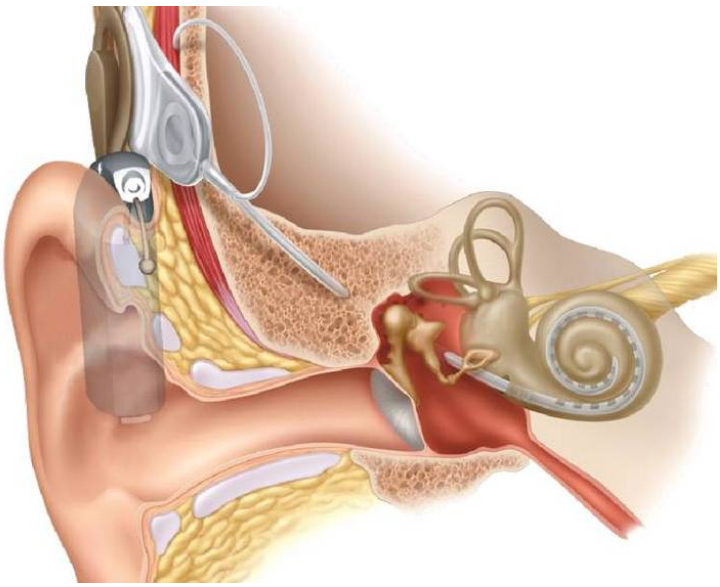

Figure 2.6: Parts of the cochlear implant system; a - behind the ear speech processor, b - transmitter coil, c - receiver, d - elecrode array
picture from: http://stvincentsent.com.au/index.php/otology-neurotology-and-lateral-skull-base-surgery/cochlear-implant-the-bionic-ear/

CI candidacy

According to Moctezuma [6] candidates for implantation can be divided into three main categories, post-lingually deaf adults, pre-lingually deaf children and post-lingually impaired people. The achievable auditory performance i.e.

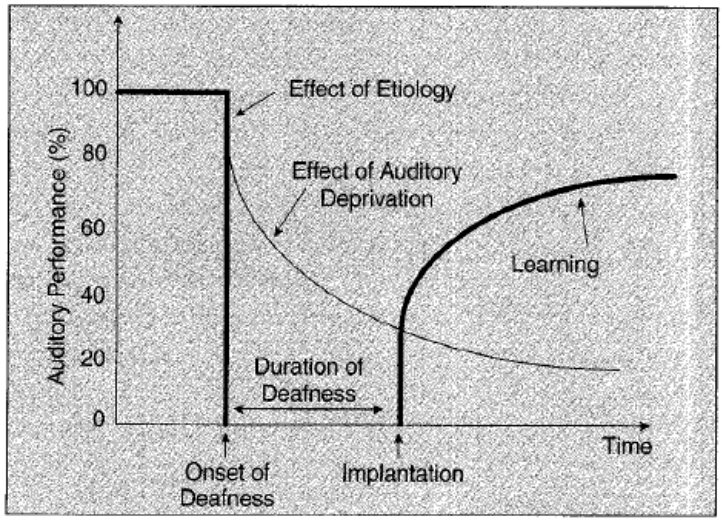

Figure 2.7: Three-stage model of auditory performance of post-lingually deafened CI users [2]

speech recognition after implantation depends on several factors [2], such as duration of deafness before implantation, age at the onset of deafness, age at implantation, duration of CI use. In general the less the duration of deafness, the better the performance of post-lingually implanted patients and they also perform better compared to pre-lingually deafened candidates. Regarding the latter type of patients it is advised to start to use a device as early as possible. Improvement in speech recognition is apparent immediately after surgery and the performance tends to increase over years of use. Figure 2.7 illustrates a three-stage model of auditory performance of post-lingually deafened CI users [2]. Besides the aforementioned factors others, like, patients attitude towards CI, physiological state of the cochlea and ANs, etc. also influence the outcome of implantation. Furthermore, speech processing technique applied by the instrument also has a great effect.

Signalprocessing techniques used in cochlear implants

The first implants in a 1970s and 1980s used only a single electrode for stimulation that was fed by a band-pass filtered and compressed signal being transmitted through a modulator/demodulator[3] transcutaneous connection.

[3]In the 1970s the device made by House/3M was even implanted without a demodulator, thus a in spite of the speech, its modulated signal was transmitted to the electrode[2].

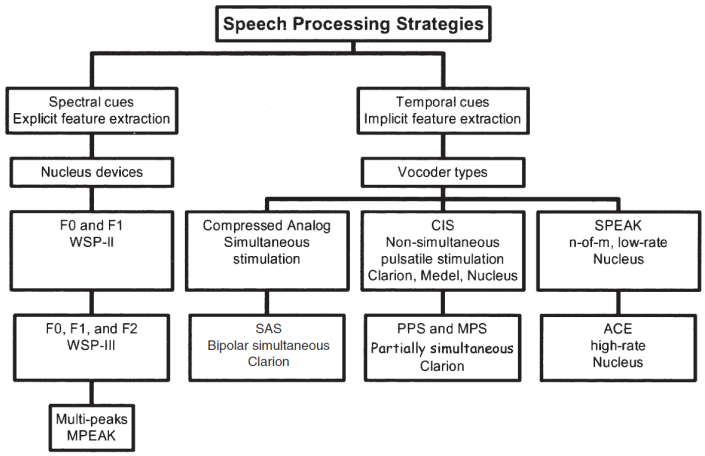

Figure 2.8: Signal processing strategies in cochlear implants[1].

These instruments can only take advantage of the temporal encoding of the speech since the electrode only excites a single location of the cochlea. Due to the refractory of the AN this can contain information up to approximately 1000Hz that is just a small part of the speech frequency range. Despite their simplicity and limited capability of speech encoding, some patient could benefit from these devices [2].

To overcome the limited frequency range of the single electrode implants, CIs with multiple electrodes were developed which can exploit the tonotopic organization of ANs. Figure 2.8 summarizes the different strategies applied by multiple electrode CIs.

All strategies must use a compression at a point to compensate for the loss of cochlear comression mentioned in section 2.2. Besides the common features of exploiting the tonotopic organization of the ANs and the usage of compression, evolved signal processing strategies can be discriminated by used stimulus type or the preserved sound information. By applied stimulus type CIs can be divided into two main groups, one using analog and the other using pulsatile stimulation. According to Zeng [1], strategis can aim to preserve spectral features, wave or envelope information.

The first developed strategy for preserving spectral information of speech is the F0/F2 strategy and uses pulsatile stimulation. The name referes to the formants extracted from the speech signal by zero crossing detection within a frequency band, i.e F0 is the fundamental frequency (below 270 Hz [2]) and

F2 is the second formant (1-4kHz [2]). The extracted fundamental frequency F0 determines the rate of stimulation[4] at the electrode corresponding to extracted second formant F2. F0/F1/F2 applies a completely similar approach but also extracts the first formant F1 in the frequency band of 280-1000Hz to obtain more low frequency information [2]. The most recent MPEAK (Multiple Peak) spectral feature extracting strategy in addition to the latter ones, also uses envelope detection at the high frequency range (2-2.8kHz, 2.8-4kHz and 4-6kHz bands [2]).

The Compressed Analog (CA) strategy starts the series of strategies preserving wave information by using analog stimulation. After separating the sound into distinct bands and applying a compression, the signal is sent directly to the electrodes simultaneously. This simultaneous activation of electrodes is the main drawback of this method due to the channel interaction i.e. electrical fields generated by adjacent electrodes influence eachother. Simultaneous Analog Strategy (SAS) tries to solve the problem of electrode interaction by using neighbouring electrodes in pairs.

The Continuous Interleaved Sampling (CIS) method extracts envelope information of the band-passed sound signal and presents it to the corresponding electrodes as pules with a constant rate. Electrodes are activated in a non-simultaneous fashion , that is, only one electrode is active at a time yielding a mitigation of the electrode interaction problem. The oredr in which the electrodes are used can vary according to patient preferences. The Paired Pulsatile Sampler (PPS) and the Multiple Pulsatile Stimulation (MPS) develop the CIS strategy further by allowing a pair or multiple distant electrodes to work at the same time. Another branch of signal processing techniques preserving envelope information uses an approach similar to MPEAK. The n of m or Spectral Peak (SPEAK) strategy procsses the input signal in m distinct freuency bands and stimulates n electrodes according to the highest observed amlitudes. The stimulation is pulsatile with a rate depending on the number of selected electrodes. The Advanced Combination Encoder (ACE) method follows the idea of SPEAK with using higher stimulation rates.

[4] During unvoiced sounds a random stimulation rate is used.

2.4 Electro Acoustic Stimulation

Recent improvements in surgical techniques make it possible to preserve residual hearing after CI implantation even with long electrode arrays reaching the residual hearing region [15]. This achievement offers patients the benefit of a combined electric and acoustic stimulation (EAS) when a CI and a hearing aid is used at the same ear.

The preserved hearing is usually restricted to the low frequency range. Usami [15] reports postoperative hearing loss form 30 dBHL to 110 dBHL in the 125 Hz to 1000 Hz region. In Figure 2.9 the area of hearing loss in which EAS can be used is represented by a gray color. It can be seen, that typical EAS candidates can have a hearing loss up to 60 dBHL in the low frequency range and must have a loss greater than 80 dBHL for the high frequency range. A normal hearing aid is used in cases with milder hearing loss and standard CI is implanted for the more severe ones.

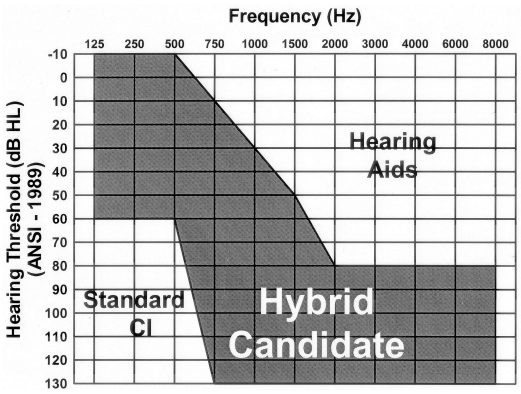

Figure 2.9: Auditory profile of EAS candidates [4]

2.4.1 EAS benefits

By taking advantage of the residual hearing EAS users show better performance than using CI or a regular hearing aid only [16]. Vaerenberg [3] also showed a great improvement of speech recognition in noise compared to using CI only.

These improvements are mainly associated with the better representation of temporal fine structure (TFS) cues. These cues are responsible for identifying small temporal variations in the sound and [17]. CI users can only use TFS cues up to 300 Hz [1], while this limit can be extended by the additional acoustic stimulation.

2.4.2 Challanges of EAS implementation

Due to the nature of the combined stimulation two main issues have to be taken into account. Temporal synchronization of electric and acoustic stimulation and placement of the electrode array compared to the actual place coding within the cochlea have to be fit.

Temporal difference arises from the different time that is needed for electrical and acoustical stimulation of the cochlea. CI has a certain processing time, then the signal is directly sent to the ANs. In comparison during acoustic stimulation, besides the hearing aid having a processing time, the sound also needs time to travel through the outer and middle ear and the final excitation point of the BM depends on the actual frequency as well. The most common solution is to delay the faster device to match the timing and obtain simultaneous stimulation [5].

The second problem is usually referred as place-mismatch. It means, that the frequencies assigned to the electrodes may differ from the ones that are actually at the place where the electrode lies in the cochlea.

A frequently investigated question is the cross-over frequency of the electric and acoustic stimulation i.e. which range should be covered by the electric stimulation for a given hearing loss. Incerti [16] concludes that patients with severe hearing loss can benefit from overlapping ranges, while patients possessing better hearing may perceive an echo and a distorted sound quality, thus the most common solution is separating the two ranges. This means that electric stimulation is only done for frequencies without residual hearing.

3 Implementation

3.1 Auditory system modeling

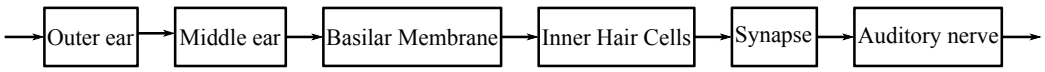

Figure 3.1: Block diagram of the applied part of Matlab Auditory Periphery model

For modeling the auditory system, the Matlab Auditory Periphery (MAP) model developed by the Hearing Research Lab at the University of Essex is used. A description of model stages depicted in Figure 3.1 is presented in the following sections. The corresponding parameters can be found in [10].

In general the model consists of a series of stages representing parts of the human auditory system. SI quantities are used for all parameters and the processing of the input sound (given a .wav file format) is done in 0.01 second time frames with a sampling frequency adapted from the input data.

3.1.1 Outer ear

The first stage of MAP takes a sound pressure wave as an input and represents the concha and ear canal resonances by band-pass filters. The output of the filters are summed and added to the direct input signal. According to section 2.1 this part can be improved by incorporating HRTFs and the band-pass amplification of the ear canal can also be improved.

3.1.2 Middle ear

The input of this stage is the sound pressure at the tympanic membrane. To fit human stapes displacement measurements first a scalar is applied to convert pressure to velocity. It is followed by a low-pass filtering (that is applied, because displacement decreases by the increase of frequency) and application of another scalar to obtain displacements accurately above 2kHz. A high-pass filter is used to introduce attenuation at the lower frequency range.

The acoustic reflex discussed in subsection 2.1.2 is introduced by a variable scalar in a negative feedback loop. This scalar is controlled by the brainstem activity in the low spontaneous rate stream (discussed in subsection 2.1.3).

3.1.3 Basilar Membrane

Stapes displacement serves as an input to this stage that models basilar membrane displacement at distinct locations by the help of a dual-resonance-non-linear (DRNL) filter. The separate locations are identified by their best frequency i.e. the frequency generating the greatest basilar membrane displacement at hearing threshold at the position [10] (same as the characteristic frequency (see 2.1.3) defined at the hearing threshold). By default these frequencies are logarithmically spaced between a user defined low and a high frequency limit.

The DRNL filter models the frequency selectivity of the auditory system and represents changes in characteristic frequency, gain and phase according the input signal level. As it can be observed in Figure 3.2, DRNL filter contains a separate linear and non-linear path.

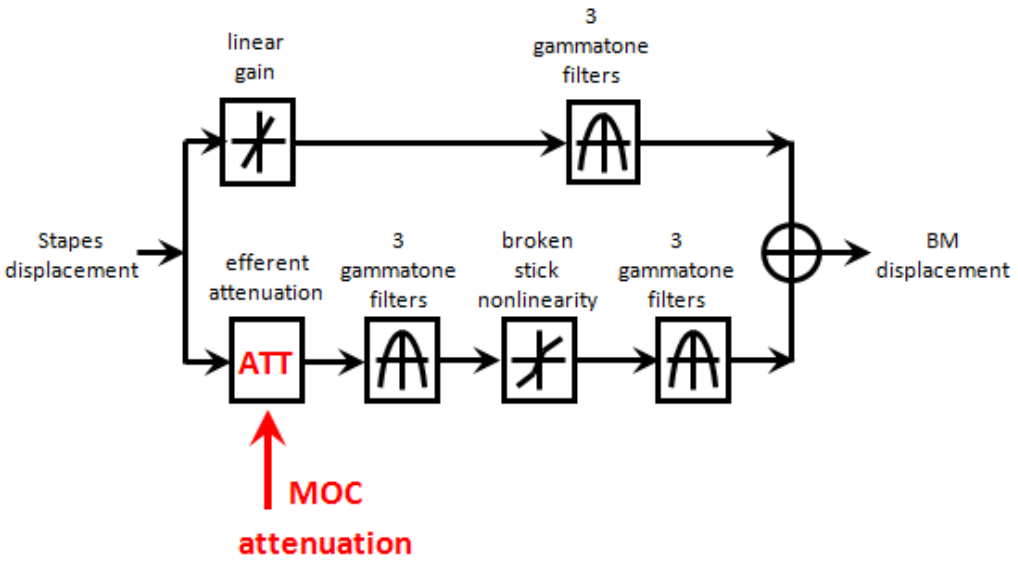

Figure 3.2: Schematic of the dual-resonance-non-linear filter [10].

In the non-linear path the signal is first attenuated by a variable scalar gain for representing the MOC reflex discussed in subsection 2.1.3. This scalar is controlled by the sum of all firing of all types (LSR, MSR, HSR) of second level neurons in a best frequency channel [10]. After the attenuation the signal is lead through a cascade of gammatone filters with center frequencies identical to the best frequencies and bandwidth increasing with the best frequency. This is followed by an input level dependent compression and finally the gammatone filters described before are applied again.

The linear path consists of a scalar gain and a cascade of gammatone filters with a characteristics similar to the ones in the non-linear path, but with different center frequencies.

The two separate paths are summed together to produce basilar membrane displacement values at the corresponding positions as an output of this stage. A more detailed description of the DRNL filter can be found in Appendix A.

This implementation is proven to model accurately the iso-intensity curves obtained from experimental data [18]. This means that the physiological behavior (e.g. non-linear input-output function, level dependent tuning) of the basilar membrane is represented properly.

3.1.4 Inner hair cell

This stage of the model produces receptor potential changes in the IHCs according to the actual BM displacements. The process can be divided into to two consecutive sections. First the conductance change of stereocilia can be obtained from the BM displacement by modeling the viscous coupling between them. After this, receptor potential change can be derived from the conductance change by applying an electric circuit model of the IHC.

The work by Shamma [19] serves as a basis for this implementation, but changes according to the usage of BM displacements in spite of BM velocities are incorporated.

Conductance change

The stereocilia of an IHC is not connected to the tectorial membrane, thus an indirect viscous coupling is realized between the basilar membrane displacement $disp_t$ and the cilia displacement u(t) that can be written as[10]:

$$\tau_c \frac{\mathrm{d}u(t)}{\mathrm{d}t} + u(t) = \tau_c C_{cilia} disp_t \tag{3.1}$$

Where C_{cilia} is a scalar and τ_c is a time constant. The derivation of Equation 3.1 can be found in Appendix B. MAP model implements this function as a high-pass filter and a scalar multiplication.

By being fit to empirical data, that shows level dependent asymmetry[1], Equation 3.2 describes the relation between the stereocilia displacement and the apical conductance change G(u)[10]:

$$G(u) = G_{cilia}^{max} \left[1 + e^{-\frac{u(t)-u_0}{s_0}} \left[1 + e^{-\frac{u(t)-u_1}{s_1}}\right]\right]^{-1} + G_a \tag{3.2}$$

[1]For high input levels, changes in conductance are much larger for positive (when smaller stereocilias deflect towards the taller ones), than for negative cilia displacement. In contrast, conductance change is similar in both directions for low input levels.

Where G_{cilia}^{max} is the maximum of apical conductance that is responsible for introducing saturation in the conductance for high stimulus levels, to be consistent with experimental data. G_a is the passive conductance of the apical membrane. The constants s_0, u_0, s_1 and u_1 are adjusted to describe the exact shape of the empirical data.

Receptor potential change

Change in receptor potential V(t) can be modeled by a passive electric circuit depicted in Figure 3.3 and its transfer function is as follows[10]:

$$C_m \frac{\mathrm{d}V(t)}{\mathrm{d}t} + G(u)(V(t) - E_t) + G_k(V(t) - E'_k) = 0 \tag{3.3}$$

Where $C_m = C_a + C_b$ is the cell capacitance, G_k is the membrane conductance, E_t is the endocochlear potential and $E'_k = E_k + E_t R_{pc}$ is the reversal potential of the ionic current (dominated by K^+[19])at the basal membrane corrected for the resistance of the supporting cells $R_{pc} = \frac{R_p}{R_t + R_p}$.

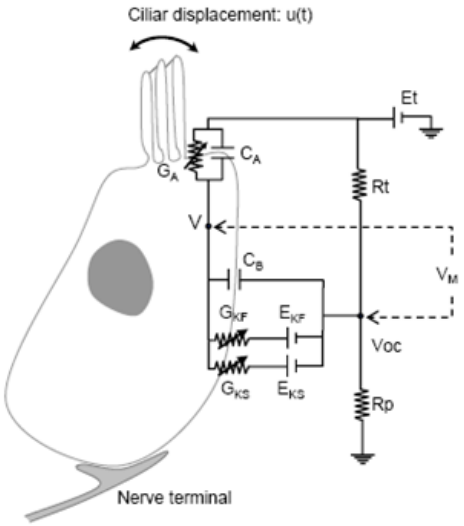

Figure 3.3: Passive electric circuit model of IHC [10]

3.1.5 Inner hair cell / Auditory nerve synapse

As it is discussed in subsection 2.1.3 the firing of AN is directly related to the transmitter release into the IHC-AN synapse. The transmitter release is controlled by the calcium concentration within the hair cell, which is related to the receptor potential.

The MAP system first models the calcium influx that is in direct relation with the receptor potential. From the calcium influx, the calcium concentration can be obtained. After the calcium concentration is available, it serves as an input to a probabilistic model that is responsible for simulating transmitter release to the synapse.

Calcium influx

The following expression determines the relation between the calcium current I_{Ca} and the receptor potential [10]:

$$I_{Ca}(t) = G_{Ca}^{max} m_{I_{Ca}}^3 \left(V(t) - E_{Ca}\right) \tag{3.4}$$

Where G_{Ca}^{max} is the maximum calcium conductance when all calcium channels are open, $m_{I_{Ca}}$ is the ratio of calcium channels being open[2] and E_{Ca} is the reversal calcium potential.

Calcium concentration

Calcium concentration Ca^{2+} near to the synapse is calculated as follows [10]:

$$\frac{\mathrm{d}[Ca^{2+}](t)}{\mathrm{d}t} = I_{Ca}(t) - \frac{[Ca^{2+}](t)}{\tau_{Ca}} \tag{3.5}$$

Where τ_{Ca} is the dwell time of calcium before the synapse and it varies according to the spontaneous rate of the actual neuron. τ_{Ca} is represented as a row vector when more than one spontaneous rate is simulated. In this case the model splits into parts according to the neuron types and continues the processing separately. This means, that MAP does not take into account the distribution of neurons (see: subsection 2.1.3) according to different spontaneous rates.

Probabilistic model of transmitter release

Figure 3.4 shows the probabilistic model of transmitter release.

[2] The actual value of $m_{I_{Ca}}$ is the low-pass filtered function of its steady state value that is determined by a Boltzmann function.

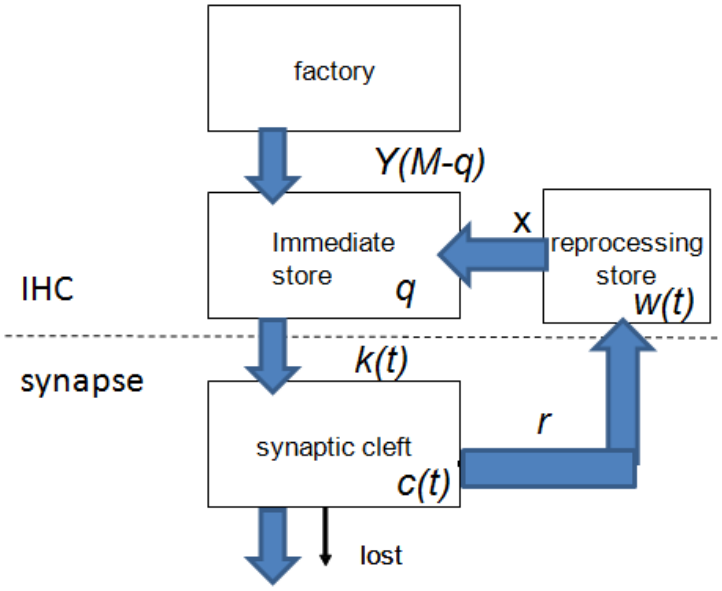

Figure 3.4: Probabilistic model of vesicle release at the IHC-AN synapse implemented in hte MAP model [10].

In Figure 3.4 q(t) represents the actual number of vesicles waiting for release in a stage referred as the immediate store, c(t) is the actual number of vesicles in the synapse and w(t) is the number of vesicles in the reprocessing store that models the reuse of released transmitters.

The transfer between the stages are represented by constants l, r, x corresponding to the rate of vesicle loss from the synapse, the rate of transmitter take back to the reprocessing store and the rate of reprocessing respectively. The rate of transmitter release rate k is expressed as [10]:

$$k(t) = z\left([Ca^{2+}]^3(t)\right) \tag{3.6}$$

Where z is a scalar for transforming calcium concentration to vesicle release rate. Finally for the rate of vesicle generation is calculated by $y(M-q)$ where M is the maximum number of vesicles that can be present in the immediate store and y is the constant of the generation rate.

According to Figure 3.4 the synapse model can be summarized in the folowing system of equations [10]:

$$\frac{\mathrm{d}q(t)}{\mathrm{d}t} = xw(t) + y\left(M - q(t)\right) - k(t)q(t) \tag{3.7}$$

$$\frac{\mathrm{d}c(t)}{\mathrm{d}t} = k(t)q(t) - lc(t) - rc(t) \tag{3.8}$$

$$\frac{\mathrm{d}w(t)}{\mathrm{d}t} = rc(t) - xw(t) \tag{3.9}$$

3.1.6 Auditory nerve

For modeling the firing of auditory nerves a probabilistic point process generation method is used. This method is the inverse distribution function technique and it is discussed further in Appendix D.

3.2 Cochlear implant model

Similarly to the auditory model, the Neurelec/Oticon Medical proprietary CI model is also built up of a series of blocks representing different parts of the device. The simulation chain takes a .wav file as an input and produces an electrodogram as an output. An electrodogram represents the activity of the electrode array for a given input signal i.e. the time course of electrical current for each electrode. An example electrodogram can be observed in Figure 3.5. The y axis represents electrode number starting from the apex, thus Eaf0 is the most apical and Eaf19 is the most basal electrode respectively. The top row in Figure 3.5 represents the frame sent to the implant.

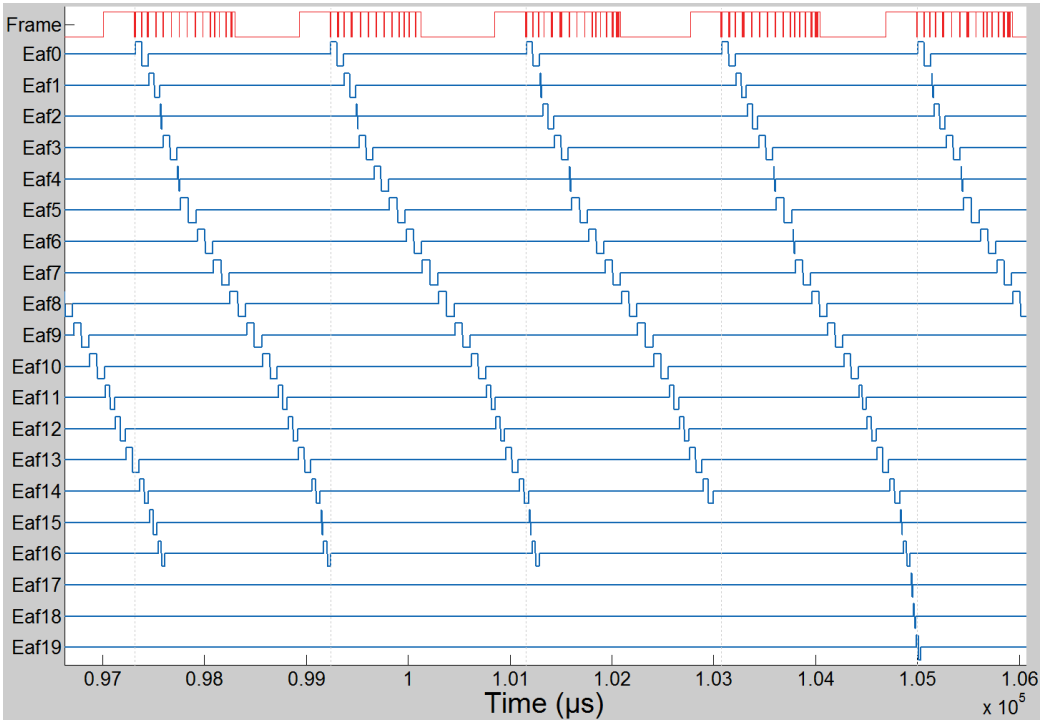

Figure 3.5: Example electrodogram

The chain consists of two main parts, the behind the ear speech processor and the implant. The schematic of the complete simulation chain is depicted in Figure 3.6. The signal types are denoted above the linking arrows between the parts. The sign .wav represent a signal being in the time domain, while .WAV denotes the signal in frequency domain. Numbers on top of the arrows shows the number of simultaneous signal paths (the lack of number indicates a single path).

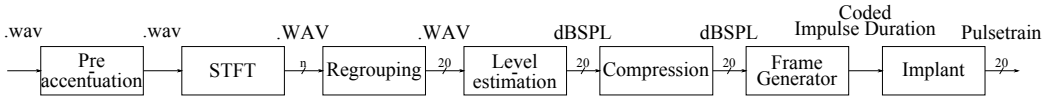

Figure 3.6: Block diagram of CI model

3.2.1 Speech processor

A wide range of speech processors is supported to be used, thus the most suitable for the desired situation can be chosen. Each different type simulates the signal processing strategy implemented in the corresponding device. Enabling certain additional features such as noise cancellation or device specific compression technique is also given.

The chosen speech processor model receives an the input signal in the form of a .wav file and filters it by a pre-accentuation filter. This filter is used for modeling the natural frequency response of the cochlea which is absent in CI users. After this, a short time Fourier transformation is used to process the signal in frequency domain further on. In the next step the frequency bands (defined by the Fourier transformation) are grouped according to the predefined frequency allocation of the electrodes. Within these frequency groups the sound pressure level is estimated and it is passed forward to a compression function. The actual values of the compressed signal is then related to impulse durations, from which a coded stimulation frame is generated.

3.2.2 Implant

Keeping in mind compatibility criteria, different implant types can be connected to the selected speech processor. The implant model is responsible for simulating the decoding of input frame (i.e. output of the speech processor stage) and the conversion to actual electrical impulses at the 20 electrodes. The output of the implant model (and also the complete CI model) consist of 20 separate vectors, containing the timing and amplitude values of the according electrode activity.

Although due to the nature of the device (i.e. can not be modified after implantation) the adjustable parameters are more restricted compared to the processor. During the current project, parameters were not adjusted after choosing the suitable implant type.

3.3 Goldwyn model

In order to obtain spike patterns as a response for electrical stimulus, a point process framework developed in [20] that is parametrized to fit reported statistics. Originally this model was developed to improve the reliability of auditory nerve response models for high stimulus rates. With this upgrade the desynchronization of auditory nerve fibers for high stimulation rates can be modeled, which is a crucial point for mimicking the spontaneous activity in the cochlea.

The statistics to which the model is adjusted includes the firing efficiency curve, chronaxie, jitter and spike history effects.

The firing efficiency curve describes the relation of the current level of an electrical pulse and the probability of nerve firing. It can be approximated by the integral of a Gaussian distribution. The threshold of a neuron θ is defined as the input stimulus level that elicits a spike with a probability of one-half, thus it is directly related to the efficiency curve. The variability in spike initiation is represented by the relative spread (RS), that is defined as the standard deviation of the Gaussian distribution (underlying the firing efficiency curve) divided by its mean.

The input pulse duration dependence of the neuron threshold (the threshold is lower for a longer pulse) is described by the chronaxie that is by definition the pulse duration at which the threshold level is twice what it would be for a much longer pulse [20]. This phenomena is caused by the ability of the neural membrane to integrate charge over time.

An additional randomness of spike timing is represented by the jitter. Jitter depends on both pulse duration and level, but for the sake of simplicity this model uses a value most commonly reported for a pulse at spiking threshold.

History effects are incorporated as transient increase in threshold and relative spread after a spike. Another implemented history related behavior is the summation effect. This represents the capacity of a neuron to integrate consecutive sub-threshold pulses for high stimulus rates (i.e. pulses with short interpulse intervals).

3.3.1 Model stages and parameters

After clarifying the statistic and point process relations, that can be found in Appendix C, a model for the conditional intensity function shown in Figure 3.7 can be built.

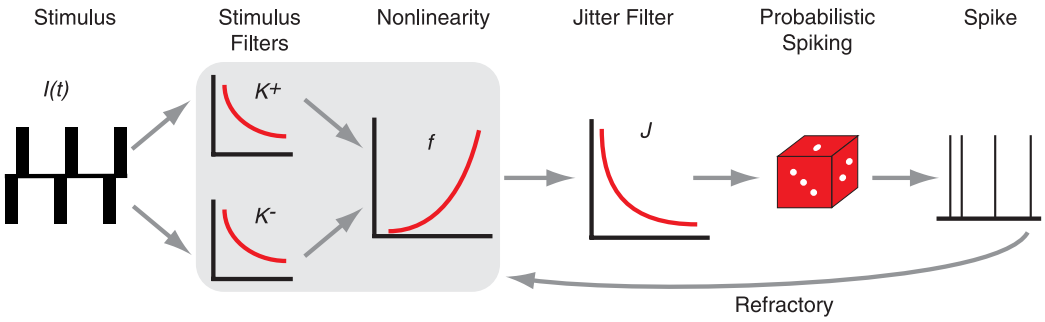

Figure 3.7: Cascade model of point process generation for electric input stimulus [20]

In Figure 3.7 it can be seen that the positive and negative parts of the input stimulus I are treated separately by stimulus filters K+(t) and K-(t). Their summed output is passed through a nonlinear function f(.) that is followed by a filter J(t) responsible for the jitter effect. The final spike train is produced by a probabilistic generator using inverse distribution function technique. By controlling the stimulus filters and the nonlinear function according to spike timing, the history effects are also included. Relying on this model the conditional intensity function is [20]:

$$\lambda(t|I,H) = [J * f(K^+ * I^+ + K^- * I^-)](t) \tag{3.10}$$

Where $*$ represents the convolution operator and each function (f(x), $K^+(t)$, $K^-(t)$ and J(t)) can be expressed by the response statistics (θ, RS, D_c, Jitter and summation time). A detailed description of these expressions can be found in Appendix C.

3.4 Model modifications and combination

In order to connect the individual acoustic and electric models within a common interface, certain modifications have to be made. Besides the necessary changes others are also made to improve the models or facilitate the problem.

3.4.1 Changes in MAP model

The MAP model is first modified at the basilar membrane part. Originally the best frequencies by which the separate locations along the membrane are identified (see subsection 2.1.3) are logarithmically spaced between a low and high frequency limit. Although the positions of BM modeling can be chosen arbitrarily, it is more natural to distribute them linearly and use the Greenwood function [11] to declare the corresponding best frequencies. This latter consideration is also necessary because as it will be described, the electrical path calculation is extended to linearly spaced locations at the BM.

The second important alternation from the original MAP model is the exclusion of efferent effects. This means that neither the acoustic reflex, nor the MOC attenuation is calculated. As it will be discussed later, the acoustic and electric models are connected before the generation of neural spike patterns. Due to this, the original neural activity reproduction of the MAP model, from which the efferent effects are calculated, is discarded thus the efferent activity can not be obtained directly. This modification of the MAP model is a facilitation. In a future work the efferent effects can be calculated from the common spike pattern of electroacoustic stimulation, but it requires a more severe modification of both the MAP and Goldwyn model (that serves as the interface for connecting the models).

A third change in the MAP model is required by the fact that for simulating real situations, an impaired ear have to be modeled in spite of a normal one, for which the MAP model is optimized in its default state. The implementation of an impaired ear requires the incorporation of several physiological considerations thus it has to be explained more in details. This explanation of the impaired ear parametrization can be found in subsection 3.4.3.

The final modifications of the MAP model are related to fiber types with different spontaneous rates. The dwell times of calcium τ_{Ca} (being directly related to spontaneous activity see: subsection 3.1.5) are corrected to have a better fit to the empirical data discussed in subsection 2.1.3. Figure 3.8 shows the instantaneous and averaged[3] firing rates of the three fiber types at a characteristic frequency corresponding to the input tone's frequency of 431 Hz. Although the MAP model is capable of modeling different types of fibers, it does not take into account their distribution, discussed in subsection 2.1.3. To overcome this deficiency the output of fibers within a CF (i.e. firing probabilities) are summarized with weightings corresponding to the their distribution. This summation can be done if firing at different fiber types are assumed to be mutually exclusive events. In this case the total probability theorem[4] can be applied [21].

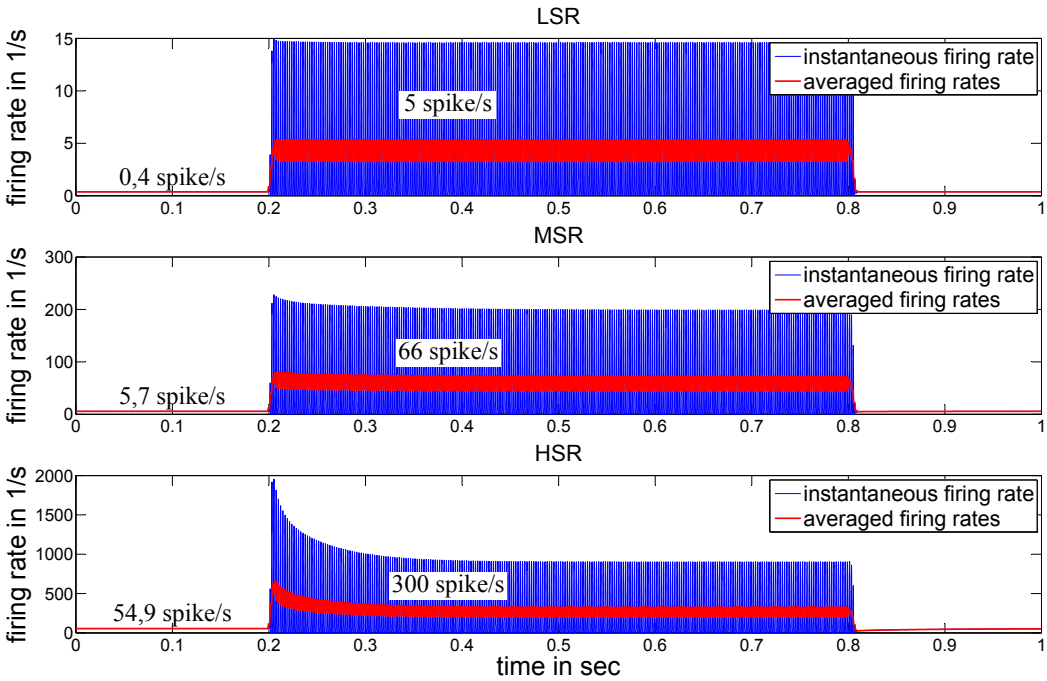

Figure 3.8: Modified firing rates for high (HSR), medium (MSR) and low spontaneous rate (LSR) fibers.

[3] a moving average is used with a time window corresponding approximately to one period of the input signal

[4] $P(B) = P(B|A_1)P(A_1) + ... + P(B|A_n)P(A_n)$ [21], where in our case $P(B)$ is the probability of firing, $P(B|A_{1...3})$ is the probability of firing at a certain fiber type and $P(A_{1...3})$ represents the distribution of the different types

3.4.2 Extension of electric model

The auditory nerve can be used as the place of combination of the electric and acoustic part, so the output of these have to be in the same form, that is firing rates of the auditory nerve.

As it is discussed in section 3.2 the output of the cochlear implant model is an electrodogram. To obtain firing rates at certain positions in the cochlea, first the effect of electrode activity on auditory nerves has to be simulated. This is done by mimicking the electrical spread caused by each pulse. Depending on the distance between the electrode and the actual auditory nerve the electrical stimulus observed by the nerve can be calculated by a predefined function. Although it is rarely the case, but for simplicity a perfect electrode placing is assumed, that is, an electrode lies above a place in the cochlea that has a CF equal to the signal's frequency delivered to the electrode.

After electric stimuli are available at the auditory nerves, the Goldwyn model (see: section 3.3) is used to convert them to firing rates.

3.4.3 Implementation of hearing impairment

Hearing loss at cochlear implant users is profound sensorineural, mostly caused by the loss of hair cells. The frequency characteristics of this loss can be specified as ski-slopes, that is residual hearing is restricted to frequencies approximately below 1000Hz and the loss shows a steep slope from 40dBHL to 100dBHL in the region of 250Hz to 1000Hz (similar hearing loss is presented in [15]). This fact can serve as a constraint for modeling, such thats representing the outer and middle ear should be kept intact and modifications should follow physiological changes occurring at patients. It has to be mentioned that damages at the level of auditory nerves (that can be present in real life situation) and beyond are not simulated.

Accorcing to Zilany [22] the loss of IHCs results in elevated tuning curves and loss of OHCs is responsible for broadening (irrespective of input levels [23]) and elevating the tuning curves, as well as reducing the nonlinear compression observed at the BM since the active mechanisms inside the cochlea are addressed to these cells [23].

According to Moore [23], the phenomena, that the BM movement excites less hair cells ,yields less neural activity in an impaired ear. This can explain the relation between elevated tuning curves and hair cell loss. Furthermore, due to its active amplification role, OHC loss is also in connection with reduced BM response for low level signals.

As a summary, the three main physiological change that should be simulated are the broadened and elevated tuning curves and the reduced compression of BM. Regarding frequency characteristics of the hearing loss, a ski-slope has to be constructed.

Parametrization for impaired ear

Bearing in mind the above mentioned aims, parameters of the BM, IHC and synapse parts of the auditory model can be changed. MAP model implements the frequency analyzer property of human hearing in the BM part. Within the nonlinear part of DRNL filter (used for modeling the traveling wave at BM, see:subsection 3.1.3) the gain can be adjusted separately for each frequency band to obtain a hearing loss resembling a ski-slope in characteristics.

The nonlinear path of DRNL filter is dominant at low input levels only (see discussion on DRNL filter behavior in Appendix A), thus changes in the linear part also have to be introduced to model hearing impairment for high input levels as well. Broadening the tuning curves can be done by reducing the order of gammatone filters applied in the both paths. These modifications of the BM part may seem controversial to the statement in the preceding paragraphs, that BM motion remains the same and the elevation and widening of tuning curves result from hair cell loss explicitly. On the other hand at the current state, the model parameters can only be set at the nonlinear path of DRNL filter with respect to frequency. Besides, the tuning and input level dependency of hair cells are similar to the BM's, thus this relation allows to change parameters at latter one. Another reasoning can be that BM motion is impeded by the presence of the electrode array within the cochlea. The BM response and the tuning curves before and after modification can be observed in Figure 3.9 and 3.10. Tuning curve elevation for the complete frequency range and broadening for high frequencies is apparent as well as the reduced compression that is an emerged property of the changed parameters. Nevertheless, the obtained graphs for absolute threshold levels in Figure 3.11

shows better hearing in the low frequency range compared to the desired ski-slope. The aforementioned graphs are obtained by a modified version of testBM and multiThreshold programs respectively. These programs are provided with the MAP model and used for evaluating it. Incorporating the modified MAP model is the only alternation from the original versions.

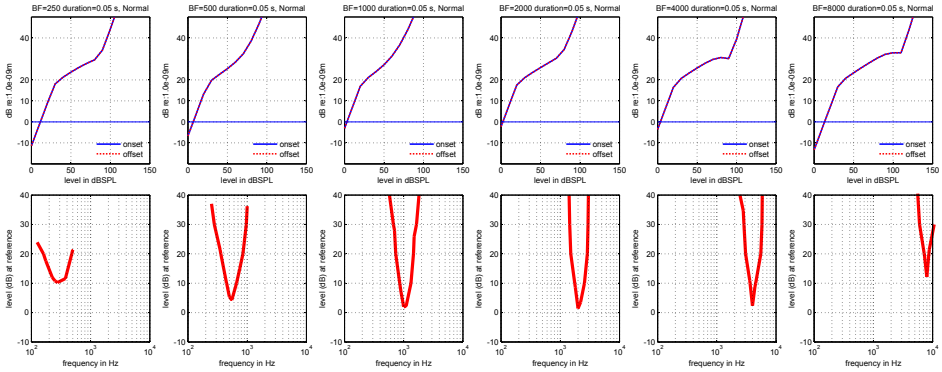

Figure 3.9: Top row - Basilar membrane response for normal hearing; Bottom row - Tuning curves for normal hearing; Data obtained by testBM program, provided with MAP

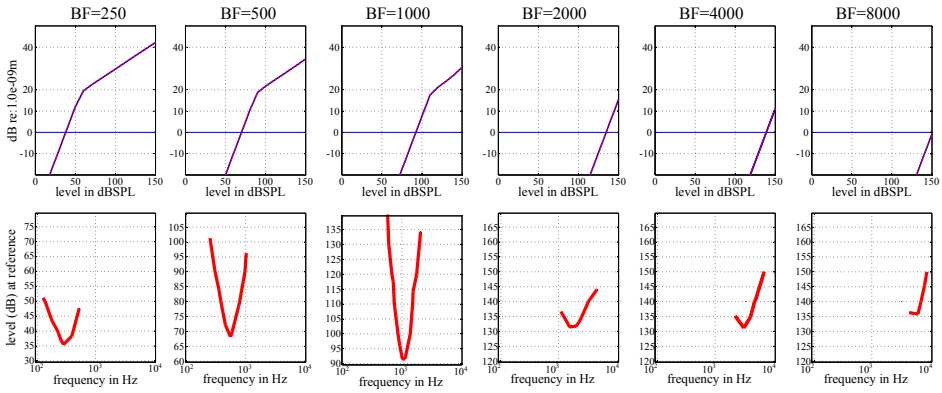

Figure 3.10: Top row - Basilar membrane response for impaired hearing; Bottom row - Tuning curves for impaired hearing; Data obtained by testBM program, provided with MAP

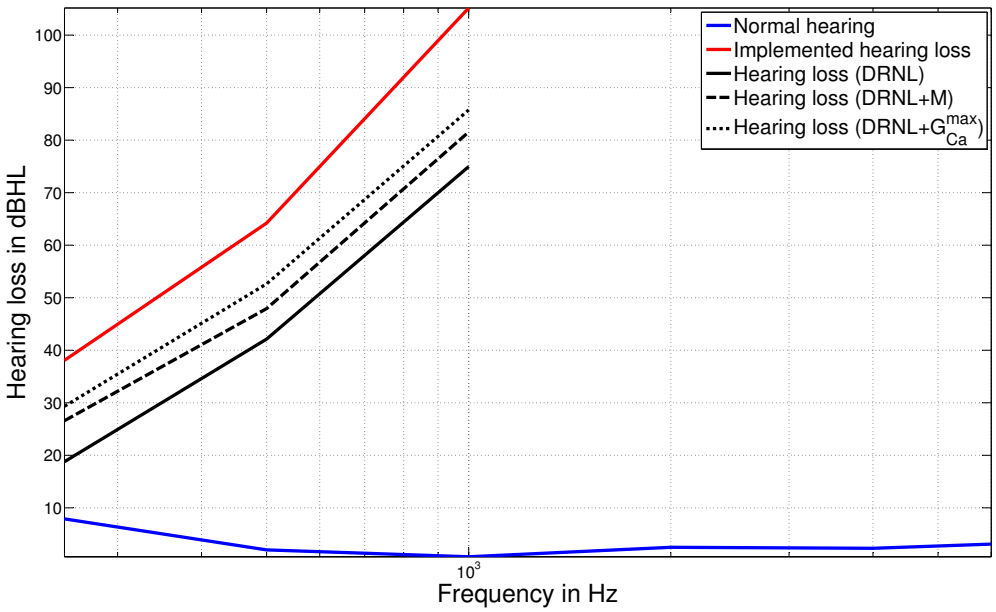

Figure 3.11: Absolute threshold levels obtained after BM modifications; Data obtaine by MultiThreshold program, provided with MAP.

To introduce further loss to the auditory model the maximum calcium conductance G_{Ca}^{max} of IHCs and the maximum number of vesicles M in the presynaptic store can be reduced. Reduction of M causes a broad band elevation of hearing threshold because it reduces the number of vesicles that can be released to the synaptic cleft at a time, thus neural activity decreases in general. Due to its nature, this change also reduces the spontaneous activity that can represent partial loss off hair cells. The indirect relation of IHC's maximum calcium conductance G_{Ca}^{max} through calcium current I_{Ca} and calcium concentration Ca^{2+} to the transmitter release rate k is shown in section 3.1. According to Sumner [24], the decrease in G_{Ca}^{max} not only increases the firing threshold of ANs, but also changes the input level dependency of firing from saturating to sloping and straight for significant decrease. Figure 3.12 depicts the change of firing curve with respect to G_{Ca}^{max} and the final absolute hearing threshold curve after application of all parameter modifications can be observed in Figure 3.11.

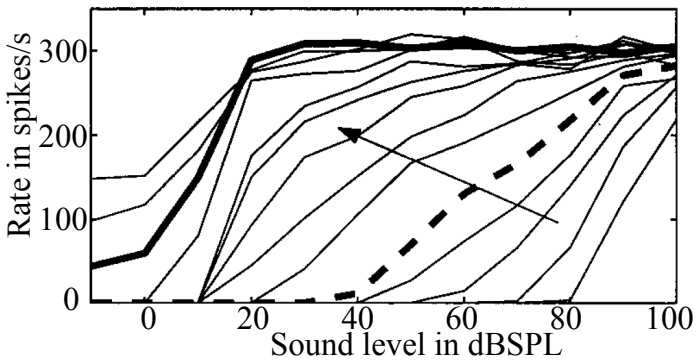

Figure 3.12: Firing curve dependency with respect to G_{Ca}^{max}. The arrow shows the direction of increasing G_{Ca}^{max}. [24]

3.4.4 Implementation of a hearing aid

A hearing aid model is built to simulate the compensation of hearing loss. As a reference the input signal level is simply raised by 50 dBSPL. With this reference hearing aid, errors introduced by more complex ones can be traced back.

Besides the reference, a hearing aid model adjusted to the given hearing loss, based on the half-gain rule introduced in subsection 2.3.1, is implemented. A 128 long finite impulse response (FIR) filter is fitted to the calculated half gains. The *fir2* Matlab function is used for generating the filter coefficients. The frequency response function of the model together with the half gains is depicted in Figure 3.13.

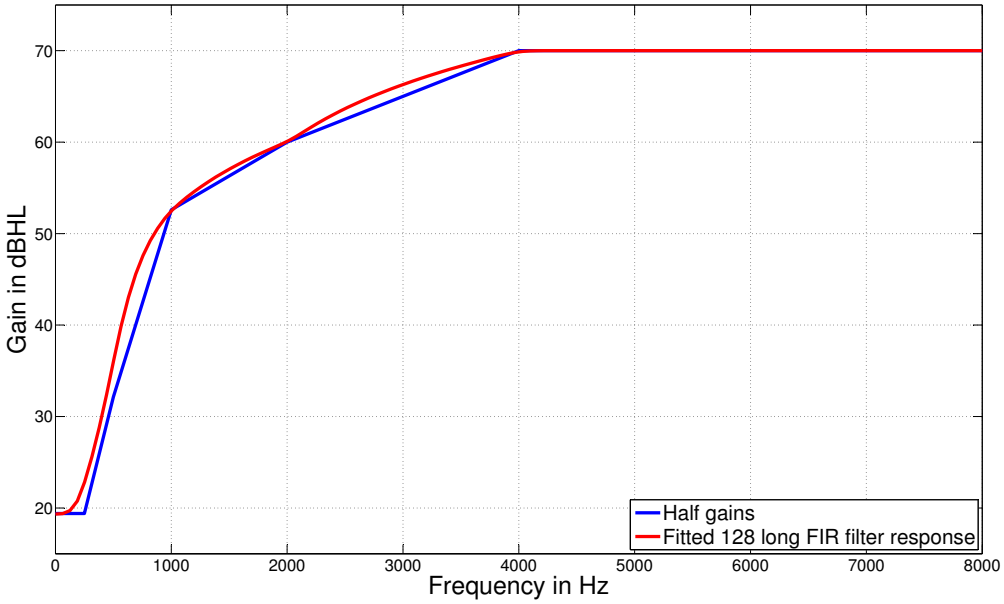

Figure 3.13: Transfer function of hearing aid (fitted with *fir2* Matlab function) and the half gains

Although the desired frequency response of the hearing aid is achieved, due to its simplicity, in the time domain it introduces a delay to the signal passing through the device. This bad design will have consequences when EAS is evaluated in chapter 5.

3.4.5 Connection of models

After modifications are implemented in the acoustic and electric model, their produced output is going to be firing intensity. After ascertaining that intensities are calculated with the same rate (Δt in Equation D.3 is the same for both cases) they are added together without calibration, since both models were adjusted to fit empirical data. From this combined output the inverse distribution function technique based probabilistic model in the Goldwyn model produces the spike pattern. Figure 3.14 depicts the technique used

for electroacoustic spike pattern generation. The usage of intensity superposition is not verified, thus the general results of this project have to be evaluated in the light of the assumption, that the superposition is considered valid. This assumption gives the possibility of a future work in which the underlying mechanisms of AN firings can be investigated during a simultaneous electric and acoustic excitation.

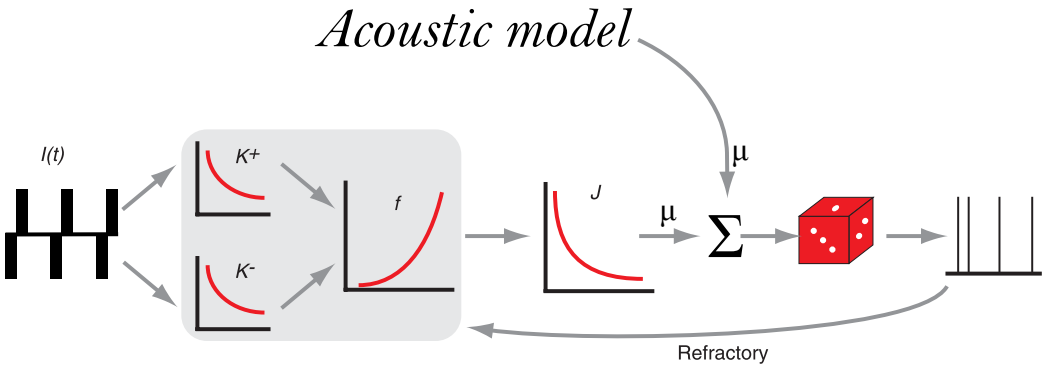

Figure 3.14: Flowchart of the modified Goldwyn model [20]; μ - firing intensity

Common input signal

To achieve correct results, a common input signal with the same amplitude have to be used for both the acoustic and electric path. Unfortunately the required format of input is different for the two parts.

The acoustic part works with quantities in SI units, thus it takes a sound wave with its amplitude in Pa as an input. The average, by means of root-mean-square (rms), level of a .wav file $input_{rms}$ can be adjusted according to a desired sound pressure level $target_{dBSPL}$ to represent the input signal as sound pressure. First the desired rms sound pressure $targert_{rms}$ is calculated from $target_{dBSPL}$ [10]:

$$targert_{rms} = p_{ref} \cdot 10^{\frac{target_{dBSPL}}{20}} \tag{3.11}$$

Where p_{ref} is the 20 μPa reference sound pressure. After $target_{rms}$ is available, the conversion of .wav file to a sound wave can be achieved as follows:

$$input_{acou,Pa} = input_{acou} \cdot \frac{target_{rms}}{input_{rms}} \tag{3.12}$$

Where $input_{acou}$ represents the vector containing the .wav file values and $input_{acou,Pa}$ is the input signal in sound pressure representation.

On the other hand the electric path requires a .wav file as an input, thus the matching can not be done outside of the model, but at the point where the level estimation is elaborated (see section 3.2). At this point the signal is converted to estimated sound pressure levels in every frequency group. To make the adjustment, first the input signal (at the same level as used for the acoustic path) is introduced to the electric model in the form of a sound wave (like it was the acoustic path). Before this signal would reach the level estimation part, the processed signal is taken out. At this point the signal is split according to the frequency groups, thus by taking the rms of each group, reference values can be obtained, that can be used for adjustment when the electric path is calculated.

$$input_{elec,Pa,adjusted} = input_{elec,Pa} \cdot \frac{target_{elec,rms}}{input_{elec,rms}} \tag{3.13}$$

Where $input_{elec,Pa}$ is the original signal in the electric path after level estimation (and being converted to Pa-s), with the rms level of $input_{elec,rms}$ for every frequency group. The $target_{elec,rms}$ is the rms obtained by the acoustic signal as described above and $input_{elec,Pa,adjusted}$ is the adjusted value of the estimated sound pressures in the electric path.

By this adjustment the usage of a same input is ensured for the distinct paths.

Figure 3.15 show the schematic block diagram of the combined model.

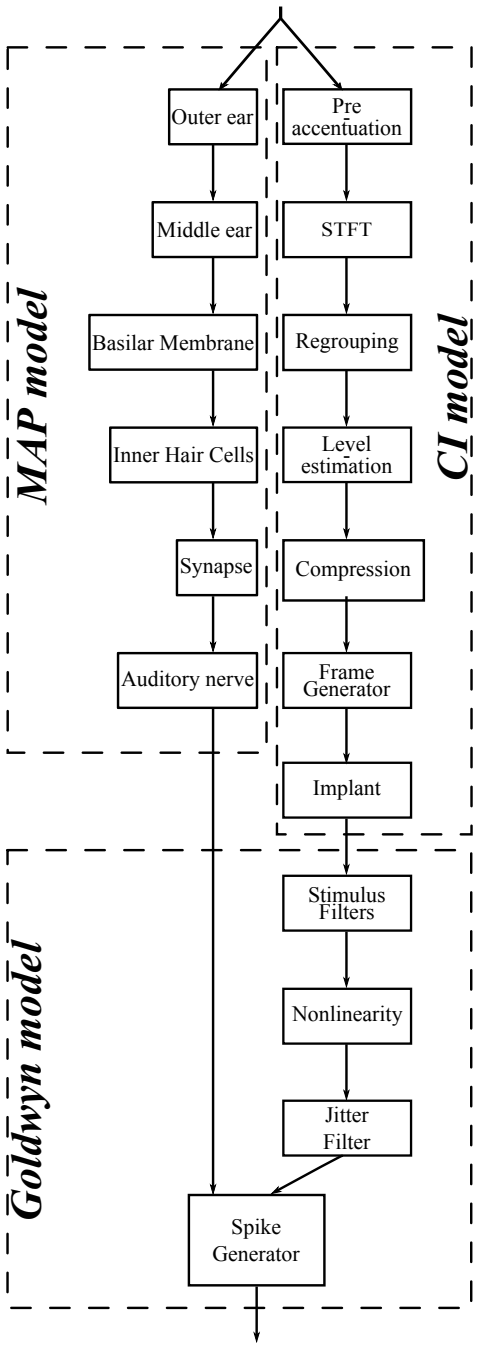

Figure 3.15: Block diagram of EAS model

4 | Validation methods

For performance evaluation of the combined electric and acoustic stimulation both objective and subjective measures can be used. By comparing the results obtained by the different measures and taking the benefit from both makes a thorough investigation available. This chapter describes the neurogram similarity index measure (NSIM) as the objective measure applied throughout the project. Development of a vocoder used for subjective listening tests is the subject of future work.

4.1 NSIM

Objective measures make it possible to compare results without any individual bias. Being able to point out differences unseen by human evaluation and the possibility of automatization are other advantages that make these measures appealing.

The auditory, electric and combined models described in chapter 3 are capable of simulating the auditory nerve spike train corresponding to various stimuli. There are several possibilities to choose from for comparing spike trains, including relative mean absolute error (RMAE), relative mean squared errors (RMSE) or cost based measures [25].

Spike patterns are built up of lines of spike trains obtained during the simulation and produced by neurons related to different frequencies along the cochlea. Spikes are occurring as black dots according to the time and in a line corresponding the neuron by which they are evoked thus as a whole it can be interpreted as black and white picture. An example of a spike pattern

is depicted in Figure 4.1. This similarity to an image allows us to use metrics originated in image processing. The chosen measure is the Neurogram Similarity Index Measure (NSIM) that is a modified version of the Structural Similarity Index Measure (SSIM) that was developed to asses perceptual image quality. NSIM is proven to be superior over RMAE and RMSE [17][26], thus after deciding on a suitable setup in terms of NSIM parameters discussed later, it is more accurate and convenient to use.

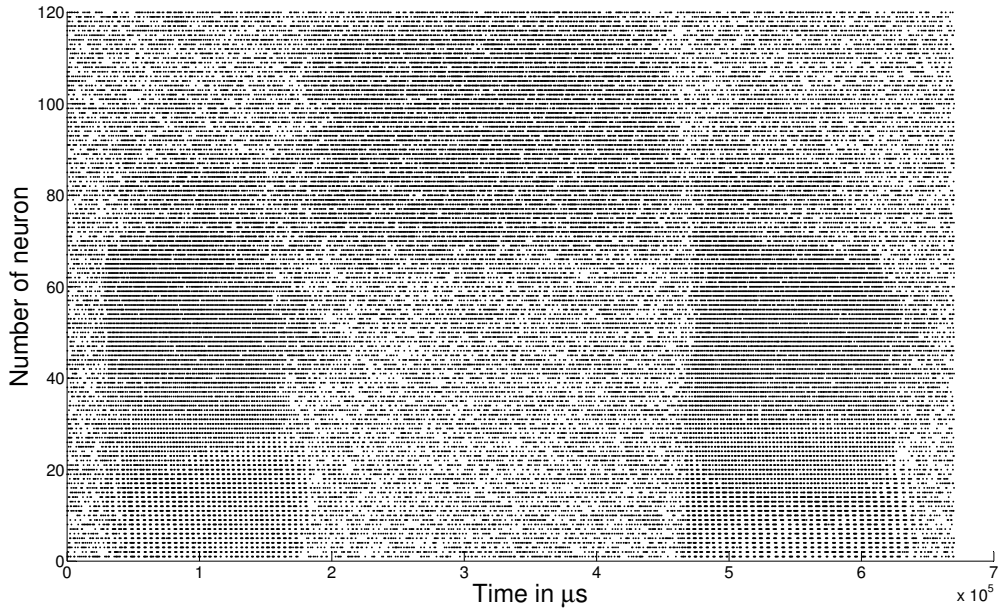

Figure 4.1: Spike pattern example. Response for 'acha' VCV word at 40dBSPL. In the y axis a lower number corresponds to a neuron with lower characteristic frequency.

4.1.1 Configuration of the similarity measure

SSIM was developed to compare images and is used to evaluate the quality degradation of pictures after compressing processes. The underlying idea of the metric is to analyze differences similarly as the human vision system would do. Due to this approach, pictures are compared by small sections and measures related to luminance, contrast and structure are used to quantify

the discrepancy[27]. The weight of the latter mentioned measures has a great influence on the behavior of the metric and have to be set depending on the actual application [28]. The result of comparison is an image, reflecting the differences between the compared pictures. By taking the mean value of the of the result a general number, showing the general similarity is obtained.

The available spike patterns are black and white images and in this form SSIM is not applicable on them, thus they need to be transformed in order to resemble more a picture. The results of the transformation are called neurograms and these are used as an input to SSIM evidently explaining the name of NSIM.

Neurograms

Neurograms are created from spike patterns line by line, following the directions given by Hines in [28]. First spikes are accumulated within a predefined time bin. This process is followed by a convolution with a 50% overlapping Hamming window with a window size given in time bins. According to the used bin and window size determining the temporal resolution of the generated picture, three main types of neurograms can be discriminated in connection with three different temporal features in speech. Envelope information (2-50Hz) is related to articulation, voicing, vowel identity and prosody of speech. Periodicity (50-500Hz) provides information of the signal's periodicity that can be related to a nasal or stop phoneme. Temporal fine structure (TFS) (600Hz-10kHz) identifies small variations of the signal and helps sound identification such as vowel formants [17]. In order to supply an adequate amount of data (in case of a sparse spike pattern the produced neurogram will not 'resemble a picture enough') either several neurons per critical frequencies or more trials have to be simulated. The used method for neurogram generation is summarized in Figure 4.2.

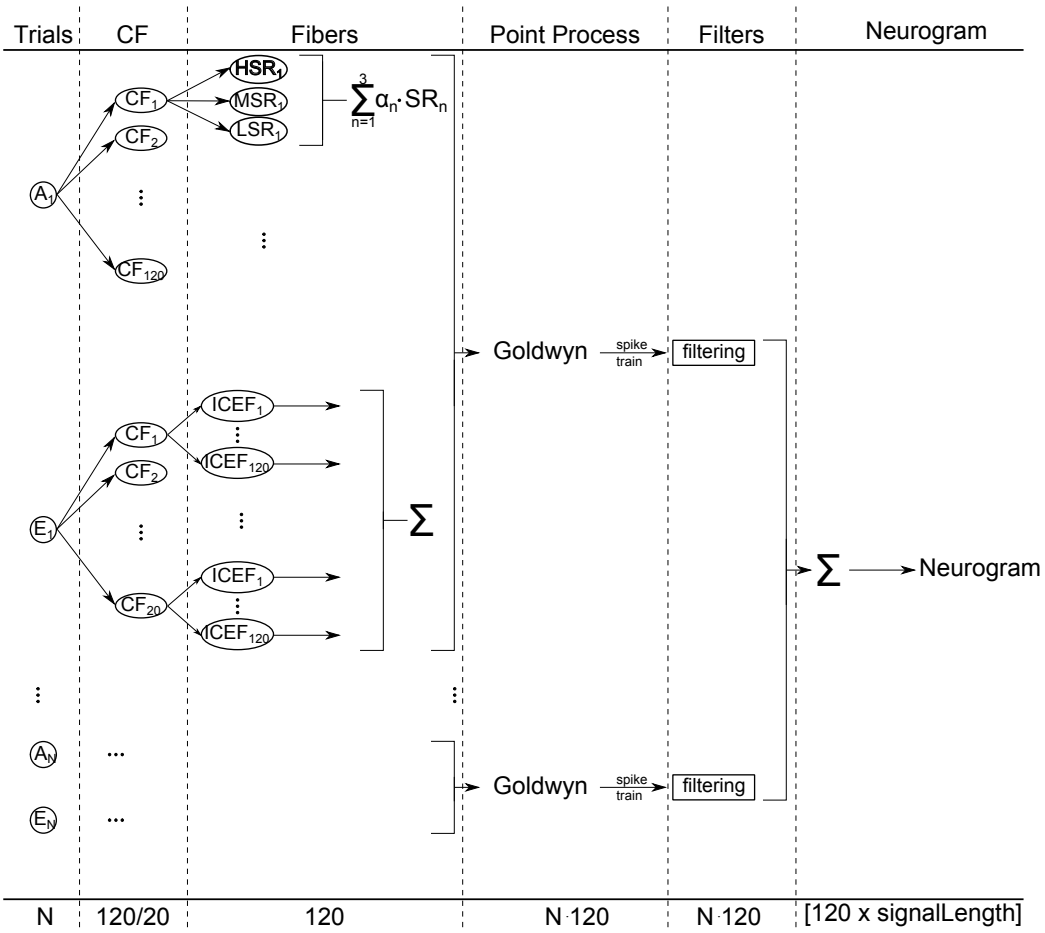

Figure 4.2: Block diagram of neurogram generation. A - Acoustic trial, E - Electric trial, N - number of trials CF - Critical Frequency, HSR - High Spontaneous Rate fiber, MSR - Medium Spontaneous Rate fiber, LSR - Low Spontaneous Rate fiber, ICEF - Intra Cochlear Electric Field

In Figure 4.2 the a line at the top indicates the stage of the process and a line at the bottom indicates the cardinality of the stage. The two parallel paths used for simulating acoustic (A) and electric (E) stimuli by the help of the modified MAP model and CI model can be observed.

The generator consists of six stages. The number of trials N i.e. repeating the whole process can be set by the user.

In the acoustic path, within a trial the input signal (a sound wave) is processed for 120 distinct location on the BM, characterized by their charac-

teristic frequency (CF). Within each location three different fiber types are simulated and their outputs (i.e. firing robabilities) are merged according to the method described in section 3.4.

In the electric path 20 CFs are representing the 20 electrodes of the implant and the intra cochlear electric field is calculated for 120 frequency points to match the acoustic path.

The two separate paths are connected within the Goldwyn model that produces 120 spike trains in each trial. Under the filtering stage the spike trains are first accumulated within a predefined time bin and then convolved with a Hamming window with a given window size. This stage produces a picture in a trial. Finally the corresponding lines obtained in every trial are added together to obtain the final image i.e. the neurogram.

NSIM parameters

After the neurograms are available SSIM can be applied on them. According to [17] a scanning window with a side length up to 5 pixels the algorithm can be used safely and a 3x3 window is proposed to yield the best results thus this size is used further on. Regarding the weights of luminance, contrast and structure an according [1 0 1] vector is used proposed to be the most suitable for phoneme discrimination [28]. These weights are adjusted such that the NSIM results would reflect listening test results thus future work to find more suitable parameters for cochlear implant simulations can be done. The function for calculating NSIM throughout the project is based on the work of Ian Bruce.

5 | Results

This chapter contains a systematic evaluation of the auditory model, the implemented impaired ear model and the electric model. These sections are followed by a thorough investigation of electroacoustic stimulation. Finally an attempt is made to find the optimal amount of electric stimulation, by means of activated electrodes, in the residual hearing range for the hearing loss at hand.

Neurograms described in chapter 4 are used for evaluating each part.

5.1 Evaluation of acoustic stimulation

First the frequency representation of the acoustic path is investigated by the help of a linear sweep signal. After this a tone burst is used for evaluating the time course of acoustic response. Investigation of a complex input signal in the form of a VCV (vowel-consonant-vowel) word finishes the section.

The auditory model with the original parameters (provided in [10]), representing normal hearing is used throughout the acoustic evaluation. The basilar membrane is represented at 120 distinct locations thus the neurogram representations contain 120 lines. To avoid difficult to read neurograms due to broad excitation of the BM a relatively low input level of 40dBSPL is used in the first two cases. When response to a VCV word is investigated, the input level is set to both 40dBSPL and 60dBSPL using the latter to simulate a general speech level.

5.1.1 Frequency course of acoustic stimulation

A one second linear sweep between frequency limits of 100Hz and 8kHz is fed to the auditory model at a level of 40dBSPL. For generating the neurogram from the spike pattern output of the acoustic model, time bin of 50 μs and a smoothing Hamming window of 32 bin length is used, resulting in a 1250Hz resolution. The procedure is repeated 100 times. The result can be observed in Figure 5.1.

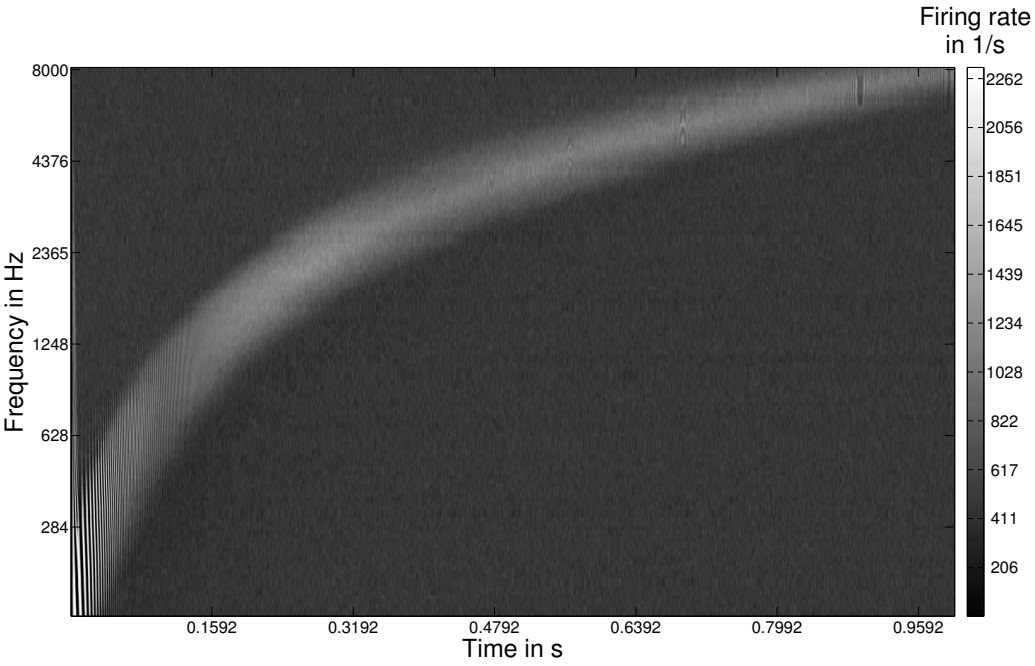

Figure 5.1: Neurogram of a linear sweep between 125 and 8000 Hz at 40 dBSPL

In Figure 5.1 the horizontal axis represents time in seconds and the vertical axis represents the neurons distributed according to the Greenwood function and identified by their corresponding frequencies. Higher firing rates are represented by brighter color, and the colorbar gives an approximation of the instantaneous firing rates. It have to be pointed out, that due to the usage of a smoothing window, it is not possible to recover the exact firing rates. Furthermore, because of the representation of instantaneous firing

rates, the maximum values tend to exceed the steady state values (see the overshoot at the beginning of the signal in Figure 3.8).

The increasing frequency of the input signal produces the white diagonal line in the figure. The nonlinear behavior of the curve is caused by the nonlinear distribution of neurons along the cochlea. The thickness of the curve (the vertical extension at a certain time) represents the tuning at the actual neuron, and it reflects well the narrowing nature of tuning curves towards higher frequencies.

The noise surrounding the actual pattern of the input signal represents the spontaneous activity of neurons.

Phase locking property of neurons is reflected by distinct vertical lines within the curve A horizontal tilt of these vertical lines can also be observed especially at the lowest frequencies i.e. at the beginning of the signal. This tilt represents the traveling wave at the BM, that is, higher frequencies are excited before lower ones.

5.1.2 Time course of acoustic stimulation

A 0.1 second 431Hz tone burst between 0.05 second silent parts is used for evaluating the acoustic path in the time domain. The 40dBSPL test signal is depicted in Figure 5.2.

For neurogram generation a time bin of 100 μs and a smoothing Hamming window of 128 bin length is used, resulting in a 156.25Hz resolution. The procedure is repeated 100 times. The result can be observed in Figure 5.3.

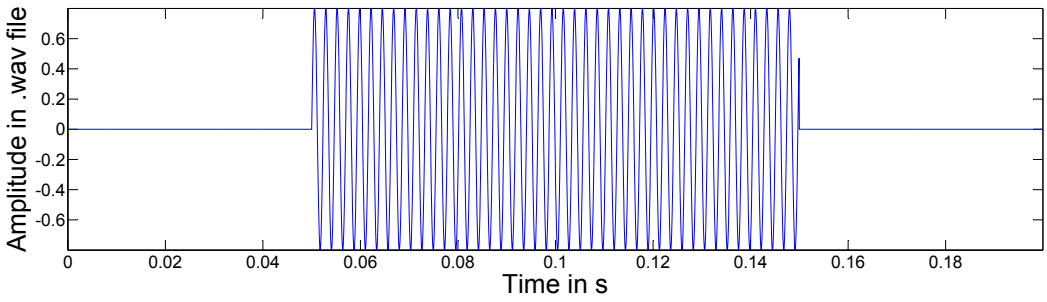

Figure 5.2: 431 Hz tone burst

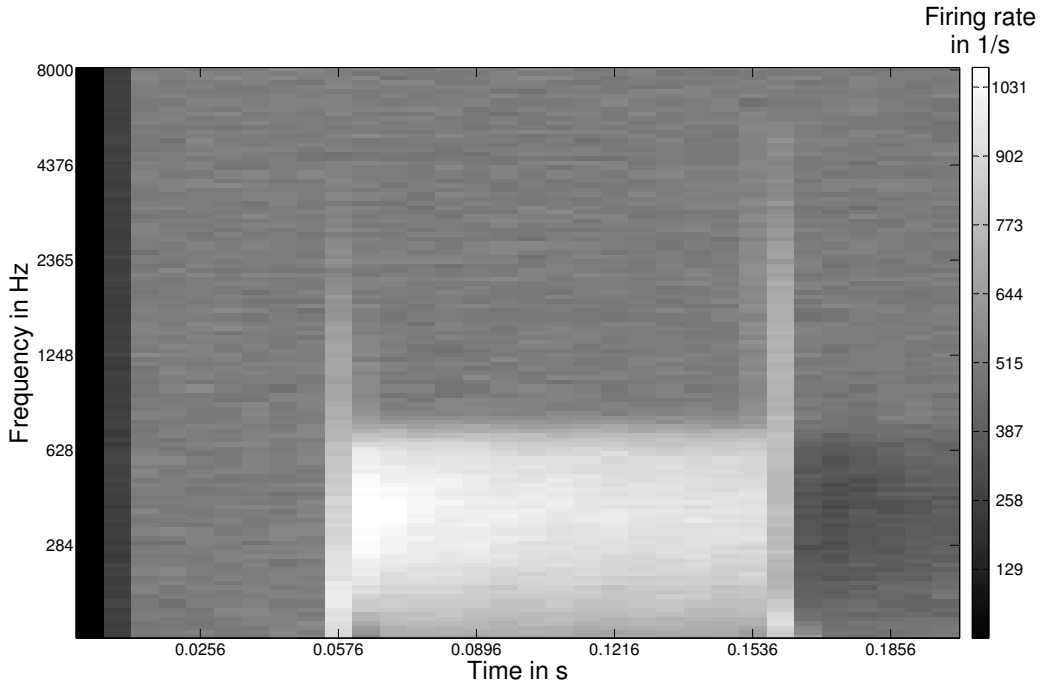

Figure 5.3: Time course of neural activity at as a response for 431 Hz tone at 40 dBSPL

Figure 5.3 have to be interpreted similarly to Figure 5.1. The lower resolution yields a more course representation, but it is necessary to reveal certain properties of the model.

The dark gray and black columns at the beginning of the neurogram are caused by the smoothing window. Activation of neurons in a wide frequency range at the onset and offset of the tone burst is an artifact. The model processes the signal in 0.01 second time frames thus at an abrupt change of the input signal results in a broad band frequency domain representation.

Neural adaptation behavior is also represented in Figure 5.3. The oversoot in firing activity can be observed as a more bright pattern after the onset signal (approximately from 0.05 to 0.08 seconds, at the frequencies around 400 Hz) compared to the rest, when firing reaches steady state values. The refractory period directly after offset of the signal is present as a darker pattern (approximately from 0.16 to 0.19 seconds an at the frequencies around 400

Hz) in the silent part following the signal where firing activity representing the spontaneous activity should be present.

5.1.3 Excitation pattern of a VCV word

In this case the VCV word 'acha' with time course depicted in Figure 5.4 is first presented to the auditory model at 40 dBSPL.

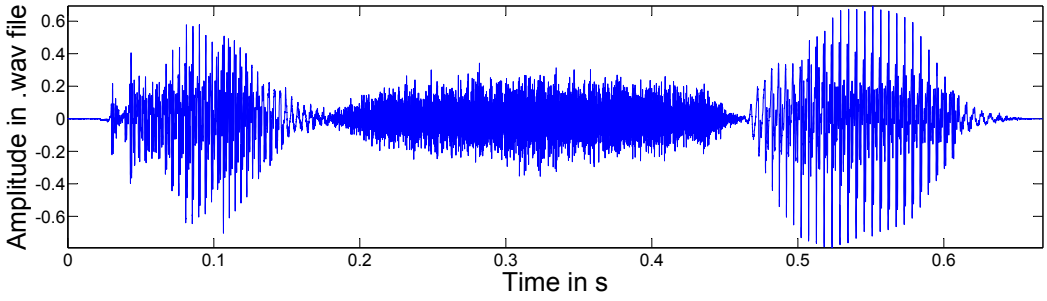

Figure 5.4: Time course of VCV word 'acha'

Its neurogram is produced by a time bin of 50 μs and a smoothing Hamming window of 32 bin length is used, resulting in a 1250Hz resolution. The procedure is repeated 20 times. The result can be observed in Figure 5.5.

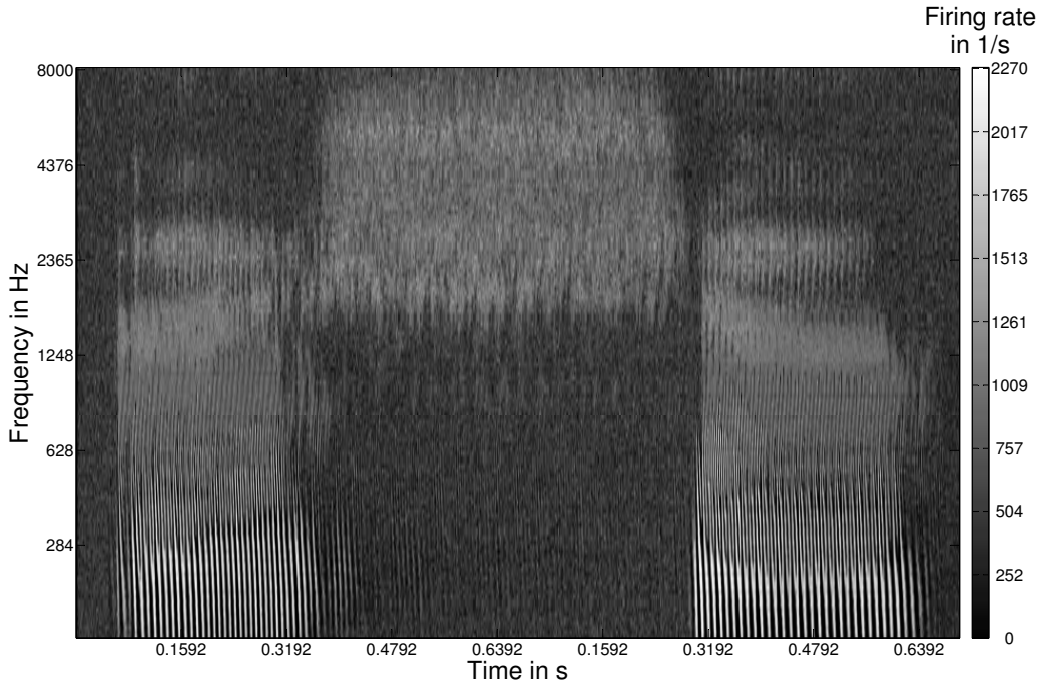

Figure 5.5: Neurogram of VCV word 'acha' at 40 dBSPL

Similarly to a spectrogram, the neurogram follows the spectral changes of the input signal. Firing activity is more pronounced at low frequency neurons at the 'a' vowels and activity is shifted to the high frequencies during the presentation of 'ch' consonant. The first formant of 'a' around 250 Hz causes the highly phase locked firing appearing as organized vertical lines in the figure. A wider firing activity pattern of vowels can be observed compared to the pattern of the consonant. This is caused by the shape of the tuning curve of the BM. As it is discussed in subsection 2.1.3 BM tuning curves have a longer tail towards the base of the cochlea i.e. the higher frequencies, thus signals with lower frequency content excite a wider range of the BM than signals with high frequency content.

Figure 5.6 illustrates a neurogram constructed in the same way as Figure 5.5, expect that the number of repetitions is 100 and the VCV word is now presented at 60dBSPL.

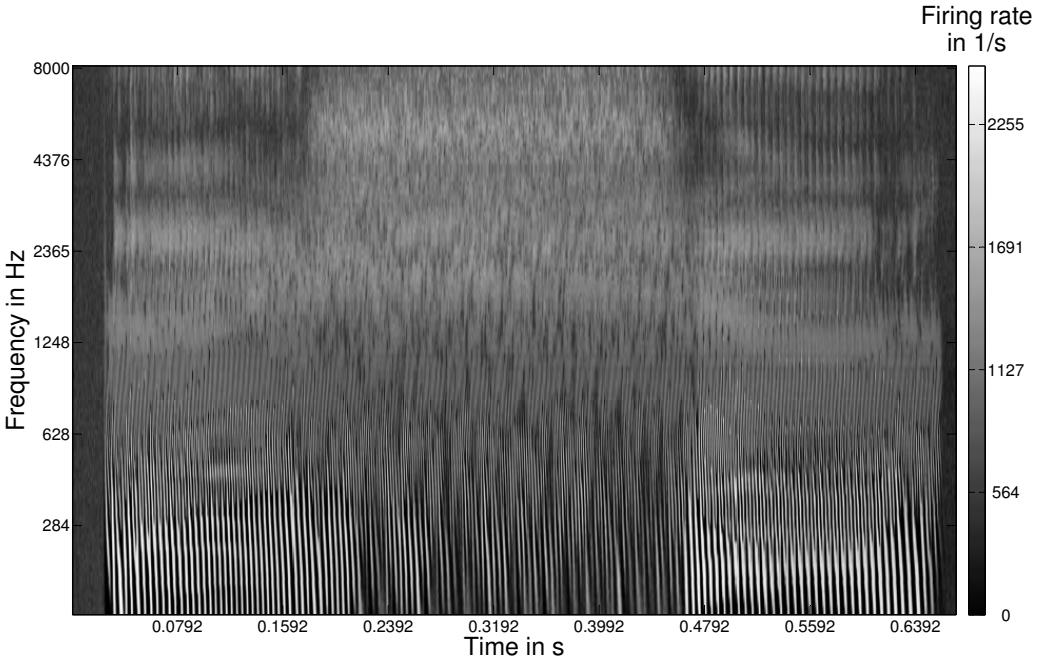

Figure 5.6: Neurogram of VCV word 'acha' at 60dBSPL

Compared to the results in Figure 5.5 a broadening of the excitation pattern can be noticed in Figure 5.6. This is the consequence of the tuning curves of the BM, that are getting wider with an increasing sound pressure level. The broad tuning curves for frequencies below 1000Hz can be seen in the lower row of Figure 3.9. The saturation of hair cells in the low frequency range can also be observed while comparing Figure 5.5 and 5.6. This effect results in the approximately same dynamic range of the two figure.

5.2 Evaluation impaired ear implementation

During evaluation of the implemented hearing loss the VCV word 'acha' is used at 60dBSPL as an input for the auditory model with modifications discussed in subsection 3.4.3. Figure 5.7 contains the results as neurograms

both for an impaired ear only in the upper graph and for a hearing aid (fitted to the half-gains) compensated loss in the lower one. A time bin of 50 μs, a smoothing Hamming window of 32 bin length and 100 repetitions are used for creating the neurograms.

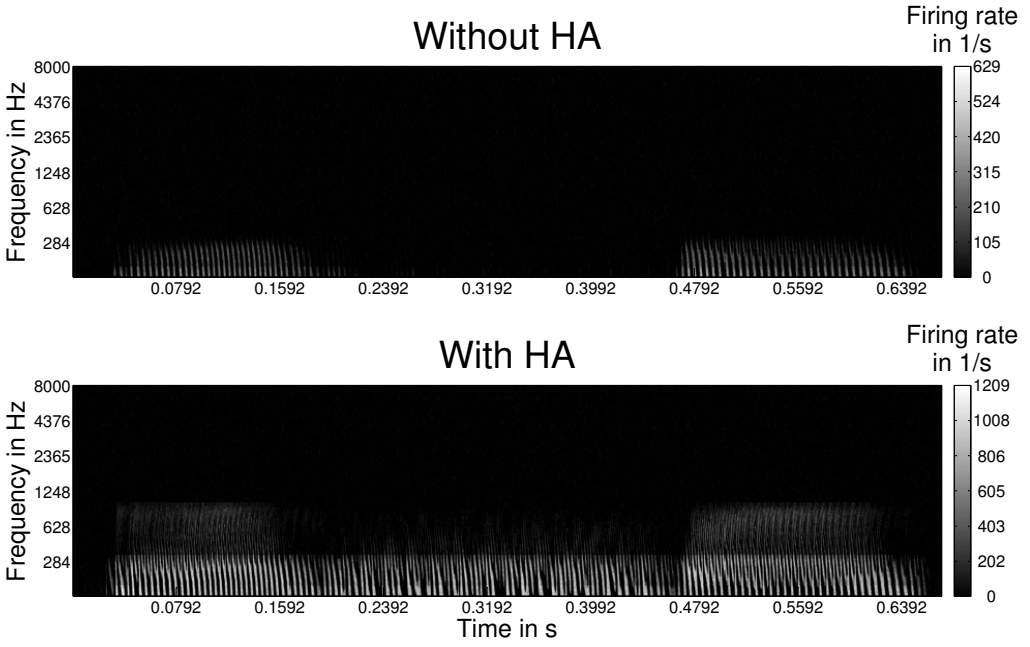

Figure 5.7: Top panel - Neurogram of VCV word 'acha' at 60 dBSPL for the implemented hearing loss; Bottom panel - Neurogram os VCV word 'acha' at 60 dBSPL for the fitted hearing aid compensated hearing loss

As it can be seen it the upper plot of Figure 5.7, most of the firing activity that is present under normal hearing conditions is ceased. The residual hearing is restricted to the low freqency range and even in this range the severe hearing loss causes a limited neural activity. The decrease in spontaneous activity is also significant, representing adequetly the lack of healthy neurons.

The situation with the fitted hearing aid compensation is presented in the lower plot of Figure 5.7. As it is expected, the firing activity is increased in the low frequency range, although the normal levels are still not met. At high input levels, due to the relatively broad tuning curves at low frequencies,

neural activity will be present also in the low frequency range for inputs with high frequency content. This effect can be observed in the 0.2 to 0.45 seconds time interval in Figure 5.7.

5.3 Evaluation of electric stimulation

Simulation of electric stimulation is completely separated from the auditory system model. Similarly to the latter evaluations the 'acha' word at 60dBSPL is used as an input for the CI model that is extended with the model of intra cochlear electric spread and the Goldwyn model that generates the neural response.

The generated neurogram with time bin of 50 μs, smoothing Hamming window of 32 bin length and 100 repetitions can be seen in Figure 5.8.

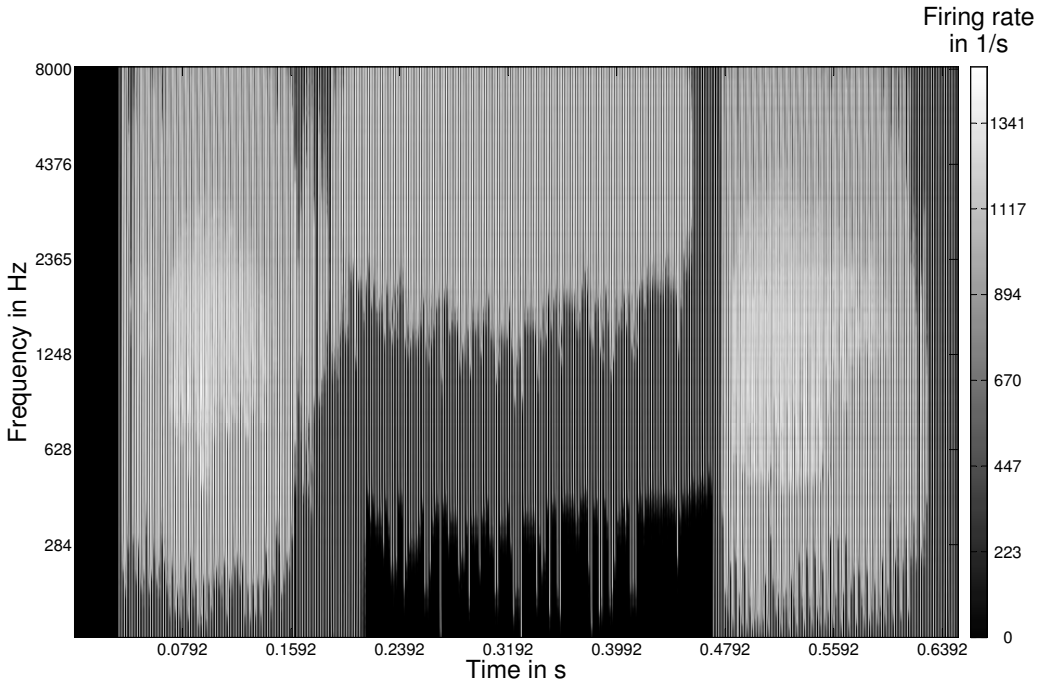

Figure 5.8: Neurogram of electric stimulation for the VCV word 'acha' at 60 dBSPL

As it can be observed in Figure 5.8 the electrode activity follows the frequency changes through the time course of the input similarly to the acoustic excitation (see Figure 5.6). The effect of electrical spread is present as simultaneous activation of a wide range of neurons besides the actual main frequency content of the signal. For example, in the time interval of 0.02 to 0.15 second the main frequency content of the VCV word is around 1500Hz and accordingly the neural response is the most vivid in this range, but the rest of the frequency range is also activated with a lower intensity.

The first main deviation from acoustic response is the lack of spontaneous activity, that is, no random noise is present during the absence of input. The second difference is the significantly higher phase locking in firing, apparent as vertical lines in the graph. As it is discussed in section 2.2, this phenomena is caused by the lack of stochastic process present at the IHC-AN synapse.

5.4 Evaluation of EAS

During evaluation of electroacoustic stimulation, the VCV word 'acha' at 60dBSPL is used as an input signal for the combined simulation chain. For the auditory path, the model with hearing impairment is used and the block of the designed hearing aid (see subsection 3.4.4) is placed before it. The electric pathway is used with the extension of simulating the intra cochlear electric field. The separate branches are connected according to the method described in subsection 3.4.5.

The generated neurogram with time bin of 50 μs, smoothing Hamming window of 32 bin length and 10 repetitions can be seen in Figure 5.8.

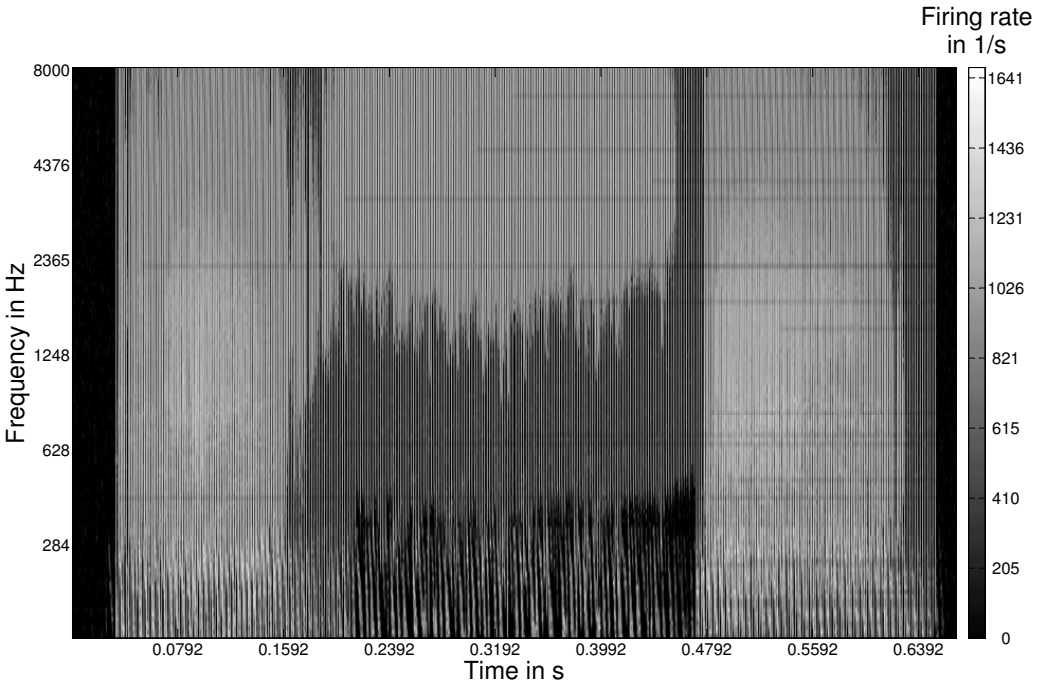

Figure 5.9: Neurogram of EAS for the VCV word 'acha' at 60 dBSPL

Figure 5.9 can be compared to Figure 5.6 and 5.8 due to the identical circumstance (by means of input signal, input level and neurogram parameters, expect the number of repetitions) used to obtain these results. Result of the hearing aid compensated hearing loss in Figure 5.7 also helps with the interpretation of the current result.

By looking at the low frequency range in Figure 5.9 the effect of the simulated residual hearing is obvious. In the time intervals, where vowels are present (0.02 to 0.2 and 0.45 to 0.65 seconds) interaction of the electric and acoustic stimulation appears as an increased firing activity reflecting well the addition of stimuli. During the representation of consonant (from 0.2 to 0.45 seconds) with frequency content above the residual hearing, the effect of the hearing aid can be seen.

The dark horizontal lines appearing in the neurogram are originated i the spike pattern. The probabilistic spike generator in the Goldwyn model produces the spike trains for each characteristic frequency separately (i.e. line-

by-line in the neurogram). Sometimes the generation of spikes ceases. So far no further investigation was taken to reveal the underlying problem of this behavior.

5.5 Evaluation of EAS performance

During evaluation of EAS the general assumption, that it performs better than an electric stimulation alone, is taken. To use this assumption, neurograms are created for normal hearing conditions, electric stimulation (ES) and EAS when the acoustic part applies the impaired ear model and the input signal is amplified with a hearing aid model. To compare performances the mean value of NSIM, with parameters described in chapter 4, is used for comparing normal hearing to ES and EAS.

VCV words with various spectral content and an tone signal with at 500Hz with changing amplitude is used as input signals to the model. Spectrograms of VCV words and the tone in time domain are shown in Figure 5.10 and 5.11 respectively. The vowel 'a' in each word contains frequencies around 1000 Hz and a formant at 250 Hz is also significant. Frequency content of the consonant varies from word to word. Frequencies around 4 kHz are dominant in 'ch' during the word 'acha', consonant 'b' in the second word, 'aba', is concentrated to the low frequencies below 250 Hz. Finally the consonant 'f' in the last word 'afa' mainly contains frequencies above 8 kHz (i.e. out of the modeling range), while exciting the whole observed frequency range at moderate level.

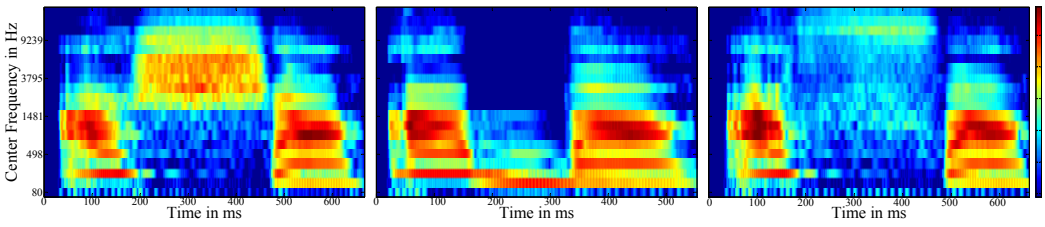

Figure 5.10: Spectrograms of VCV words. In a left to right order: 'acha', 'aba', 'afa'

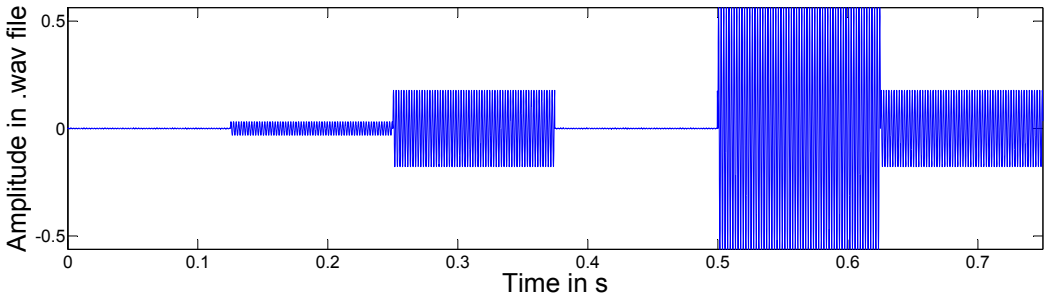

Figure 5.11: Time domain representation of a 500Hz tone with varying level

5.5.1 EAS performance with respect to varying time resolution

Before further evaluations EAS compared to ES performance has to be investigated with respect to the temporal resolution of neurograms. Later results can only be considered valid if EAS outperforms ES at the applied resolution, otherwise results would be inconsistent with listening test data (see: section 2.4).

During simulations with varying temporal resolution the 'acha' VCV word at a level of 60 dBSPL and 10 trials per neurogram is used. The EAS model contains the reference hearing aid described in subsection 3.4.4 to avoid complications that differences in the acoustic response for normal hearing and impaired hearing can cause. The used combinations of time bin and smoothing window size, resulting in different temporal resolutions can be seen in Table 5.1.

Resolution in Hz	156.25	312.5	625	1250	2500
Time bin in μs	100	100	50	50	50
Window size in bin	128	64	64	32	16

Table 5.1: Time bin and smoothing window sizes used for different temporal resolutions

Figure 5.12 illustrates the evolution of NSIM results for ES (dashed line) and EAS (continuous line)with increasing temporal resolution. The graph shows a monotonous decrease above 625Hz. This behavior can be associated with the noise like spontaneous activity, that is present in the neurogram of

normal hearing and getting more pronounced with increasing resolution. In other words, as the picture resolution is getting higher an individual neural spike is contributing more to one pixel. Both ES and EAS are compared to the normal hearing neurogram thus the noise will affect both results. Besides the general tendency of curves, a varying difference between ES and EAS results is also noticeable.

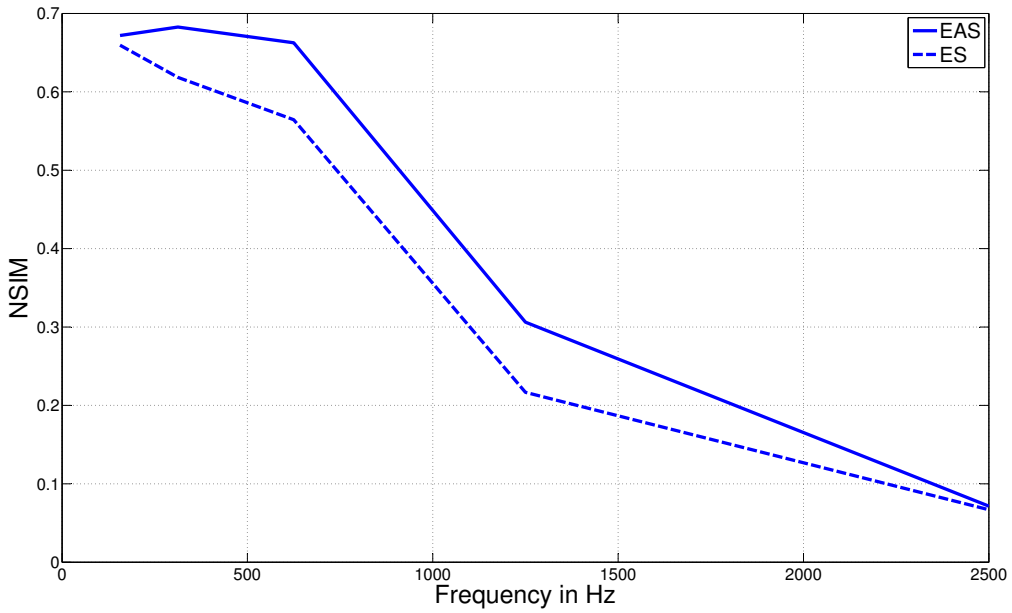

Figure 5.12: NSIM results for EAS and ES for the VCV word 'acha' at 60 dBSPL, with respect to neurogram temporal resolution

Figure 5.13 shows the difference between ES and EAS results normalized to the actual EAS value. The diagram shows a peak at 1250 Hz, indicating that the deviation is the most distinct at this resolution.

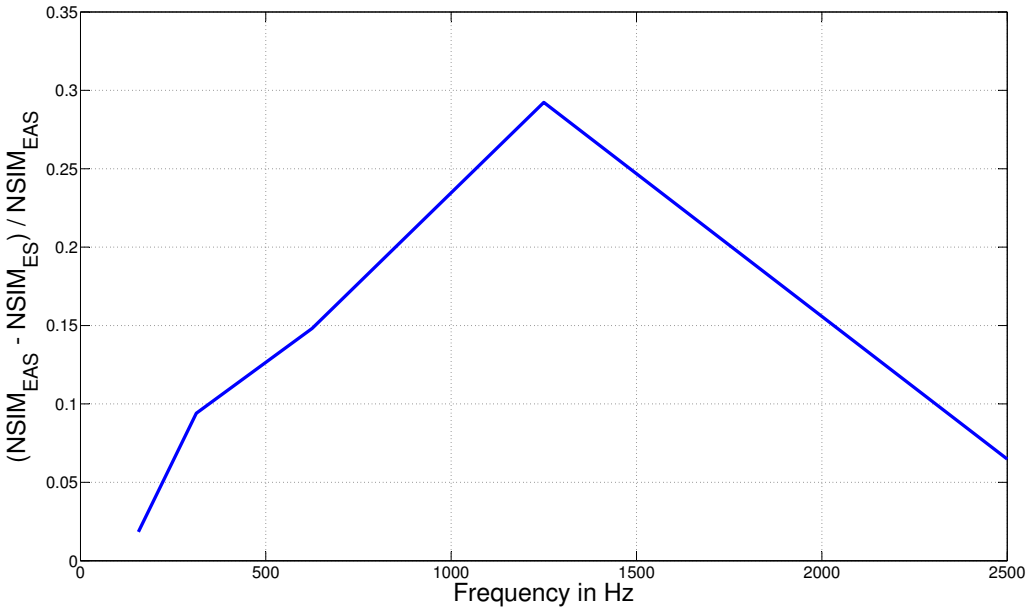

Figure 5.13: Normalized advance of EAS over ES for the VCV word 'acha' at 60 dBSPL, with respect to neurogram temporal resolution

5.5.2 Optimal activation of electrodes during EAS

To find the optimal number of active electrodes for EAS, apical electrodes, responsible for the low frequencies, are deactivated systematically and NSIM results for EAS as well as differences between EAS and ES are investigated. Unfortunately the used CI model is not capable of deactivating the electrodes inherently. To bypass this problem, the electrical spread within the cochlea is not calculated for electrodes which are assumed to be switched off. It means, that the neurons only receive electric stimulation from the active electrodes.

Throughout the simulations conducted for this evaluation 10 trials are applied to generate neurograms at an input level being set to 60dBSPL to simulate normal speech conditions, and a resolution of 1250Hz is used. By applying the latter parameter, TFS information is present in the neurograms

while the difference between ES and EAS results is definite. Investigation at a temporal resolution, that can reflect TFS information is necessary because these cues are related to the superiority of EAS over ES (see section 2.4), that is the general assumption for evaluating EAS performance.

EAS with reference hearing aid

For the first simulations the reference hearing aid is used. Figure 5.14 shows the evolution of NSIM results with respect to the number of switched off electrodes. Various input signals are marked with different colors ('acha'- blue, 'aba' - red, 'afa' - green, 'tone' - black) while dashed and continuous line show the results for ES and EAS respectively.

Results for VCV word show a decrease in similarity index with more electrodes being switched off. ES and EAS results are clearly separated in all cases with EAS showing better performance. The 'acha' word gives better NSIM values compared to the other inputs, independently of the number of switched off electrodes. Results for the tone signal are only depicted up to three turned off electrodes, because switching of more yields a lack of electrode activation since the first four apical electrodes covers the range where the tone lies. Further more, NSIM values for the tone signal are lower than the ones obtained for VCV words, although EAS is still better than ES.

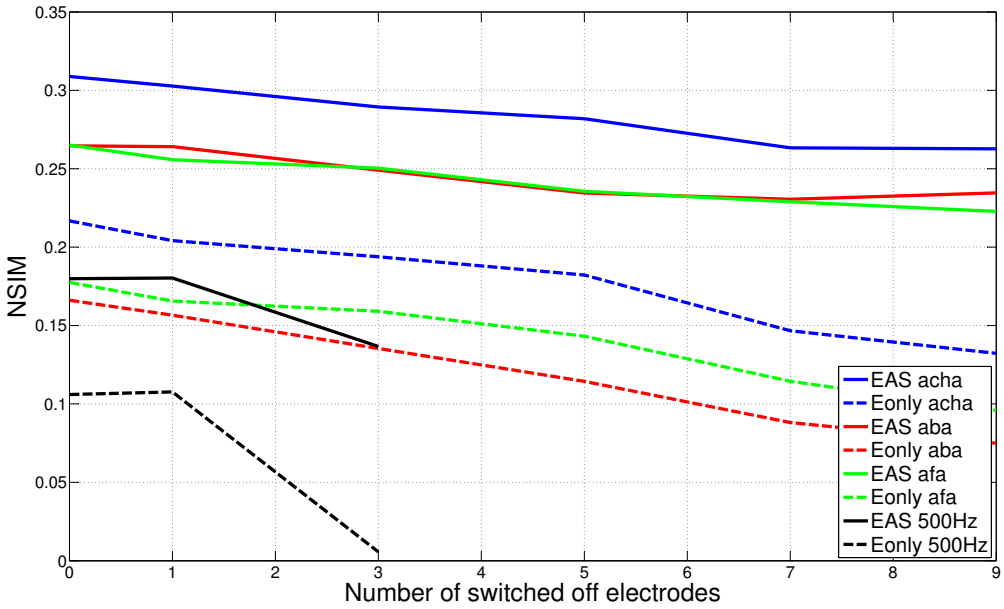

Figure 5.14: NSIM results for EAS and ES with respect to the number of switched off electrodes for various input signals at 60 dBSPL. Reference hearing aid is used in the EAS model.

These results indicates, that according to the objective measure EAS has the best performance when all electrodes are activated.

The differences in results regarding the VCV input stimuli are originated in their frequency content. During EAS the electric stimulation is dominant if the whole frequency range is investigated. The frequency content of 'acha' is mostly out of the range of residual hearing, thus electric stimulation is responsible for restoring it, yielding the overall superiority of NSIM results.

The lowest results obtained by the tone signal is due to the fact, that its frequency lies completely in the residual hearing range. It means that both the electric and acoustic part of EAS is restricted to the low frequency range, and the spontaneous activity of neurons, that is present during normal hearing conditions, in the high frequency range is missing. NSIM compares the whole neurogram, thus the overall result for the tone stimulus is going to be

low due to the differences in the dominant high frequency range.

To investigate further the effect of electric stimulation in the residual hearing range, NSIM can be calculated only for the low frequency range of the neurograms. Figure 5.15 shows the results when neurograms are compared for frequencies below 1000 Hz. This figure shows increasing NSIM values for EAS as the electric stimulation leaves the residual hearing range. This can be interpreted such that the electric stimulation distorts the acoustic during simultaneous stimulation. ES results show decreasing NSIM values, which is natural, since there is no residual acoustic hearing in this case, thus neurons are not stimulated without the presence of electric stimulus.

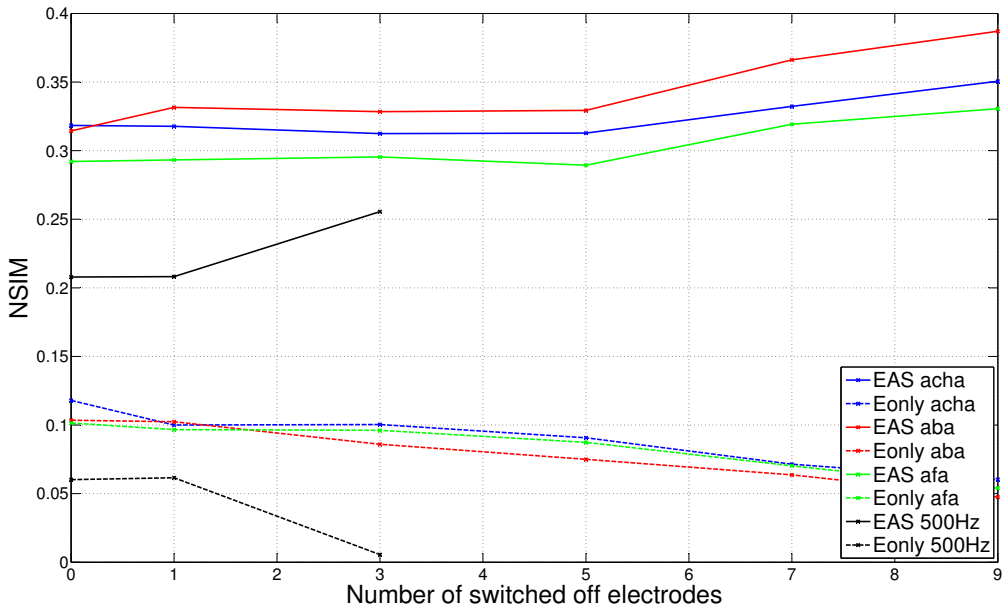

Figure 5.15: NSIM results for EAS and ES with respect to the number of switched off electrodes for various input signals at 60 dBSPL. Investigating the frequency range below 1000 Hz. Reference hearing aid is used in the EAS model.

Figure 5.16 shows the evolution of advantage of EAS over ES NSIM results normalized to the actual EAS values with respect to the number of switched off electrodes. All curves show an increasing tendency with the increas-

ing number of deactivated electrodes. Result for tone excitation with three turned off electrodes show great deviation form the others, because in this situation the electrical stimulus is limited due to the tone getting out of the frequency range covered by the residual electrodes. Among the VCV words 'aba' possesses the highest values, while 'afa' and 'acha' run close to each other , the latter being smaller in every situation.

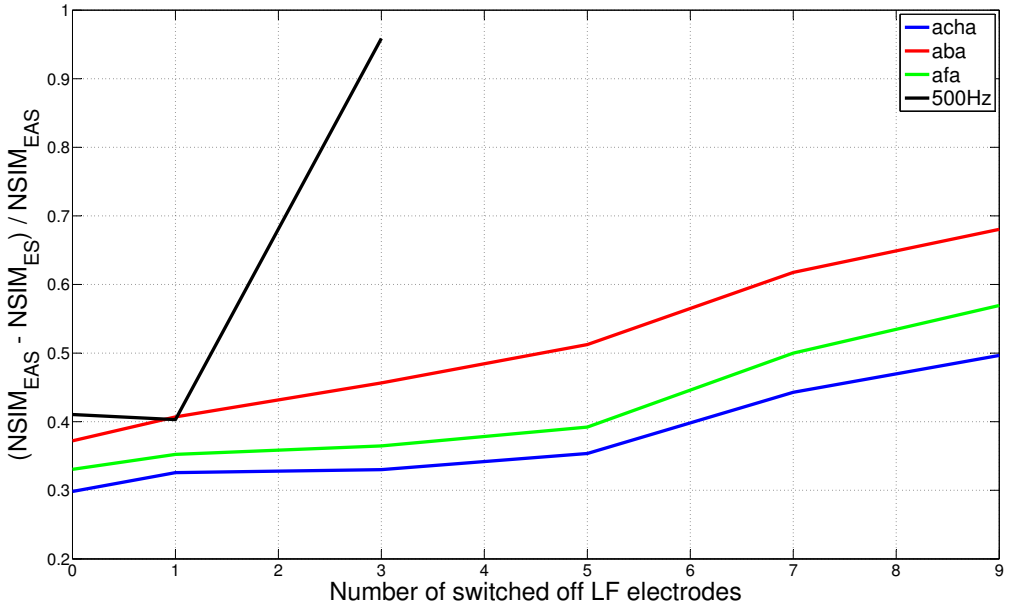

Figure 5.16: Normalized advance of EAS over ES with respect to the number of switched off electrodes for various input signals at 60 dBSPL. Reference hearing aid is used in the EAS model.

Positive values for relative differences between EAS and ES indicate the superiority of EAS in all cases. Now the order of results is the revers compared to the one depicted in Figure 5.14. This inversion can be associated with the frequency content of input signals. Signal with more frequency content in the low frequency range will benefit more from the residual hearing. The increase of relative differences with less electric excitation in the low frequency range is caused by the acoustic hearing being exploited during EAS. E.g. when no electric excitation is present in the low frequency range, ES will have a

blank band in the neurogram, while EAS will present similar excitation to the normal hearing in this range.

EAS with fitted hearing aid

The second series of simulations apply exactly the same setup as the first, only the hearing aid is changed to the half-gain rule based model.

Figure 5.17 shows the evolution of NSIM results with respect to the number of switched off electrodes. Result for the tone signal are similar to the ones presented in Figure 5.14. Result for VCV words show the same tendency and distribution by means of input stimuli. An important deviation from the results obtained in Figure 5.14 is the small difference between ES and EAS results, furthermore ES is shown to be better in representing normal hearing conditions in the case of 'aba' and 'afa'. These latter observations can also be seen in Figure 5.18, where differences are normalized to the actual EAS value.

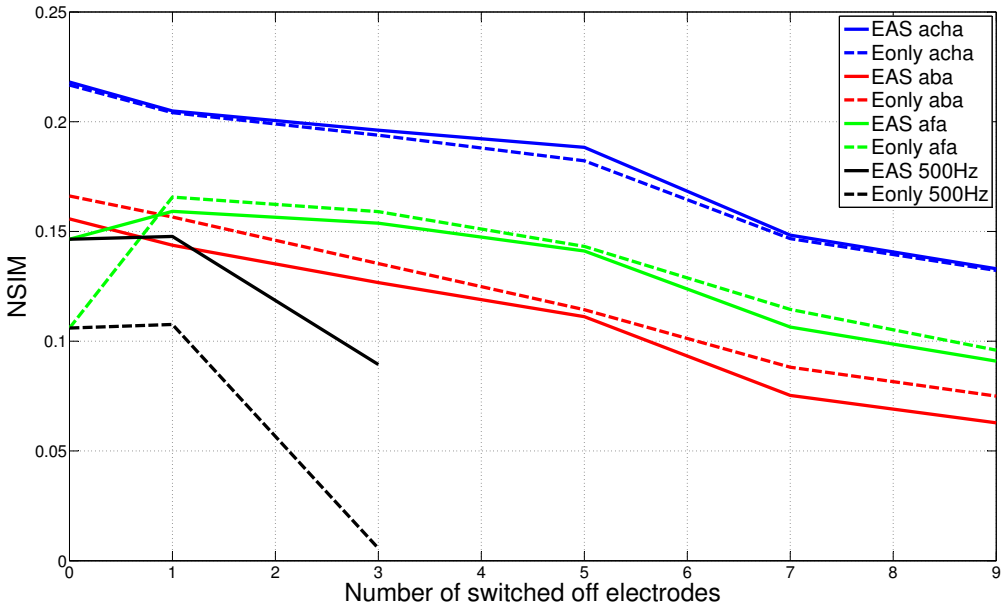

Figure 5.17: NSIM results for EAS and ES with respect to the number of switched off electrodes for various input signals at 60 dBSPL. Fitted hearing aid is used in the EAS model.

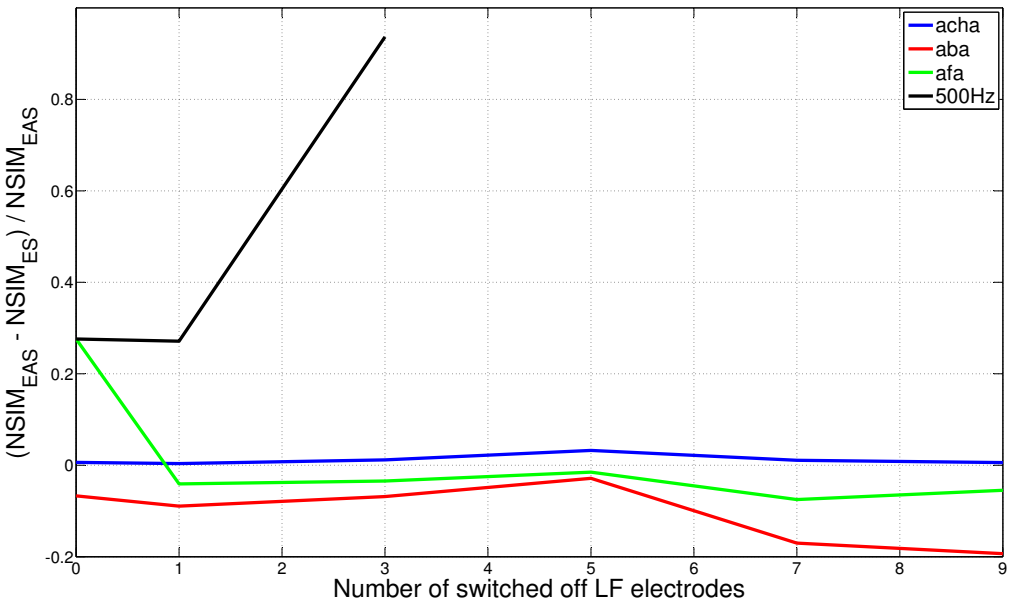

Figure 5.18: Normalized advance of EAS over ES with respect to the number of switched off electrodes for various input signals at 60 dBSPL. Fitted hearing aid is used in the EAS model.

These results are caused by the bad design of the hearing aid. Figure 5.19 shows sections of neurograms in the low frequency range at the onset of a VCV word. Normal hearing and hearing aid compensated hearing loss conditions can be observed in the top and bottom panel respectively. The delay Δt introduced by the hearing aid is illustrated by red vertical lines. As it can be seen Δt approximately coincides with the time period corresponding to the fundamental frequency of vowel 'a'. This coincidence will result in comparing black vertical patterns to white ones during calculation of NSIM. Eventually cancelling out the positive effect of additional acoustic part in EAS. Furthermore, this delay will result in lowering the performance of EAS below ES, represented by negative values in Figure 5.18. This figure also shows that input signals with more low frequency content[1] are affected more by the delay.

[1]'Although 'afa' contains high frequencies, those are out of the investigated range, thus the low frequency part is dominant in its the neurogram.

NSIM results for the tone signal also show a decrease compared to the simulation with reference hearing aid, but since the period time of the tone is larger, the complete cancellation will not be present.

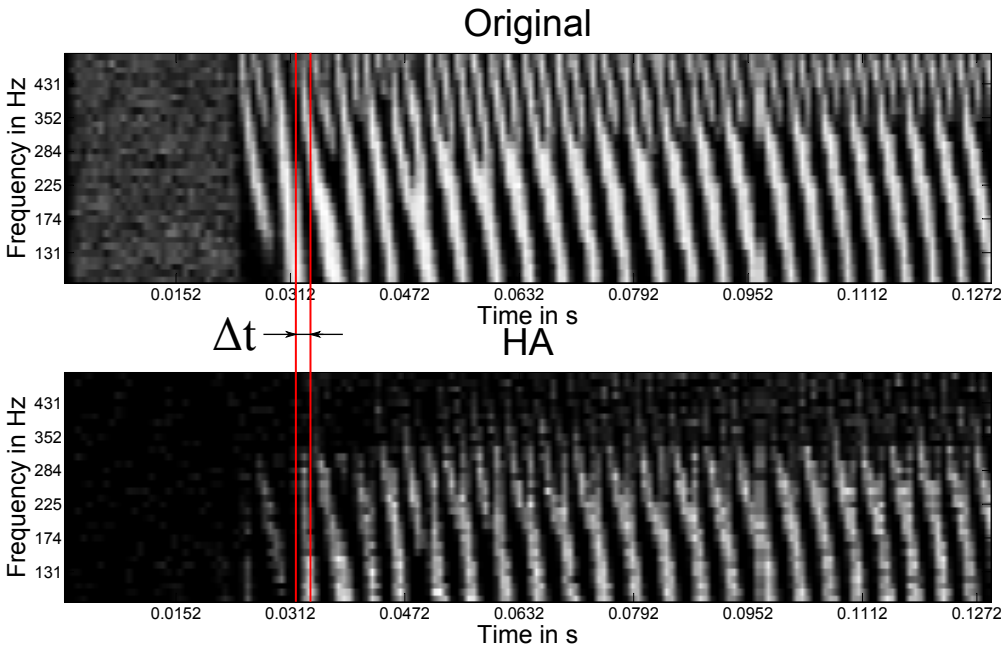

Figure 5.19: Illustration of delay introduced by the fitted hearing aid. Top figure - Neurogram for the VCV word 'acha' at 60 dBSPL for normal hearing conditions; Bottom figure - Neurogram for the VCV word 'acha' at 60 dBSPL for hearing loss compensated by the fitted hearing aid.

6 | Discussion

Result for simulating the auditory system reflects well the physiological data discussed in section 2.1. Behavior of basilar membrane (tonotopic organization, tuning curves, traveling wave), inner hair cells (saturation), and auditory nerves (neural adaptation - overshoot and refractory period) can be observed in neurograms.

Although the representation of auditory system by neurograms is qualitatively correct, the obtained firing rates are higher than the reported empirical data (discussed in subsection 2.1.3). Firing rate results shown in Figure 3.8 are consistent with empirical data, thus the problem can be related to the spike generation method implemented in the Goldwyn model. Nevertheless the same spike generator is used for acoustic, electric and EAS simulation, thus the quantitative error of firing rates does not effect the final results.

Implementation of an impaired ear can be considered successful. A better representation can be done by adjusting the bandwidth of gammatone filters in the non-linear path of DRNL filter, that would result in the widening of tuning curves for lower input levels. This would be noticeable in the residual hearing range.

Results for electric stimulation represents well the case of a CI usage, i.e. high phase locking of neurons, lack of spontaneous activity and presence of intra cochlear electric field.

The problem of non-optimal simulation of electrode deactivation (see: subsection 5.5.2) can distorts the results. The actual solution for the problem rather simulates damaged electrodes, than deactivated ones. Probably by simulating turned off electrodes properly, NSIM results would change in the same direction for both ES and EAS and the final results would be the same.

This part must be investigated further in a future work.

EAS results show the properties discussed both for the acoustic and electric part. Based on the assumption, that is taken in section 3.4 the connection of paths is successful. EAS neurograms are similar to ES and acoustic neurograms in the corresponding frequency ranges and the interaction in firing activity appears where both stimuli are present.

Results obtained by deactivation of electrodes suggest that the best performance can be achieved by all electrodes being activated. This result have to be handled with care, because of the non-optimal electrode control. Further investigation is needed in this case. On the other hand, investigation restricted to the low frequencies, where the residual hearing is present, shows that hearing is preserved in a better way without electric stimulation. This can be caused by the nature of the combined stimulation or it can indicate a synchronization problem in the electric stimulation. Results with the fitted hearing aid show that NSIM can be used to evaluate temporal synchronization. This property of the objective measure later can be used to evaluate proper fitting of CI and hearing aid.

7 | Conclusions

The main goal of implementing a working model that combines electric and acoustic hearing can be considered successful, but the assumption of combination (see section 3.4) have to be validated by further experiments. A hearing loss implementation according to a specific impairment, typically found in CI users, is also done. The NSIM objective measure is used to evaluate the built model and its eligibility for pointing out temporal synchronization problems in EAS is shown. An attempt to find optimal electrode activation in CI during EAS is made, but further work is needed to obtain reliable results.

As future work both the auditory model and the acoustic model can be improved and further studies on EAS can be made. A better representation of the outer and middle ear transfer function can be incorporated in the acoustic model and reintroducing the efferent effects would also improve it. The presence of an electrode array inside the cochlea also changes the behavior of the inner ear, thus the changes according to this also have to be investigated. The electric stimulation model in the current state assumes a perfect placement of electrodes, thus simulation of place-mismatch can serve as a further development. A study on the optimal compression applied in both the acoustic and electric part can be conducted besides investigating the optimal crossover-frequency. The proposed method for connecting the electric and acoustic path can be validated by a study using an EAS based vocoder and an evaluation by CI users can also be conducted.

Bibliography

[1] Fan-Gang Zeng. Trends in cochlear implants. *Trends in Amplification*, 8(1), 2004.

[2] Philipos C. Loizou. Mimicking the human ear. *IEEE Signal Processin Magazine*, 1998.

[3] Bart Vaerenberg, Vincent Péan, Guillaume Lesbros, Geert De Ceulaer, Karen Schauwers, Kristin Daemers, Dan Gnansia, and Paul J Govaerts. Combined electric and acoustic hearing performance with zebra® speech processor: Speech reception, place, and temporal coding evaluation. *Cochlear Implants International*, 14(3):150–157, 2013.

[4] Bruce J. Gantz, Christopher Turner, Kate E. Gfeller, and Mary W. Lowder. Preservation of hearing in cochlear implant surgery: Advantages of combined electrical and acoustical speech processing. *The Laryngoscope*, 115(5):796–802, 2005.

[5] Hugh J. McDermott Tom Francart. Psychophysics, fitting, and signal processing for combined hearing aid and cochlear implant stimulation. *Ear and Hearing*, 34(6):685–700, 2013.

[6] Jane Tu Ariel Moctezuma. An overview of cochlear implant systems. *BIOE 414*, 2011.

[7] Paul J. Abbas. *Otorhinolaryngology - Head and Neck Surgery*, chapter 143, Physiology of the Auditory System. Mosby, 2001.

[8] Chritoffer Bovbjerg, Rune Egedal, and Attila Fráter. Simulating surround sound in headphones using binaural techniques, 2013. AAU project at the department of Acoustics and Audio Technology.

[9] Brian C. J. Moore. *An Introduction to the psychology of hearing.* Academic Press, 2003.

[10] Ray Meddis. *Matlab Auditory Periphery (MAP) Model technical description.* Hearing Research Lab, University of Essex, 2011.

[11] Donald D. Greenwood. A cochlear frequency-position function for several species-29 years later. *The Journal of the Acoustical Society of America*, 87(6):2592–2605, 1990.

[12] Fan-Gand Zeng. Compression and cochlear implants. In SidP. Bacon, RichardR. Fay, and ArthurN. Popper, editors, *Compression: From Cochlea to Cochlear Implants*, volume 17 of *Springer Handbook of Auditory Research*, pages 184–220. Springer New York, 2004.

[13] Christian Sejer Pedersen. Anatomy and physiology of the human hearing, 2013. AAU lecture notes.

[14] Christian Sejer Pedersen. Hearing diagnosis and hearing disorders, 2013. AAU lecture notes.

[15] Shin-Ichi Usami, Hideaki Moteki, Nobuyoshi Suzuki, Hisakuni Fukuoka, Maiko Miyagawa, Shin-Ya Nishio, Yutaka Takumi, Satoshi Iwasaki, and Claude Jolly. Achievement of hearing preservation in the presence of an electrode covering the residual hearing region. *Acta Oto-Laryngologica*, 131:405–412, 2011.

[16] Robert Cowan Paola V. Incerti, Teresa Y.C. Ching. A systematic review of electric-acoustic stimulation: Device fitting ranges, outcomes, and clinical fitting practices. *Trends in Amplification*, 2013.

[17] Naomi Harte and Andrew Hines. Evaluating sensorineural hearing loss with an auditory nerve model using a mean structural similarity measure. 2010.

[18] Enrique A. Lopez-Poveda and Ray Meddis. A human nonlinear cochlear filterbank. *The Journal of the Acoustical Society of America*, 110(6):3107–3118, 2001.

[19] Shihab A. Shamma, Richard S. Chadwick, W. John Wilbur, Kathleen A. Morrish, and John Rinzel. A biophysical model of cochlear processing: Intensity dependence of pure tone responses. *The Journal of the Acoustical Society of America*, 80(1):133–145, 1986.

[20] Joshua H. Goldwyn, Jay T. Rubinstein, and Eric Shea-Brown. A point process framework for modeling electrical stimulation of the auditory nerve. *Journal of Neurophysiology*, 5:1430–1452, 2012.

[21] Kokoska S. Zwillinger D. *Standard Probability and Statistics Tables and Formulae*. CRC press, 2000.

[22] Muhammad S. A. Zilany and Ian C. Bruce. Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery. *The Journal of the Acoustical Society of America*, 120(3):1446–1466, 2006.

[23] Brian C.J. Moore and Brian R. Glasberg. A revised model of loudness perception applied to cochlear hearing loss. *Hearing Research*, 188(1–2):70 – 88, 2004.

[24] Christian J. Sumner, Enrique A. Lopez-Poveda, Lowel P. O'Mard, and Ray Meddis. A revised model of the inner-hair cell and auditory-nerve complex. *Acoustic Society of America*, 2001.

[25] Keith P. Purpura Jonathan D. Victor. *Analysis of Parallel Spike Trains*, chapter 7, Spike Metrics. Springer, 2010.

[26] Andrew Hines and Naomi Harte. Speech intelligibility from image processing. *Speech Communication*, 52(9):736 – 752, 2010.

[27] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, April 2004.

[28] Andrew Hines and Naomi Harte. Speech intelligibility prediction using a neurogram similarity index measure. *Speech Communication*, 54(2):306 – 320, 2012.

[29] Don H. Johnson. Point process models of single-neuron discharges. *Journal of Computational Neuroscience*, 3:275–299, 1996.

A | Dual-Resonance-Non-Linear filter

The DRNL filter used in the MAP model (section 3.1) is based on the work presented in [18], that aimed to fit the filter to empirical human psychophysical data. This model uses stapes velocity as an input and basilar membrane velocity as an output as it is depicted in Figure A.1. This a deviation from the MAP model that uses displacements results in a simplification in the parametrization, such that constant values can be used along the basilar membrane and the low-pass filters are not needed.

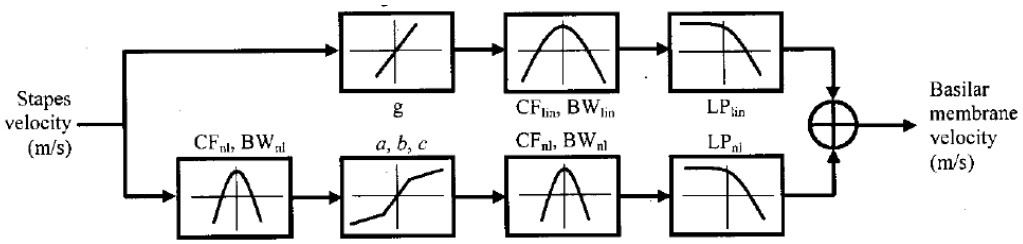

Figure A.1: Schematic of the dual-resonance-non-linear filter.[18]

A.1 Parts of DRNL filter

The gammatone filters both in the linear and non-linear part are used to approximate frequency selectivity of ear. The linear gain g is responsible for adjusting the level when the basilar membrane response becomes linear again for high input levels. The broken-stick nonlinearity contributes to the basilar membranes input-output function both at the low input stimulus and

the compression range, and it is defined as follows[18][1]:

$$y(t) = sign[x(t)] \cdot min[a|x(t)|, b|x(t)|^c] \tag{A.1}$$

Where x(t) is the input signal, y(t) is the output signal and a,b and c are parameters of the function.[2] As it can be seen in Figure A.2, this function behaves linearly for low input levels and a compression introduced by c starts only for higher x(t) input.

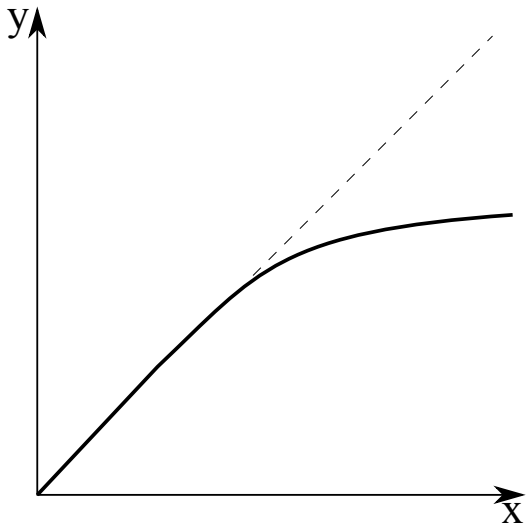

Figure A.2: Schematic representation of broken-stick compression.

A.2 Output of DRNL filter

Output of the DRNL filter at certain basilar membrane position for low and high input sound level can be observed in Figure A.3.

The dominance of the nonlinear path for low level stimuli and the dominance

[1] The implementation of this function in MAP differs from the one represented here because of the velocity-displacement difference. Nevertheless the underlying idea of including a broken-stick nonlinearity is the same.

[2] a, b and c are constant along the basilar membrane in MAP model due to the velocity-displacement difference

of the linear path for high level stimuli is obvious from this illustration. Even though the output of both paths grow with increasing input levels, the one in the nonlinear path changes in a slower pace due to the compression and that explains the exchange of superiority.

The change of bandwidth with level is also apparent in Figure A.3, as well as the shift of characteristic frequency. The presence of these behaviors can be related respectively to the difference in the gammatone filter bandwidths and center frequencies used in the parallel paths.

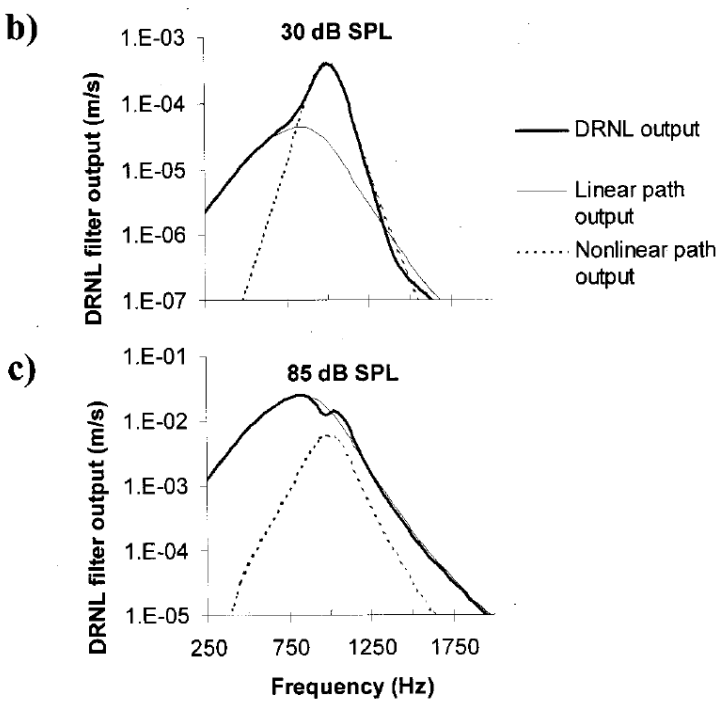

Figure A.3: Output of DRNL for low- and high-level input stimulus.[18]

A.3 DRNL filter parameter roles

The basilar membrane input-output function described in subsection 2.1.3 is attempted to be simulated for arbitrary frequencies by the DRNL filter.

The linear part of basilar membrane response for low input levels is controlled

by the nonlinear path of the DRNL filter, since this is the dominant in this input level range (see Figure A.3). As it can be seen from Equation A.1 the output of this path is linear for low levels and parameter a takes the main role and it is adjusted to fit empirical data of basilar membrane response at hearing threshold for normal hearing. The tuning of the membrane is controlled by CF_{nl} the bandwidth of gammatone filters[3].

The starting point of the compressed part is controlled by parameter b[4]. Although parameter c is responsible for the level of compression in the broken-stick function (see: Equation A.1) it is just indirectly related to the nonlinear part of the filter output, since it is the result of the whole system.

At the highest input levels (approximately above 85dB [19]) the linear part of the filter becomes dominant as it can be seen in Figure A.3. The level at which the filter behaves linearly again is set by parameter g and CF_{lin} the bandwith of the gammatone filters in the linear path. To obtain the psychophysical property of less tuning high stimui (see subsection 2.1.3), CF_{lin} is set to be less than CF_{nl}. At last a downward shift of best frequencies with input levels between 65 and 95dB [19] is also introduced by setting a constraint of CF_{lin}/CF_{nl} ratio being less than one.

[3]During implementation of an impaired ear, parameter a can be changed to raise hearing threshold and bandwidth of gammatone filters in the nonlinear path can be increased to lower the tuning of the membrane

[4]The MAP implementation uses a variable ctBM that explicitly sets the border of linear and compressed part.

B | Viscous coupling between basilar membrane displacement and stereocilia displacement

The underlying idea of the modeling of stereocilia displacement u as a responses to basilar membrane displacement ω is visualized in Figure B.1 and can be summarized in the following expression [19]:

$$\tau_c \frac{\partial u}{\partial t} + u = \tau_c C \frac{\partial \omega}{\partial t} \tag{B.1}$$

Where τ_c is a constant proportional to the geometry of stereocilia and the viscosity of endolymph which are consider to be constant along the cochlea for simplification. C is a constant depending on the geometry of space between the hair cell and the tectorial membrane.

Equation B.1 represents the viscous coupling. The basilar membrane movement evokes a fluid flow that is represented by Q the mean transverse component of fluid velocity in Figure B.1.

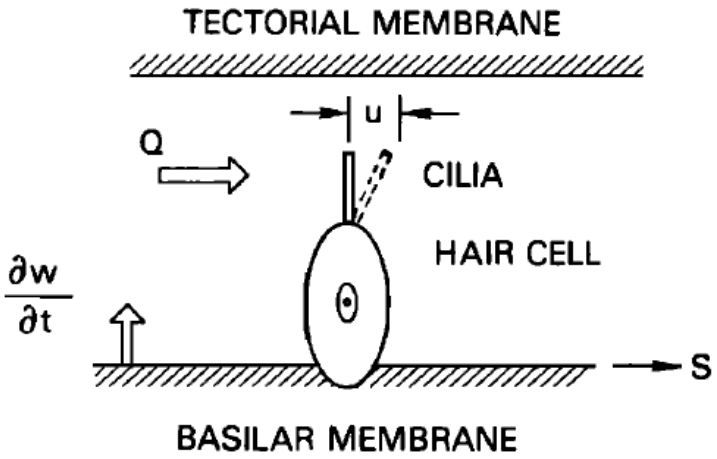

Figure B.1: Schematic of basilar membrane displacement and cilia displacement relation. [19]

Because of linearity the velocity[1] of basilar membrane $\frac{\partial \omega}{\partial t}$ is proportional (by C) to Q and it is proportional (by τ_c) to the force deflecting the cilia that takes place in the right part of Equation B.1.

The restoring forces in the left part of Equation B.1 are the dissipative force and the elastic force proportional to the cilia velocity $\frac{\partial u}{\partial t}$ and displacement respectively.

This model of coupling is independent of position along the cochlea. In reality stereocilia length l varies with position [19] that could be represented by a variable τ_c, because only the elastic force is independent of l and τ_c at the same time.

[1]Note that in the MAP model uses basilar membrane displacement directly at this point (See Equation 3.1). Because of this, the constants used in that model differs from the ones used in [19], but the underlying idea remains the same.

C | Relation of Goldwyn model to point processes

This appendix contains the relation between response statistics used in the Goldwyn model and point processes. Besides, expressions for functions used in the Goldwyn point process model can be found in the second section. [20] is used for creating the list of expressions and Table C.2 containing the related parameter values.

C.1 Relation of response statistics to point process

The conditional intensity function $\lambda(t|I,H)$[1] (I and H stands for the amplitude and the history of the input pulse respectively) completely defines the point process as it is discussed in Appendix D and its integration can be written as follows [20]:

$$\Lambda(t_1,t_2|I,H) = \int_{t_1}^{t_2} \lambda(s|I,H)\,\mathrm{d}s \tag{C.1}$$

The probability distribution function that can be interpreted as the probability that a spike will be produced in a time interval of $[t_1,t_2]$ can be derived

[1] Note that in Appendix D $\mu(.)$ was used as a sign for the intensity function and $\lambda(.)$ denoted the rate of discharge, but as it was explained, for a Poisson process these two values are equal.

from $\Lambda(t_1, t_2|I, H)$ and known as the lifetime distribution function [20]:

$$L(t_1, t_2|I, H) = 1 - e^{-\Lambda(t_1,t_2|I,H)} \tag{C.2}$$

Observe that Equation C.2 is similar to the desired probability distribution function (D.9) applied in section D.4. It can also be seen that if C.2 is the function of I, it represents the firing efficiency curve. Likewise if C.2 is the function of t_2, then it represents the probability that a stimulus will elicit a spike before t_2 i.e. the temporal dispersion of spiking is expressed in this case. Consequently if t_2 goes to infinity, C.2 shows the probability that a spike will ever occur.

The relation among response statistics and lifetime distribution function $L(t_1, t_2|I, H)$ (or the integrated intensity function $\Lambda(t_1, t_2|I, H)$) is summarized in Table C.1.

Where $\theta(D_c)$ is the threshold for chronaxie and $\theta(\infty)$ is the threshold for monophasic long pulse duration. The probability density function w.r.t I is the derivative of the lifetime distribution is defined as [20]:

$$l_I(0, \infty|I) = \frac{\mathrm{d}}{\mathrm{d}I}[\Lambda(0, \infty|I)]e^{-\Lambda(0,\infty|I)} \tag{C.3}$$

and similarly the probability density function w.r.t t is the derivative of the lifetime distribution [20]:

$$l_t(0, t|\theta) = 2\lambda(0, t|\theta)e^{-\Lambda(0,t|\theta)} \tag{C.4}$$

[2]

$$f(x) = \begin{cases} x^\alpha & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases} \tag{C.5}$$

[2]The multiplier of two is originated in the definition of chronaxie that can be interpreted such as the probability of producing a spike at any time for an input stimulus level of θ is twice the probability of producing a spike in the [0,t] time interval i.e. $2L(0, t|\theta)$.

Response statistics	Definition	Relation to point process
Threshold (θ)	Current level for which the firing efficiency curve is one half	$\frac{1}{2} = L(0,\infty\|\theta)$ or $\Lambda(o,\infty\|\theta) = \log 2$
Relative Spread (RS)	Standard deviation of density function w.r.t. I ($l_I(0,\infty\|I)$), divided by its mean	$\frac{\sigma(l_I(0,\infty\|I))}{\mu(l_I(0,\infty\|I))}$
Chronaxie (D_c)	pulse duration for which the threshold is twice the value as for a long pulse	$\theta(D_c) = 2\theta(D_\infty)$
Jitter	Standard deviation of density function w.r.t. t ($l_t(0,t\|\theta)$)	$\sigma(l_t(0,t\|\theta))$
Summation effect	Same statement as for threshold, but θ is defined by a pair of pulses	$\Lambda(0,\infty\|\theta_{pair}) = \log 2$
Refractory effect	Dependency of θ on spike times	History condition (H) is introduced

Table C.1: Summary of the relation of response statistics - used for configuring the point process framework - to point process expressions. The top four statistics are measures for a single pulse and the bottom two are measures including history dependence.

Response statistics	Value	Model Parameter	Value
Threshold	0.852 mA	κ	9.365
Relative spread	4.87%	α	24.52
Chronaxie	276 μs	τ_κ	325.4 μs
Jitter	85.5 μs	τ_J	94.3 μs
Summation time	250 μs	β	0.333

Table C.2: Values for respose statistics and the corrsponding parameters

$$K^+(t) = \begin{cases} \frac{\kappa}{\tau_\kappa} e^{-\frac{t}{\tau_\kappa}} & \text{if } t \geq 0 \\ 0 & \text{else} \end{cases} \tag{C.6}$$

$$K^-(t) = \begin{cases} \frac{\beta_\kappa}{\tau_\kappa} e^{-\frac{t}{\tau_\kappa}} & \text{if } t \geq 0 \\ 0 & \text{else} \end{cases} \tag{C.7}$$

$$J(t) = \begin{cases} \frac{1}{\tau_J} e^{-\frac{t}{\tau_J}} & \text{if } t \geq 0 \\ 0 & \text{else} \end{cases} \tag{C.8}$$

In equations C.5 to C.8 the parameters (κ, α, τ_κ, τ_J and β) can be directly related to the response statistics as follows:

D | Point process theory

Neural activity is represented as action potentials occurring at the auditory nerve. Point process is used for measurement and analysis of a sequence of action potentials that represents a neural code [29], thus it is important to have a basic understanding of it.

D.1 Definition of point process

The occurrence of an action potential in a time interval of Δt in a regular point process can be defined as a constraint probability function as follows [29]:

$$P[\text{one event in}[t, t+\Delta t)|N_t, \mathbf{w}_t] = \mu(t; N_t, \mathbf{w}_t)\Delta t \tag{D.1}$$

where N_t is number of events occurred before time t, $\mathbf{w}_t$ is is a vector containing the times occurrences and $\mu(.)$ is the intensity of the point process. It can be observed that the probability of occurrence is directly proportional to the length of the time interval. The intensity function is related to the instantaneous rate (defined as events per second) of the process and its variation can represent deterministic or stochastic external influences. $\mu(.)$ is also dependent on N_t and $\mathbf{w}_t$ i.e the history of the process that can account for dynamic neural behaviors such as refractory effect.

D.2 The Poisson point process

The Possion process can be considered as the base of any more complex point process. Its intensity function described in Equation D.2 does not contain the part responsible for the process's history, in other words the Poisson process has no memory. For defining a stationary process, the rate of discharge $\lambda(t)$ is constant, while variation of $\lambda(t)$ over time accounts for a doubly stochastic Poisson process.[29]

$$\mu(t; N_t, \mathbf{w}_t)\Delta t = \lambda(t) \tag{D.2}$$

The rate of discharge $\lambda(t)$ is defined as the limit of the expected number of events occurring in a time interval divided by the length of the time interval [29]:

$$\lambda(t) = \lim_{\Delta t \to 0+} \frac{\varepsilon[N_{t+\Delta t} - N_t]}{\Delta t} \tag{D.3}$$

For a Poisson process the expected value can be expressed as follows:

$$\varepsilon[N_{t+\Delta t} - N_t] = \int_t^{t+\Delta t} \lambda(\alpha)\, \mathrm{d}\alpha. \tag{D.4}$$

Due to this property the instantaneous rate is equal to the intensity [29].

D.3 Renewal process

By definition a renewal point processes can be described as a point process that depends on the occurrence time of the last event:

$$\mu(t; N_t, \mathbf{w}_t) = \mu(t; N_t, w_{N_t}) \tag{D.5}$$

If the intensity is independent of the absolute time, then it is only the function of time since the last event:

$$\mu(t; N_t, w_{N_t}) = h(t - w_{N_t}) \tag{D.6}$$

For constant $h(.) = h(\tau)$, the Poisson process is obtained, that can be rewritten to describe a process with absolute refractory period:

$$h(\tau) = \mu_0 u(\tau - \Delta) \tag{D.7}$$

Where u(.) is a unit-step function and Δ accounts for the absolute refractory period during which the probability of an event is zero. After time exceeds Δ the process behaves similarly to a Poisson process.

D.4 Point process generation

To produce a point process, calculation of subsequent interevent interval durations is used. The most basic practice relies on Equation D.1 i.e. a biased coin is flipped as a predefined Δt passes and an event is produced when 'the coin comes up heads'. In this method the probability of 'head' is the product of the intensity function and the interval duration.

Despite being the most straightforward method, it has undesirable properties. First to accurately mimic neural discharges the probability of 'heads' should be significantly less than one, that is challenging to produce with a random generator. Secondly the small probability of event generation results in a huge computational loss for not producing an event.

D.4.1 Inverse distribution function technique

A more sophisticated and less computationally heavy technique uses the inverse distribution function for generating interevent time intervals. This

method exploits the fact that the probability distribution function of a random variable applied to the random variable itself results in a uniformly distributed random variable (this statement can be extended to be applicable for constrained random variables)[29]:

$$U = P_X(X) \tag{D.8}$$

First a uniformly distributed random variable U is generated. Then X is calculated such that it satisfies Equation D.8. In this way X is going to have a probability distribution function described by P_X.

Applying this method for point process generation, the steps can be substituted as follows:

- define the desired probability distribution function:

$$P_{\tau_{n+1}|\tau_n,\ldots,\tau_1}(\tau_{n+1}|\tau_n,\ldots,\tau_1) = 1 - e^{-\int_0^{\tau_{n+1}} \tilde{\mu}(\alpha;n,\tau)\,\mathrm{d}\alpha} \tag{D.9}$$

- generate exponentially distributed random variables E_n[1]
- calculate τ_n interevent intervals until the following integral stands:

$$E_{n+1} = \int_0^{\tau_{n+1}} \tilde{\mu}(\alpha; n, \tau)\,\mathrm{d}\alpha \tag{D.10}$$

For a stationary renewal process the integral in Equation D.10 can be analytically expressed as:

$$E_{n+1} = \begin{cases} 0, & \tau_{n+1} < \Delta \\ \lambda_0 \cdot (\tau_{n+1} - \Delta), & \tau_{n+1} \geq \Delta \end{cases} \tag{D.11}$$

Equation D.10 yields an 'integrate and fire' [29] property, where the integrated value is the intensity function of the process that incorporates all the history related constraints of the process and an exponentially distributed random variable is the threshold for firing is in this case.

[1] Although the original method uses uniformly distributed variables, due to the direct relation ($E = -\ln U$) an exponentially distributed variable can be used as well