Lund University

Master's Thesis

# Optimizing Stimulation Strategies in Cochlear Implants for Music Listening

*Author:*
Petra Maretic

*Supervisors:*
Maria Sandsten
Alfred Widell
Søren Kamaric Riis
Manuel Segovia-Martinez

*Examiner:*
Andreas Jakobsson

*20th October, 2015*

Mathematical Statistics
Centre for Mathematical Sciences

# *Abstract*

## Optimizing Stimulation Strategies in Cochlear Implants for Music Listening

by PETRA MARETIC

Most cochlear implant (CI) strategies are optimized for speech characteristics while music enjoyment is significantly below normal hearing performance. In this thesis, electrical stimulation strategies in CIs are analyzed for music input. A simulation chain consisting of two parallel paths, simulating normal hearing conditions and electrical hearing respectively, is utilized. One thesis objective is to configure and develop the sound processor of the CI chain to analyze different compression- and channel selection strategies to optimally capture the characteristics of music signals. A new set of knee points (KPs) for the compression function are investigated together with clustering of frequency bands. The N-of-M electrode selection strategy models the effect of a psychoacoustic masking threshold.

In order to evaluate the performance of the CI model, the normal hearing model is considered a true reference. Similarity among the resulting neurograms of respective model are measured using the image analysis method Neurogram Similarity Index Measure (NSIM). The validation and resolution of NSIM is another objective of the thesis. Results indicate that NSIM is sensitive to no-activity regions in the neurograms and has difficulties capturing small CI changes, i.e. compression settings. Further verification of the model setup is suggested together with investigating an alternative optimal electric hearing reference and/or objective similarity measure.

**Keywords:** cochlear implant, normal hearing model, signal processing, neurogram, compression, N-of-M selection, NSIM

# *Acknowledgements*

I would like to direct a thank you to Oticon Medical for giving me the opportunity to write this thesis. The expertise and knowledge shared from the teams in Copenhagen and Nice has been invaluable for the working process and a very inspiring learning experience. A special thanks to Alfred Widell for taking the time to give his unconditional support and guidance throughout the project and to Søren Karmaric Riis for continuous revision and always brining fresh ideas up for discussion. I would also like to thank my supervisor Maria Sandsten for the helpful comments and valuable feedback throughout the thesis. Last but not least, thank you to family and friends for all the support and encouragements along the way.

# Contents

# CHAPTER 1

# Introduction

## 1.1 Background

The idea of using electricity to stimulate hearing arose already in year 1800 with Alessandro Volta. He conducted an experiment connecting batteries to two metal rods and inserting them in his ear canal through which he managed to create an auditory sensation.

It would take until year 1950 before a direct stimulation of the auditory nerve was performed on a human being. During a neurosurgical operation the Swedish neurosurgeon Lundberg used sinusoidal electric current to stimulate a patient's auditory nerve who perceived it only as noise. The first implant in the cochlea, allowing the auditory nerve to be stimulated by a multiple electrode device, was performed by American surgeons John M. Doyle and William F. House in 1961 [2, pp. 6-9]. The recipients reported auditory percepts in loudness with a change in stimulation level. The observations with the first cochlear implant prototype led to several detailed experiments in the upcoming decades to increase the knowledge in the cochlea functions and sound perception.

During the 70s research refined on clinical applications and implant technology, making the market open up for CIs as a single channel implant was developed and the first to be commercially marketed. In the last half of the 80s a multichannel implant was introduced and immediately became successful with its capabilities of capturing spectral information and speech recognition. Despite the CIs prevalence there is a long way with many years of research to go before their performance can be compared to that of a normally functioning human ear.

Since the most important sound for human communication is speech, research in CIs has for long been focused on designing the CI to emphasize speech performance. Developing the next generation CIs the improvements lie in new signal processing and encoding strategies, bilateral cochlear implants as well as combined electric and acoustic stimulation in order to extract, encode and deliver important acoustic features in different types of environments and situations. Speech covered in background noise, sound localization and sound segregation are some of the targeting points [23].

One area where there has been less directed attention to is perception of non-speech sounds, including music and tonal languages. Many CI users are unsatisfied with their ability to perceive music after implantation. An improved music hearing would not only increase the quality of life for these users but is also believed to benefit understanding speech in noisy environments [16].

This thesis report is written in close collaboration with Oticon Medical in Copenhagen which is a growing global organisation in implantable hearing solutions. As a branch from the R&D team this project works as an initial study in how music appreciation can be improved in the recent CI sound processor Saphyr Neo.

## 1.2 Objective and aims

With the context of the previous background the aim of this thesis is to investigate and optimize the electrical stimulation strategies in cochlear implants for music listening, using a simulation model provided by Oticon Medical. The focus will lie on configuring and developing the simulation model to analyze compression settings and optimally capture the characteristics of music signals in the sound processor. Another approach will be to find an alternative N-of-M strategy to today's selection of the N largest envelope channels.

The simulation chain consists of two paths running in parallel simulating normal hearing conditions and electrical hearing respectively. In order to evaluate the performance of the CI model, the model simulating normal hearing will serve as a reference. We assume that the more similar the CI model output is to the output of the normal hearing, the better is the user's hearing performance. An objective measure will then be applied to quantify the similarity. Another focus area of the project will be to validate the performance and resolution of the objective measure.

## 1.3 Project outline

The three following chapters will give the reader knowledge and understanding of the background theory leading up to the implementation and results. Chapter 2 introduces the human hearing including a deeper insight in hearing devices with an emphasis on cochlear implants. Chapter 3 focuses on the theory behind the simulation model while Chapter 4 describes the main objective evaluation measure. Chapter 5 includes a description of the model configuration for a typical CI user as well as listing the proposed optimizing strategies implemented in the model. The final chapters includes a summary and discussion of the main results and figures in Chapter 6, followed by conclusions and recommendations for future studies in Chapter 7.

CHAPTER 2

# Theory

## 2.1 Human hearing

The first theory section describes the anatomy and functions of the human auditory system followed by a discussion on different types of hearing loss and treatments for hearing impairment.

### 2.1.1 Auditory System

The human auditory system is a sensory system responsible for the sense of hearing. It consists of two subsystems; the Peripheral auditory system including the outer, middle and inner ear, and the Central auditory system where the processed input stimulus is carried further to the brain stem through neural response. The sound travels a complex path through the ear before it is transformed to a transmitter in the synapse and can be perceived by the brain as sound. The emphasis in the following sections will lie on the functions and anatomy of the main parts of the ear, see Figure 2.1.

### Outer ear

Changes in acoustic pressure caused by vibrations of a medium is the definition of what we commonly refer to as sound. The sound waves are first collected by the visible part of the outer ear called *pinna*. The pinna is formed as folds of cartilage and serve as a protection towards the more delicate inner parts of the ear as well as its shape help detect the direction of the sound source. From the pinna the sound enters the external auditory canal (*meatur*) where it is amplified and directed to the eardrum or *tympanic membrane* which starts to vibrate.

Before sound enters the pinna it passes over the torso and head, which due to their shape and structure provide obstacles to the sound, causing spectral and temporal changes in the original sound wave decomposition. To measure the spectral changes of the sound transmission the so called *Head-Related Transfer Functions* (HRTFs) can be used. These transfer functions are describing changes between the source and outer ear in terms of attenuated amplitudes of the spectral components and introduced phase shifts of the originating sound. The HRTFs are important to consider when discussing sound localization.

### Middle ear

The middle ear, separated from the external ear by the cone-shaped tympanic membrane is an air-filled cavity located in the temporal area of the skull. The tympanic membrane is attached to the inner ear via a series of the three smallest bones in the body (*ossicles*) named *malleus, incus* and *stapes*. The bones are suspended in the middle ear cavity by means of axial ligaments and the *tensor tympani-* and *stapedial muscle*. The ossicular chain acts as a lever, converting the lower-pressure tympanic membrane vibrations to higher-pressure vibrations at another smaller membrane called the *oval window* through the translation of sound pressure waves to a mechanical motion of the bones. For the bones to move efficiently the middle ear must not be a completely closed cavity in order for air pressure build up or reduced to be able to release. The *eustachian tube* has the important role of equalization of pressure difference across the tympanic membrane by providing another path for the air. Without the function of the eustachian tube, changes in air pressure that we experience frequently in air planes or under water, would cause the tympanic membrane to move more in one direction than in the other and thus stretching the membrane due to the unequal pressure. This would lead to pressure changes caused by sound waves not to be as successful in vibrating the tympanic membrane.

The base of the stapes pushes the oval window which vibrates in response, causing the dense fluids of the inner ear to move. This mechanical movement is responsible for compensating for the acoustic impedance mismatch between the air and the inner ear fluids, i.e. the inner ear fluid is denser than the air. If the oval window was directly driven by air, the system would lose some of its sensitivity. The compensation is done by increasing the pressure at the oval window which is made possible by the significant size difference between the tympanic membrane and the stapes footplate attaching the oval window.

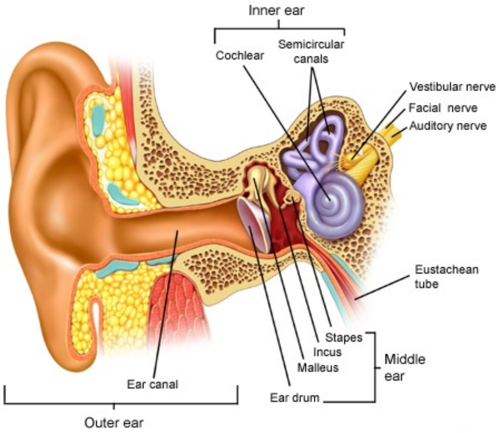


Figure 2.1: Cross-section of the human ear anatomy.

Source: `http://www.healthgalleries.com/anatomy-ear-learning-activity`

### Inner ear

The inner ear can be divided into three sections; the *semicircular canals*, the *vestibule* and the *cochlea*. The semicircular canals and the vestibule together make up the *vestibular system* which is the structure that affects the sense of balance. This structure will not be discussed further in this project. The cochlea is a small snail-shaped structure in the inner ear being the primary sensory organ for hearing where the information in the sound waves are transformed into neural form.

### The cochlea

The cochlea appears as a coiled tube of decreasing diameter with approximately two and a half turns in humans. If it were to be unravelled and stretched out it measures about 35 mm. One fundamental principle of the cochlea is its *tonotopic organisation*, meaning that incoming sounds waves deform the *basilar membrane* (BM) at a position that is specific to the frequency of the vibration. High frequencies cause movements in the *base* (near the oval window) of the cochlea and low frequencies work at the *apex* (at the top of the cochlear spiral) which results in a spatial representation of frequencies along the cochlea [22, pp. 91-103]. See Figure 2.2 for a tonotopic map representation of the cochlea.

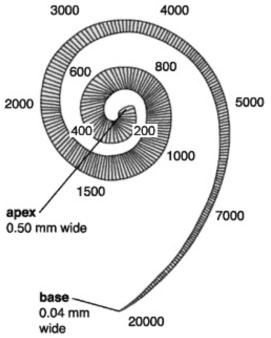


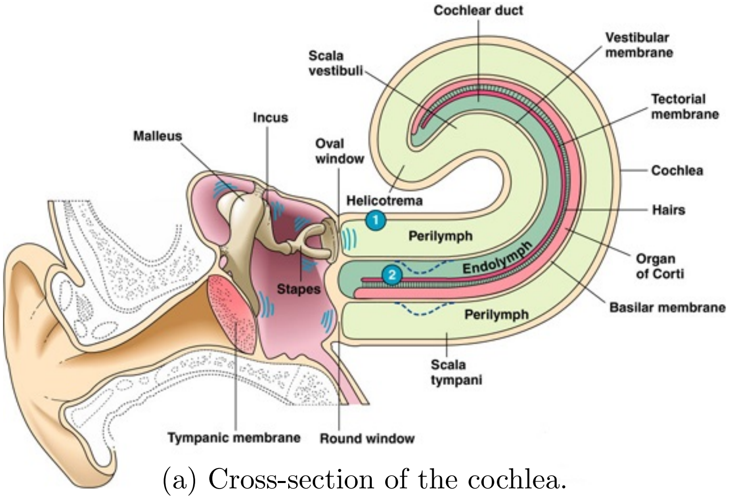
Figure 2.2: Tonotopic represenation of the cochlea. High frequencies cause basilar membrane movements in the cochlear base and low frequencies in the apex.

Source: `https://science.education.nih.gov/supplements/nih3/hearing/guide/info-hearing.html`

The cochlea is spilt into three distinct ducts known as the *scala vestibuli*, the *scala timpani* and the *scala media*. All three sections contain chemical fluids called *perilymh* in the scala vestibuli and scala tympani, containing a low concentration of $K^+$ and high concentration of $Na^+$. The scala media is filled with a very different ionic fluid called *endolymph* that has a high concentration of $K^+$ and low $Na^+$. The lower passage of the cochlear canal, i.e. the scala tympani, has an opening known as the *round window* which is covered by a thin membrane. The *Reissner's membrane* separates the scala vestibuli from the scala media and the BM separates the scala media from the scala tympani, see Figure 2.3a for a detailed sketch of the cochlear anatomy.

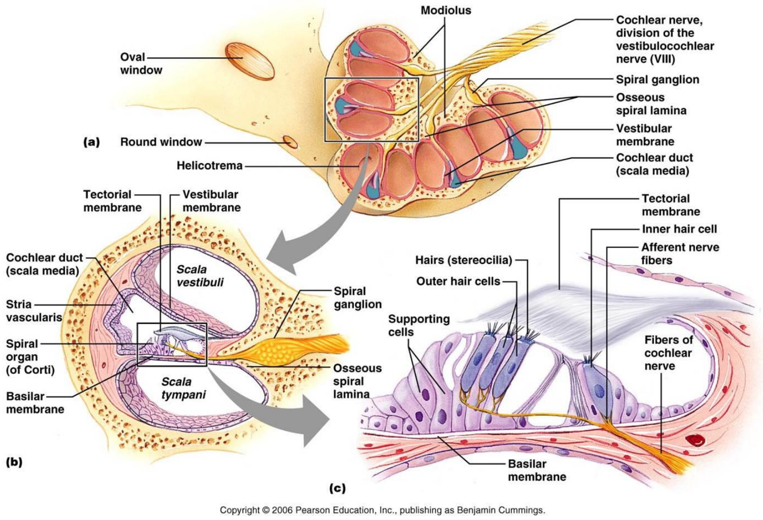
### Organ of Corti

Positioned on the BM and running lengthwise down the cochlea's scala media surface is the *organ of Corti*, see Figure 2.3b. The organ of Corti is composed of *hair cells* which are responsible for generating the nerve impulses required for hearing. The general function of the cochlea is to translate the mechanical vibrations of the stapes and the inner ear fluids into neuroal responses in the auditory nerve. The vibratory patterns of the basilar membrane act as a key factor in this process.



(a) Cross-section of the cochlea.

Source: `http://medicsindex.ning.com/m/blogpost?id=5826870`



(b) Organ of Corti.

Source: `http://healthfavo.com/organ-of-corti.html`

Figure 2.3: Cross-section of the cochlea including a detailed skecth of the organ of Corti.

### Basilar membrane

As mentioned in the previous section the BM is essential for understanding the frequency analyzing ability of the cochlea. The membrane runs from the oval window at the end of the middle ear to a small opening in the apex called the *helicoterma*.

The membrane varies in thickness and elasticity along the cochlear spiral. It is wider and under no tension at the apical end and narrower and stiffer at the base. Each point along the basilar membrane that is set in motion due to sound waves vibrates at the same frequency as the stimulus. However, some locations of the membrane respond more strongly to the stimulus than others depending on the stimulus frequency and input level. The narrow and stiff basal end will resonate maximally to high frequencies (short wavelengths) while the less tightly stretched wide apical end is resonating most strongly to low frequencies (long wavelengths). Thus, the membrane's natural frequency of vibration decreases as the membrane becomes wider and more flaccid.

The entire motion that occurs on the BM in respond to sound can be seen as a travelling wave. An example of this at an instant in time can be studied in the top image in Figure 2.4. Another instantaneous vibration pattern for three successive given times is seen in the bottom image in Figure 2.4, together with the envelope where the maximum displacement is determined by its peak value. The travelling wave motion is in fact an alternation of upward and downward displacement of the BM, propagating from the base to the helicoterma at the apex. How far the wave travels depend on the stimulation frequency; lower frequencies travel further and stimulate both basal and apical end whereas higher frequencies only stimulate the basal end. Worth noting is that the motion of the BM is not entirely linear. One consequence of the non-linearity is that the membrane displacement may not be completely linearly related to the input stimulus level [14, pp. 24-34].

If two different frequencies are received by the cochlea simultaneously they will each create a maximum displacement at different points along the basilar membrane. The separation of complex signals into different maximum displacement points means that the membrane is performing a type of spectral analysis. The membrane will move up and down at different amplitudes in synchrony with vibrating stimulus creating a temporal pattern of displacement following the one of the incoming sound. The input frequency that causes the highest amplitude displacement along the membrane is called the *characteristic frequency* (CF) of that particular location.

The relationship between the CFs of a signal and the position in the cochlea was developed empirically by Greenwood in 1961 [6]. The frequency-position function is described as

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

where

$F$ = characteristic frequency of the sound [Hz]

$A$ = scaling constant between the CF and upper frequency limit of the specific species = 165.4 average for humans.

$a$ = slope of the straight-line section of the frequency-position curve = 0.06

$x$ = length of the cochlea measured from the apex to the region of interest [mm]

$k$ = integration constant = 1

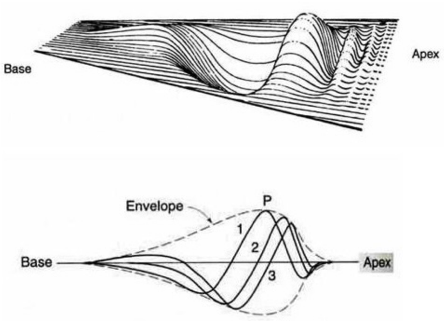


Figure 2.4: Basilar membrane vibration in response to a travelling sound wave.

Source: `http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/pitch/pitch.html`

### Hair cells

The hair cells are arranged in four rows in the organ of Corti along the whole cochlea coil. There are two types of hair cells with different functions, three rows of outer hair cells (OHCs) and one row of inner hair cells (IHCs) which also has various supporting cells. The upper surface of the hair cells are tiny hair like projections called *stereocilia* and a smaller part is the basal body called *kinocilium*. Depending on the type of hair cell it contains different number of stereocilias in each bundle, ranging from 40-150. In humans, there are about 12,000 OHC each with approximately 140 stereocilia and 3,500 IHC, each with about 40 stereocilia [14, pp. 34 -35] The cilias are arranged in parallel graded rows with the shortest stereocilia on the outer rows and the longest in the center. This formation is an important anatomic feature as it allows tuning capability of the hair cells into the chain of chemical events that take place within the cells. The tips of the tallest row of cilia of each outer hair cell are holding a structure of fibres known as the *tectorial membrane*, see Figure 2.3b. The tectorial membrane moves in response to sound vibrations.

Despite the significantly larger number of OHC than IHC positioned along the BM it is the latter ones that have a crucial function to convey acoustical information into neural information, as most of the fibres in the auditory nerve connect to the IHCs. Hence, the IHCs act as transducers of sound vibrations from the BM to electrical activity in the nerve fibres. The OHCs main function is instead to amplify the motion of the BM by feeding back mechanical energy and by that increasing the system's sensitivity.

### Auditory nerve

The auditory nerve consists of a bundle of nerve fibres, or *neurons*, connecting to the previously described IHCs along the entire BM through a synapse. In humans, the auditory nerve bundle averages around 30,000 fibers. The main purpose of the auditory nerve is to carry the information from acoustic stimulus further to the auditory cortex in the brain.

There are two basic types of nerve fibers: *afferent* and *efferent*. Afferent fibers are sensory nerves carrying information from the organ of Corti to the brain. Efferent fibers typically have the opposite direction of activity flow. Most of the afferent neurons are said to have a many-to-one connection to the IHC since each IHC may be innervated by 16- 20 type I afferent neurons. A smaller part of the afferent fibers type II are said to have a one-to-many connection as they innervate the outside row of the OHCs.

Each auditory nerve fibre responds only to a narrow range of frequencies, matching the vibrational pattern of the BM. The neuron activity is initiated by the hair cells which produces so called *receptor potentials* causing a change in the chemical concentration of the cell. The liberation of the chemical transmitter initiates an action potential at approximately 150 mV to be emitted through the auditory nerve fibers that innervate the base of the hair cell. The action potential of a single neuron is called *spike* and the rate of which spikes are fired is proportional to the velocity of the basilar membrane.

Characteristics of the auditory nerve fibers are regularly described in terms of *spontaneous rate*, *threshold* and *tuning curves*. The first property is defined as the neural activity (discharge) that occurs when no stimulus is present, i.e in a sound-free environment. The spontaneous activity is expressed in spikes/second of a single neuron. The low spontaneous rate (LSR) fibers discharge less than 0.5 times per second and constitute approximately 16 % of all nerve fibers. About 61 % fibers have high spontaneous rates (HSR) (>18 spikes/second) and the remaining 23 % of the fibers groups under medium spontaneous rates (0.5-18 spikes/second). The threshold of a neuron is defined as the lowest sound level at which a change in response of the neuron can be measured. HSR are usually associated with low thresholds and vice versa. The acoustic tuning curves shows the frequency selectivity of each auditory neuron. The curve shows the sound intensity that will cause the fiber respond as a function of frequency. The lowest threshold of the tuning curve is referred to as the characteristic frequency [14, pp. 38 -40].

Another mechanism of the auditory nerve fibers is their ability to lock onto the phase of certain input stimulus and fire action potentials. Phase locking encodes temporal structure of stimuli and is generally used in the context of pure tones where the auditory fibers will then fire at the same frequency as the tone. The quality of phase locking decrease with increased frequency. For frequencies below 1 kHz nerve firings are in synchronous with input stimulus but becomes progressively inaccurate at higher frequencies (1-5 kHz) [14, pp. 44-50].

### 2.1.2 Hearing impairment

Hearing impairment can be broadly categorized in two main types by which part of the auditory system is damaged; *conductive hearing loss* and *sensorineural hearing loss*. The degree of hearing loss is defined as the minimum detectable level (threshold in dB) of a sound relative an average hearing threshold specified for ”healthy” listeners. A commonly used classification system includes the levels: mild (26- 40 dB), moderate (41- 70 dB), severe (71- 90 dB) and profound (91 dB+).

### Conductive hearing loss

Conductive hearing loss usually occurs when there is a problem with the ear canal, ear drum or middle ear that reduces the sound transmission to the cochlea. Three common causes are fluids in the middle ear as a result of infection, stapes immobilization as a result of bone growth over the oval window and wax in the ear canal [22, pp. 61-63]. Typically the cause of conductive hearing loss can medically treated to restore hearing partially or completely. Following medical treatment hearing aids are usually effective in correcting the remaining hearing loss.

### Sensorineural hearing loss

Sensorineural hearing loss, also known as "nerve deafness", most commonly occurs from a defect in the cochlea (sensory) like poor hair cell function in the organ of Corti or a defect in the auditory nerve (neural). Hearing loss due to abnormalities higher up in the auditory system is known as *retrocochlear loss*. People suffering from sensineuroal hearing loss often cannot be treated solely by using hearing aids. For more severe hearing loss or even complete deafness, cochlear implants are often an effective solution [22, pp. 61-63].

## 2.2 Hearing solutions

This section summarizes the different types of hearing solutions available on the market. Which treating option to chose depends on the cause, severity and time course of the hearing loss.

### 2.2.1 Hearing aids

Hearing aids are devices whose main function is to amplify sounds. Most hearing aids are built up by similar components including a microphone picking up the sound, an amplifier with gain control and a loudspeaker. Examples of four different styles of hearing aids are the Behind-The-Ear (BTE) aid, In-The-Ear (ITE), In-The-Canal (ITC) and Completely-In-the-Canal (CIC) aids. Hearing aids are usually recommended to people suffering from mild to severe hearing loss.

### 2.2.2 Bone anchored hearing systems

The bone anchored hearing device is an implantable solution utilizing the body's natural way of conducting vibrations through the skull to the inner ear and thus bypassing the damaged parts of the external auditory canal and middle ear. These devices work well for people with conductive hearing loss, single-sided deafness (the "head-shadow" frequency effect is treated by routing the signal via bone conduction to the opposite cochlea) and people with mixed hearing loss.

### 2.2.3 Middle ear implants

Another alternative to conventional hearings aids are the middle ear semi-implantable solution which directly vibrates the small bones in the middle ear, bypassing the ear canal and tympanic membrane. These devices are an option for people suffering from moderate to severe hearing loss and are also an alternative for people who cannot use hearing aids because of medical reasons or in any other way are dissatisfied with their hearing aid.

### 2.2.4 Cochlear implants

Cochlear implants are also examples of implantable hearing solutions with a complex electronic device. The device is based on direct electric stimulation of the auditory nerve fibers in the cochlea, bypassing the damaged sensory hair cells. Cochlear implants are typically recommended for patients suffering from moderate to profound hearing loss including complete deafness. A throughout description of the cochlear implant components and signal processing encoding strategies are listed in the next section.

## 2.3 Cochlear Implants

Cochlear implant devices are based on the idea that there are enough auditory neurons left for stimulation, despite the loss of hair cell function. The hair cells can thus be bypassed by direct electrical excitation of the auditory neurons, improving hearing for people suffering from a profound hearing loss.

Generally a CI consists of an external part and an implantable part. The external part includes a microphone, a sound processor and a transmitting coil. The internal part consists of a decoder, a receiving coil and an electrode array. The incoming sound is picked up by the microphones and processed by the sound processor located in the BTE device. The encoded sound is then transmitted through the coil attached to the skull by a magnet to the implanted receiver. The implant is responsible for converting the encoded sound to electric pulses, sent to the electrode array in the cochlea. Activated electrodes stimulate auditory nerve fibers located at the vicinity of the the electrodes allowing propagation of neural impulses further to the brain. In Figure 2.5 the main parts of a cochlear implanted device are illustrated.

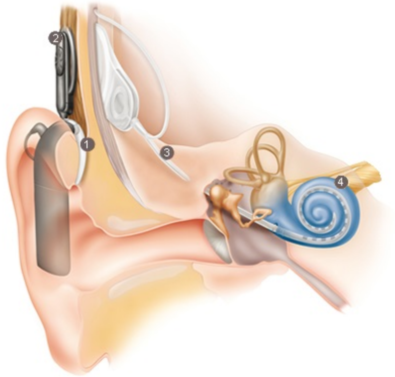


Figure 2.5: The main parts of a cochlear implant; 1. BTE sound processor. 2. Transmitter (external)/ receiver(internal) coil. 3. Implant. 4. Electrode array.

Source: `http://www.cochlear.com/wps/wcm/connect/au/home/understand/hearing-and-hl/hl-treatments/cochlear-implant`

### 2.3.1 Hearing with cochlear implants

It is important to note that a cochlear implant does not restore normal hearing but instead give useful representation of sound to a hearing impaired person. However, there are today a lot of restrictions on the electrical representation of sound in a CI compared to the sounds a normal functioning ear pick up. Two important limitations to consider is the limited dynamic range and the loss of stochastic nerve firings.

The normal frequency hearing range is commonly given as 20- 20,000 Hz. The frequency range of a CI patient is limited by the length and insertion depth of the electrode array. Depending on the patient's shape and size of the cochlea the implantation success differ from patient to patient. Due to the narrow passage in the apex patients often experience unsatisfactory in low frequency hearing.

A patient's dynamic range is defined as the range in electrical amplitudes between a barely audible threshold level(T) and loudness uncomfortable level(C). The dynamic range is expressed in terms of dB and for normal hearing the interval scales between about 1- 140 dB [22, pp. 27-28]. In acoustic hearing the dynamic range may be 30 dB wide for conversational speech whereas CI users may have a range as small as 5 dB [10].

Another difference is the lack of stochastic effects in the auditory nerve firings among CI users due to the bypassing of the damaged hair cells. Electric stimulation excites neurons in a highly synchronous way whereas different rates of spontaneous nerve firing are present in acoustic hearing.

### 2.3.2 Electrode design

The design of CI electrode arrays is a highly focused researched area. Some of the associated issues are electrode placement, number of electrodes and electrode configuration [10].

Commonly, the electrode array is inserted in the scala tympani to bring the electrodes in close proximity with the auditory neurons along the organ of Corti. Typically the electrode array can be inserted 22- 30 mm from the base of the cochlea [10]. Examining the electrode position, Greenwood's frequency-position function is used to estimate the CFs of the CI [6].

The number of electrodes as well as their spacing affects the place resolution for coding frequencies. Optimally, the larger the number of electrodes the finer place resolution to the corresponding frequencies. However, the design of the electrode array is constrained by two inherent factors; the spread of excitation in response to electrical stimulation and the number of surviving neurons at a particular location in the cochlea. Thus, including a large number of electrodes will not generally result in better performance.

### 2.3.3 Stimulation

There are generally two types of stimulation of electrodes depending on how information is presented. The stimulation is referred to as analog if the acoustic waveform itself is presented simultaneously to the electrodes in analog form. One disadvantage of the analog approach is that channel interactions may occur due simultaneous stimulation.

The other stimulation strategy is referred to as pulsatile as the information is delivered to the electrodes using a pulse train. In this case, the pulses can be delivered in a non-overlapping way at a certain pulse rate and thereby minimizing channel interactions. The rate at which the pulses are delivered to the electrodes has been found to affect speech recognition performance [10].

Pulsatile electrodes can be configured in different ways to deliver stimulation; monopolar, bipolar and common ground, see Figure 2.6. The monopolar mode comprises one intracochlear active electrode and several extracochlear electrodes that are located further away. These electrodes serve as a return current path for several discrete active electrodes. In bipolar stimulation currents are passed between an active electrode and a return electrode in the cochlea. The third configuration type can be seen as an intermediate between the monopolar and bipolar configurations. The common ground stimulation allows one active electrode and several or all the remaining electrodes to be used as return path for the current [12].

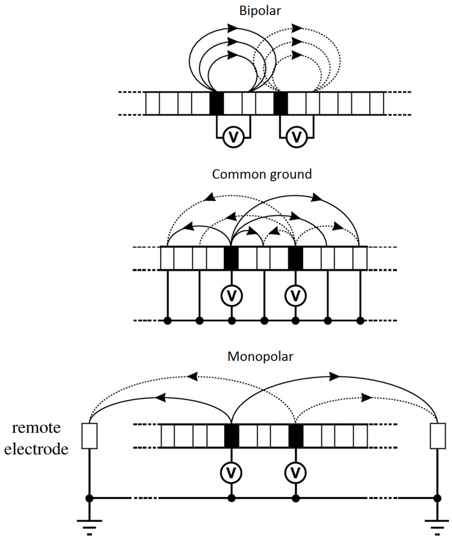


Figure 2.6: Sketch showing the current flow for three different type of modes of stimulation.

### 2.3.4 Signal Processing

The encoding strategy in the sound processor plays an extremely important role in maximizing the CI users overall speech perception and communicative ability. There are a number of strategies that have been put forward over the years for transforming the incoming audio signal to electrical stimuli and determining which electrodes should be activated at each time. The aim for many such strategies is to encode the signal to mimic the firing patterns of a healthy auditory system.

In the early 1970s the first single-channel implants were implanted in human patients. These implants used a single electrode for stimulation and did not exploit the place coding mechanism in the cochlea for encoding frequencies. When stimulating a single site in the cochlea the temporal encoding of frequencies was restricted to 1 kHz due the refractory period of the auditory neurons. However, despite the poor transmitted frequency information a handful of patients did experience some speech perception [10].

The modern multichannel implants were first introduced in 1980s and has been designed to provide electrical stimulation at multiple sites in the cochlea using an electrode array. Different electrodes are thus stimulated depending on the frequency content of the signal. Electrodes close to the base of the cochlea are stimulated with high frequency signals while electrodes near the apex are stimulated with low frequencies.

The number of electrodes in multi-channel implants differ among the manufacturers. Some devices uses a large number of electrodes (up to 22) but stimulate only a few in each cycle, while other devices use only 4-12 electrodes but stimulate all of them. Depending on how the signal information is extracted and transmitted to each electrode, the signal processing strategies can be divided into feature-extraction strategies and waveform strategies.

The first devices coding for spectral information employed a feature extraction approach which was based on the extraction of formants from the sound signal. As the name suggests, the F0/F2 strategy is based on the extraction of the fundamental frequency (F0) and the second formant (F2) using zero crossing detectors. The F0/F2 strategy was later modified to include also the first formant F1. F1 and F2 convey information about the identity of vowels and other voices speech sounds. Further refinements of the feature extraction algorithm were included in the MULTIPEAK (MPEAK) strategy where high-frequency information could be captured [12].

The development of spectrum estimating pulsatile strategies led to the abandonment of the feature extraction techniques. One such approach called Compressed-Analog (CA) aims at delivering band-specific amplitude-compressed analog stimulation to different electrodes after the signal has been filtered into distinct frequency bands. A drawback of this method is that the analog waveforms are delivered simultaneously to the electrodes which causes channel interaction and may distort spectrum information [10].

The Continous Interleaved Sampling (CIS) method addressed the interaction problem by delivering non-simultaneous interleaved pulses. The pulse amplitudes are derived by extracting the envelopes of band passed waveforms [10]. Another spectrum-estimating scheme is the Spectral Peak (SPEAK) algorithm which has a larger number of bandpass filters than CIS and N of M ($N < M$) envelope channels are selected for stimulation in each period at a rate of 250 Hz per channel. The Advanced Combinational Encoder (ACE) follows the steps of SPEAK but includes a much higher stimulation rate (14.4 kHz) [10].

### 2.3.5 Music enjoyment

Speech and music may be considered to be two very different signals in the way that they operate on acoustical principles [21]. The perception of music for CI users rests on the assumption that music can be categorized as a sequence including fundamental frequencies (perceptional: pitch), rhythm, melody and timbre [12]. Often the transmission of fine temporal frequency information is poor in CIs due to the large excitation spread. An improvement of pitch perception is believed to improve the perception of music which in turn can lead to improved quality of life for the users and additional improved speech perception in noise [16].

CHAPTER 3

# Models

This chapter gives a technical description of each of the two simulation chains running in parallel. Both the normal hearing path and the CI model are extended with a point process model called the Goldwyn model which is simulating the auditory nerve firings. The final model output is represented by neurograms which are to be compared between the normal hearing reference and the CI model. A block diagram of the two independent simulation paths is found in Figure 3.1. Each model description is followed by the necessary changes made to create a common interface for comparison purposes.

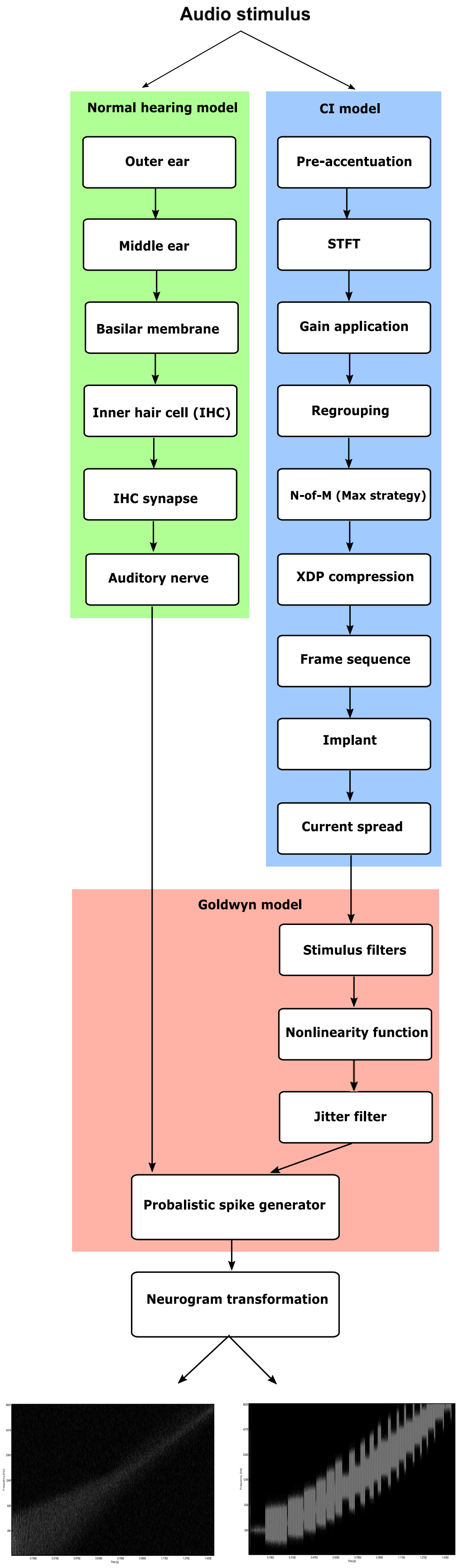


Figure 3.1: Block diagram of the normal hearing chain and electrical hearing chain running in parallel and extended with the Goldwyn model to finally generate comparable neurograms.

## 3.1 Auditory Periphery Model

The Matlab Auditory Periphery (MAP) is a computational model simulating all stages of the auditory periphery, from the outer and middle ear up to the auditory nerve and the brainstem. The model design is based on measurements from human patients [13]. The model can be used among several things to process acoustic waveforms to generate representation at different levels in the auditory periphery and demonstrate physical phenomena such as absolute threshold. In this project it is used to simulate normal hearing conditions.

A schematic view of the MAP path can be seen in Figure 3.2 where each stage is simulated by computational formulae. Observe that the model will be interrupted after the stage called "Auditory nerve" and brain cell responses will not be taken into account.

The model input is a mono acoustic pressure waveform expressed in Pascals and evaluated in 10 ms long time frame segments.

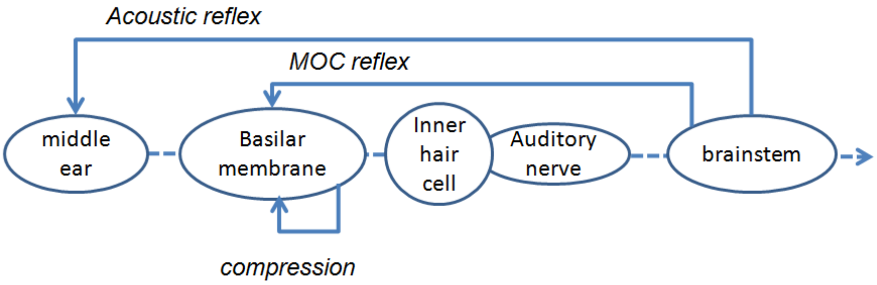


Figure 3.2: Implemented sections of the auditory periphery path and parameters can be found in [13].

### 3.1.1 External ear

In the first block of the model two independent band-pass filters, representing the ear canal resonance and the concha [1] resonance respectively, are applied to an input sound pressure wave. The output from the filters are summed and applied on the original sound wave.

The output from this substage and thus input to the next is the sound pressure at the tympanic membrane. The stapes response in the middle ear is modelled as a displacement measure where the stapes velocity is first represented as being proportional to the sound pressure. To convert the velocity to displancement, a frequency variable is introduced where a doubling in frequency results in a halving of displacement. In practice, a low pass filter is applied resulting in matching human stapes measurements at frequencies exceeding 2kHz. To limit displacements at lower frequencies a high pass filter is introduced.

### 3.1.2 Basilar membrane

The input to the next stage which models the basilar membrane movement is the stapes displacement. To simulate the physical displacements of the basilar membrane a *Dual-Resonance-Non-Linear* (DRNL) filter is used. The filtering technique models the basilar membrane at discrete locations, each identified by its logarithmically spaced *best frequency.* The term best frequency is equivalent to the previously described characteristic frequency and here referred to as the frequency generating the greatest basilar membrane response close to the hearing threshold. The most responsive frequency however changes with the level of input stimulus.

[1] The concha is the part of the external ear nearest the ear canal.

The DRNL filter bank serves as a powerful model of the human non-linear cochlear behaviour consisting of two branches, one linear and one the order non-linear, with different band-pass responses. A schematic overview of the DRNL stimulation at a single location can be seen in Figure 3.3.

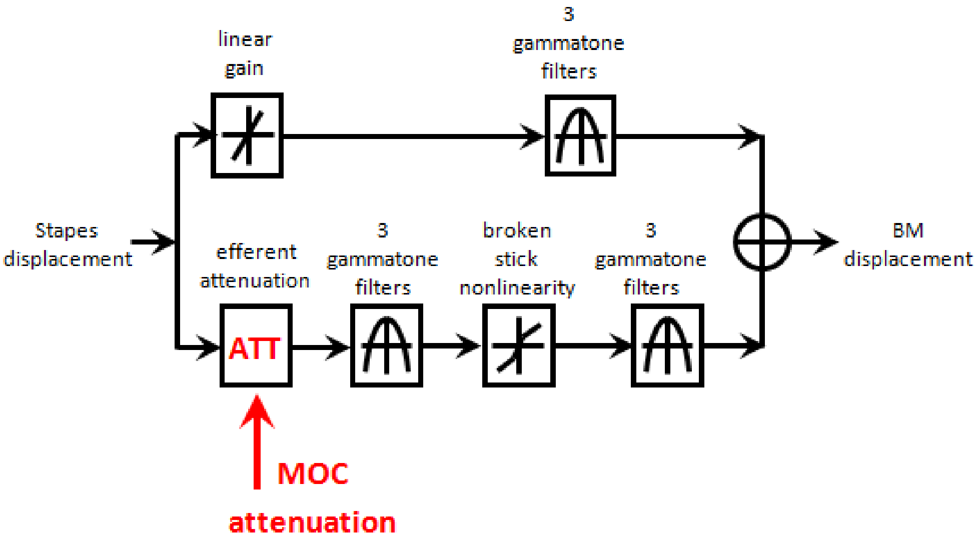


Figure 3.3: A schematic description of the DRNL filter simulating the displacement of the basilar membrane at a specific location in the cochlea.

The non-linear path (lower in Figure 3.3) starts with an attenuation of the input signal by a variable scalar representing the MOC reflex. The MOC response is based on the sum of all three firing rates (HSR,MSR, LSR) at a particular best frequency channel.

Following the attenuation are several gammatone filters with center frequencies close to the best frequencies at the respective location and increasing band widths with the best frequencies. Further, a so called broken stick compression function with linear response below a compression threshold is applied to the signal before gammatone filters are applied once again.

The parallel linear pathway consists of a scalar representing attenuation (or gain) and similarity to the non-linear pathway, a cascade of gammatone filters. However, there is an important difference between the paths being the non corresponding center freqencies accounting for observed shifts of best frequencies of the DRNL filter for higher stimulus level.

Finally the output of each pathway is summed to produce an output modelling the displacement of the cochlear partition (basilar membrane and organ of Corti) at individual locations. Evaluation of the DRNL filter modelling shows that its output results accurately matched iso-intensity curves from experimental data, see Figure 3.4.

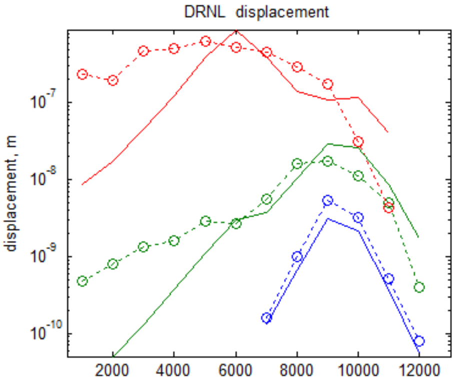


Figure 3.4: Iso-intensity contours showing the intensity of a sinusoid required to produce a spike rate in the neuron as a function of the frequency of the sinusoid. The iso contours are shown for three different sound levels (dB SPL). The solid curves represent the DRNL filter modelling by the MAP model and the dotted lines with open circles experimental chinchilla measurements.

### 3.1.3 Inner Hair Cell

The next stage of the model is simulating the inner hair cell response taking the basilar membrane displacement as input and producing an IHC stereocilia receptor potential change. The process can be divided into two sub stages being the change of conductance in the stereocilia and the receptor potential changes in the cell body.

#### Stereocilia conduction changes

As the BM moves in response to stimulation, it indirectly causes the IHC stereocilia displacement through coupling with the tectorial membrane located above the IHC. The movements of the tips of the steoereocilia modifies the conductivity of the local ion channels by either depolarization (influx $K^+$, $Ca^{2+}$) or hyperpolarization (outflux $K^+$) of the hair cell, see Figure 3.5. The formula of the procedure represents a high-pass filter and follows as

$$\tau \frac{du(t)}{dt} + u(t) = \tau \; C_{\text{cilia}} \; \text{disp}_t$$

where $\tau$ is a time constant, $C_{\text{cilia}}$ is a scalar converting BM displacement, $\text{disp}_t$, to cilia displacement, $u(t)$. The cilia displacement further determines the apical conductance $G(u)$.

#### Receptor potential changes

The membrane potential of a IHCs body is referred to as V(t) and is modelled with an passive electric circuit with the transfer function

$$C_m \frac{dV(t)}{dt} + G(u)[V(t) - E_t] + G_k[V(t) - E'_k] = 0$$

where $C_m$ is the cell capacitance, $G_k$ is a fixed membrane conductance $E_t$ is the endocochlear potential and $E'_k$ is the reversed potential of the basal current. An important characteristic of the transfer function is that it is asymmetric at high stererocilia displacements, meaning that a negative cilia displacement causes only a small shift in the receptor potential whereas a positive displacement gives rise to higher potential shifts.

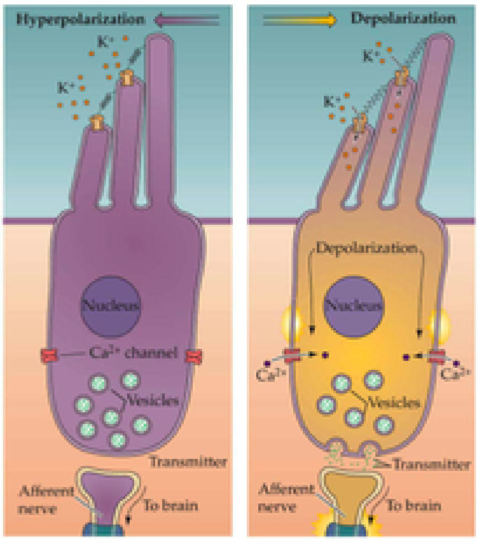


Figure 3.5: Hair cell motion, hyperpolarization (left) and depolarization (right).

Source: `http://www.rci.rutgers.edu/~uzwiak/AnatPhys/Audition.htm`

### 3.1.4 Auditory Nerve

The IHC receptor potential influences the firing of the neurons in the auditory nerve through controlling the concentration of calcium ions in the synaptic region. The process is treated in two stages where the calcium influx is first modelled followed by modelling the neurotransmitter release to the synapse.

### Ca$^{2+}$ influx

As the receptor potential rises due to stererocilia displacement, calcium ions (Ca$^{2+}$) flows into the cell causing the the small packets called vesicles located near the synapse to have an increased probability of generating an action potential. The calcium current $I_{Ca}$ is based on the membrane potential V(t)

$$I_{Ca} = G_{Ca}\ m^3_{I_{Ca}}(t)[V(t) - E_{Ca}]$$

where $G_{Ca}$ is the maximum calcium conductance when all the calcium channels are open, $m_{I_{Ca}}(t)$ is the fraction of all open channels and $E_{Ca}$ is the reversal calcium potential.

### Calcium concentration

The pre-synaptic calcium concentration is modelled as function of calcium current

$$\frac{dCa^{2+}(t)}{dt} = I_{Ca}(t) - Ca^{2+}(t)/\tau_{Ca}$$

where $\tau_{Ca}$ is a time constant reflecting the time from the calcium canals open and the vesicle release in the synapse. The value of $\tau_{Ca}$ controls the release characteristics of the synapse as it varies according to the spontaneous rate of the neurons which are allowed to operate in parallel.

### Probabilistic model

Vesicle release holding the neurotransmitter can in the MAP model be simulated both in an quantized way, modelling individual vesicles, or in a probabilistic manner, the latter one being more approximative however less computationally heavy.

In response to the released neurotransmitter is the auditory nerve where a model assumption is that a single vesicle transmitter release is sufficient to trigger an action potential in the auditory nerve fibers. The auditory nerve firing rate simulations are based on the on the quantity of transmitter of the synaptic cleft between the cell body and the neuron.

### 3.1.5 MAP modifications

Several modifications were made in the original MAP model as a part of the former thesis work by Attila Fráter [4].

The first change in the MAP model is related to the best frequencies corresponding to discrete locations along the basilar membrane. In section 3.1.2 it was described that the positions are by default logarithmically spaced between two values. However, in [6] it is suggested that the Greenwood function could be used to accurately map linearly spaced basilar membrane positions to corresponding cochlea frequencies.

The second deviation from the original configuration is the exclusion of the attenuation feedback from the brainstem. Neither the acoustic reflex nor the MOC reflex are included in the model as it is connected to Goldwyn before reaching the brain stages. The consequence of the simplification is that the efferent activity of the system in not be modelled.

## 3.2 Cochlear Implant Model

Running in parallel with the MAP model is the Oticon Medical CI model developed under the ABCIT Project, simulating each consecutive stage of electrical cochlear stimulation in a 2D cochlear model. The model takes an audio file as input and generates an electrodogram with corresponding intra cochlear electric field. An overview of the implemented platforms can be found in the right path in Figure 3.1. The main blocks are the BTE signal processor, which is the area of focus in this project, and the implanted part of the device.

### 3.2.1 Signal processing strategy

The signal processing part of a cochlear implant system plays an important role in how the recipient perceive the incoming sound. Today, the BTE processor can be equipped with a wide range of additional features depending on the lifestyle and demands of the user, for example noise cancellation and wireless connections.

In the first block of the Oticon Medical BTE Saphyr Neo sound processor the audio signal is acquired by the processor microphone and then applied to a pre-accentuation filter modelling the hear related transfer functions discussed in section 2.1.1. The filter output is passed on to the short-time Fourier transform where the signal is transformed from time to frequency domain, see Figure 3.6.

In the next step the STFT-transformed signal is passed through an envelope detector and regrouped according to the predefined frequency bands corresponding to each electrode's channel. The regrouping configuration is patient specific and can vary for each electrode. The N-of-M block represents the pulsatile CIS strategy but where only N (N < M) electrodes are stimulated in each time frame using non-simultaneous, interleaved pulses delivered at a constant rate at 520 Hz. It is worth mentioning that the design of the Oticon Medical sound processor allows stimulation of electrodes in either ascending (apex to base) or descending order (base to apex). This restriction will have an impact on the spectral content of the signal.

Finally, the N envelope outputs are level estimated and then compressed through the XDP compression function which aims at transforming each electrode's acoustic dynamic range to electrical stimulation used to modulate biphasic pulse trains in the electrode array [17].

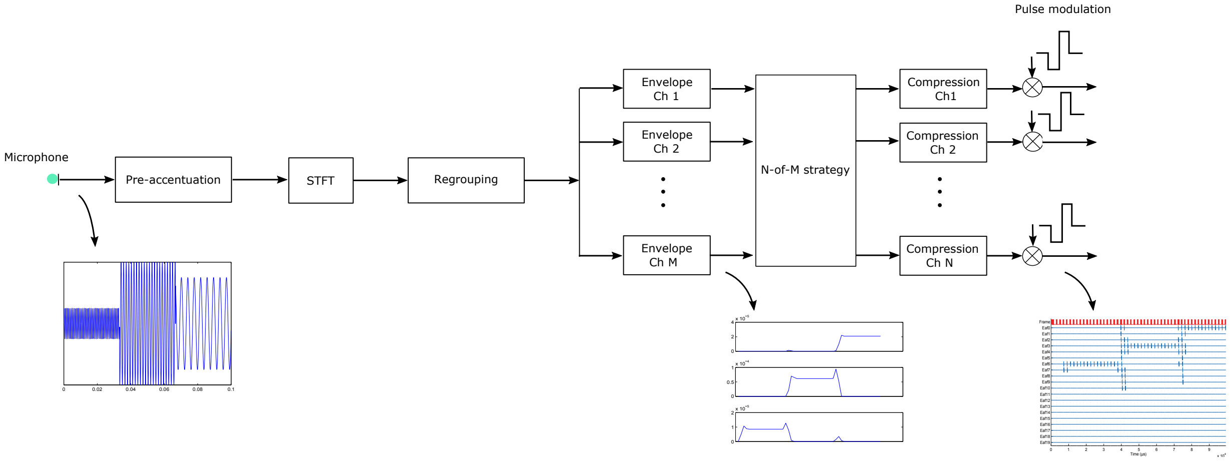


Figure 3.6: The main blocks of the Oticon Medical signal processing strategy.

## 3.2.2 Implant

The stimulation pattern computed from the signal processor is transmitted to the implantable part. In this implanted stage the information is decoded and current pulses are generated through the electrode array, which in the next phase serves as input to the electrode-auditory nerve interface. The output from the implant is a stimulation frame structure including both time and amplitude information of the current pulses.

The implantable part of the cochlear device is considered more or less a fixed structure (due to the nature of the actual hardware) in contrary to the BTE part where parameters variable and different configurations tested.

### 3.2.3 Intra-cochlear electric field (ICEF)

Connected to the implant block is the platform modelling the intracochlear potential map estimating the electrical field generated from the current spread, from each point source known as the electrodes. The potential field problem is solved analytically by a partial difference equation whose solution is a series of modified Bessel functions. An approximation which can be used for the actual implementation of the current spread is a decaying exponential function. This is justified by the effect produced by the tapering of the cochlear ducts.

The exponential spatial filter is given by

$$\Phi(x) = [x - x_i]\ e^{-(x-x_i)/\tau_a} + [x_i - x]\ e^{-(x_i-x)/\tau_b}$$

where $\tau_a$ is a spatial decay constant along the apical axis and $\tau_b$ along the basal axis. $x$ represents the cochlear position in mm and $i$ stands for the electrode index. $[x]$ denotes the threshold linear function as $x$ if $x > 0$ and 0 otherwise.

### 3.2.4 Model modifications

To match the mapping from the MAP model, the Greenwood function is used to model the cochlear frequency-position interface for the electrical hearing chain. Several additional modifications are made in the CI blocks to make the neurogram output look as expected. These changes will not be described in detail but includes gain adjustment along the complete CI path and configuring the current spread parameters.

## 3.3 Connection of models

When including the modifications in each simulations chain their respective output needs to be matched before entering the Goldwyn model. The output from the MAP model includes probabilities of the synapse transmitter release in the auditory nerve for each best frequency. The CI model produces current values at each best frequency as modelled by the implementation of the ICEF. Hence, each model results in slightly different outputs and will therefore be entering the Goldwyn model at different stages. The MAP model connects to the spike generation stage in Figure 3.7, bypassing all the previous stages of the chain. As for the CI model it enters the Goldwyn model with current input in the form of biphasic pulse trains expressed in $\mu s$.

## 3.4 Goldwyn Model

Both the MAP and CI model are extended with a point process model used to generate auditory nerve patterns in response to the electrical stimulation from the cochlear electrode array. The stochastic Goldwyn model is completely defined by its conditional intensity function, simulating the response of a single auditory nerve using a cascade of linear and non-linear stages and producing a spike pattern as output [5].

The Goldwyn model is based on fundamental statistics from recordings of physiological data from cats and accounts for acoustic threshold, jitter, refractory period and summation effects. Below sections give an overall description of the point process' framework and parameters.

The *firing efficiency curve* is a function that relates the current level of a single pulse to the probability of a nerve firing. It can be approximated by a Gaussian distribution. In close relation to the efficiency curve is the threshold of a neuron, defined as the half the probability of a stimulus eliciting a spike. A measure of variability of the spike initiation is the *relative spread* which is defined as the standard deviation of the underlying Gaussian distribution divided by its mean.

For longer pulse duration the threshold current level is typically smaller than for a single pulse of stimulation due to the neuron capacity to integrate current over time. The dependence of pulse duration on threshold is described by the *chronixe*.

An additional stochastic measure to the firing efficiency curve (spike initiation variability) is the timing of spikes described by *jitter*. Jitter depends both on pulse duration and pulse level but in this model used as the value measured for a pulse at spiking threshold.

For the model to generate realistic results it is crucial to include the effect of previous spikes. One such historic effect is the *refractory effect* which reduces the excitability of a neuron immediately after firing. The refractory period is implemented as an increase in the threshold following a spike.

Another implemented feature in the Goldwyn model is the summation effect of several consecutive pulses. This effect is relevant for high carrier pulses and accounts for multiple pulses being more likely to evoke a spike than single pulses acting independently on the neuron.

A schematic view of the model can be studied in Figure 3.7 and summarized as follows: An incoming biphasic pulse train $I(t)$ is passed through several linear filters, accounting for stimulus and spike time variability, and a non-linear function. The filter output defines the instantaneous probability of spiking which is then used to generate a random sequence of spike times. Previous spikes provide feedback to control the the stimulus filters and non-linearity.

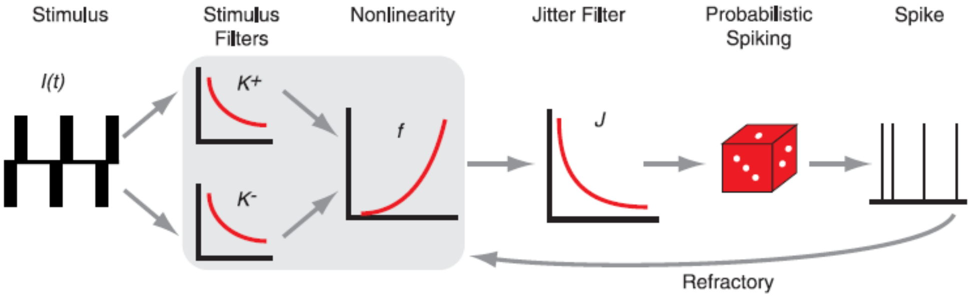


Figure 3.7: A schematic diagram of the Goldwyn point process model modelling the auditory nerve fibers response to electrical stimulation.

## 3.5 Neurograms

A commonly used graphics tool to represent the neural response of the auditory nerve is the neurogram [7]. It presents the intensity of neural activity from multiple auditory nerve fibers in the time-frequency domain, completely analogous to a spectrogram. The frequency axis shows the Greenwood -spaced characteristic frequencies between a lower and upper frequency bound. The neurogram is created by analyzing each row in the spike pattern output from the Goldwyn model. Within each time frame, spikes are accumulated and smoothed by convolving with a 50% overlapping Hamming window.

CHAPTER 4

# Objective measures

As described in the previous section, both the acoustic hearing model and the electrical hearing model are extended with an additional model go generate neural spike patterns and then neurograms. Their discharge patterns can be evaluated subjectively by visual inspection, however to statistically determine the differences between the neurograms a quantitative objective performance measure is needed. In this chapter three different objective measures will be discussed, two based on image analysis methods and one on vocoder transformation. A common denominator for the three is that they are developed to predict speech intelligibility. The Neurogram Similarity Index Measure (NSIM) has been extensively used within this model setup and will therefore be the focus of testing and validation.

## 4.1 NSIM

The acoustic and electric hearing model respectively respond differently to input stimulus and both generate spike patterns. There are numerous methods that can be used to compare the spike trains including simple quality metrics based on error sensitivity like the mean squared error (MSE)[20].

Some of the more complex methods that originate from image processing, NSIM being one of them, evolve around that the neurograms from the model output can be treated as images. NSIM was originally developed as a modified version of the Structural Similarity Index Measure (SSIM) used to evaluate JPEG compression quality between compressed and uncompressed images. The modified version is described as a technique to access speech intelligibility by effectively ranking the information of degradation from different amount of simulated hearing loss in the acoustic hearing model [7]. Throughout this project we will make the assumption that NSIM can be used as a comparison measure between the acoustic output considered the reference and the CI model.

The underlying theory of SSIM is to extract structural information from the image used to characterise its composition. It follows that a measure of feature extraction change quantitative can capture the perceived image distortion. The SSIM measure is a straightforward comparison of two signals directly. For our application it will not serve as a measure of image degradation but instead measure the similarity between the reference neurogram and CI output.

SSIM is computed as a weighted sum of three structural parameters; luminance (l), contrast (c) and structure (s) between two neurograms. Luminance compares the mean intensity values of the neurograms. The contrast is a variance measure and the structure parameter equivalent to the cross correlation coefficient. The SSIM between two neurograms $r$ and $d$ is defined as

$$SSIM(r,d) = \left(\frac{2\mu_r\mu_d + C_1}{\mu_r^2 + \mu_d^2 + C_1}\right)^{\alpha} \cdot \left(\frac{2\sigma_r\sigma_d + C_2}{\sigma_r^2 + \sigma_d^2 + C_2}\right)^{\beta} \cdot \left(\frac{\sigma_{rd} + C_3}{\sigma_r\sigma_d + C_3}\right)^{\gamma} \tag{4.1}$$

### 4.1.1 NSIM parameters

At each point of the neurogram the local statistics $(\mu_r, \sigma_d, \sigma_{rd})$ are computed within a $3 \times 3$ square window. The local statistics are used to compute an SSIM map over all regions. Finally, the mean of the SSIM map represent the overall similarity value [7].

Each component in (4.1) also contains constants, $C_1, C_2, C_3$, which are chosen somewhat arbitrary and have negligible influence on the results but are included to ensure stable boundary conditions. The weighted coefficients $\alpha$, $\beta$ and $\gamma$ are used to adjust the relative importance of each of the three SSIM components.

In [7] the relative contribution from each component is investigated through human listener test scores. The scores which resulted in the best phoneme discrimination had $\alpha$ and $\gamma$ close to full weighting whereas $\beta$ had almost no contribution. These results serve as a strong argument to excluding the contrast component in the optimally weighted SSIM and thus $[\alpha, \beta, \gamma] = [1, 0, 1]$. One could argue that you should find new weights when having a CI application, but as listener test are not feasible within the given time, the same configurations as above mentioned will be used throughout the project. The simplified version of SSIM is referred to as NSIM:

$$NSIM(r,d) = \frac{2\mu_r\mu_d + C_1}{\mu_r^2 + \mu_d^2 + C_1} \cdot \frac{\sigma_{rd} + C_2}{\sigma_r\sigma_d + C_2} \tag{4.2}$$

## 4.2 Alternative measures

As it will show in section 6, the performance measure NSIM will give inconclusive results for certain types of controlled model configurations. It is therefore desirable to investigate alternative objective measures.

### 4.2.1 NOPM

A recently proposed alternative speech intelligibility measure is the Neurogram Orthogonal Polynomial Measure (NOPM) [11]. This metric suggests the use of orthogonal moments as a feature extractor to predict speech intelligibility for listeners with hearing loss and has a wider dynamic range than the previous discussed NSIM.

Similarly as for NSIM, the NOPM is developed to predict speech intelligibility for a range of hearing loss under quiet and noisy conditions. The metric makes use of orthogonal moments which has been used in various image processing areas for pattern recognition, image segmentation and multi-resolution analysis to mention a few. One of the most important properties of orthogonal moments is the ability to localize in space, meaning that they allow analysing and reconstructing a certain part of the original image, which in turn could contain relevant perceptional feature information. Moreover orthogonal moments are good signal descriptors as they can capture small intensity pixel changes efficiently.

Signals can be transformed from time to moment domain, where they have more compact representation, using a set of basis functions. A commonly used set of orthogonal moments is the Discrete Krawtchouck Transform based on the orthogonal Krawtchouck polynomials. The n$th$-order normalized Krawtchouk polynomial is given by

$$K_n(x) = \sqrt{\frac{\binom{N-1}{x}\, p^x\, (1-x)^{N-1-x}}{(-1)^n \left(\frac{1-p}{p}\right)^n \frac{n!}{(-N+1)_n}}}\,{}_2F_1\left(-n,-x;-N+1;\frac{1}{p}\right)$$

where $p$ controls the moments' localization in the moment of interest and can extract different frequency components. The function ${}_rF_s$ is a hyper-geometric series. The Krawtchouck transformation for a neurogram block $f(x,y)$ of size $N \times N$ is given by

$$\psi_{nm} = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} K_n(x)K_m(y)f(x,y)$$

To compare similarity between two neurogram transformed to moment domain cross-correlation is applied to provide quantitative results.

### 4.2.2 Vocoder + STOI

A different approach than the above mentioned image analysis method is to synthesize the sound from a given spike pattern using a spike-based vocoder. This method has no reference from the acoustic model, instead the processed signal from the CI model is compared with the original unprocessed sound.

The implemented spike-vocoder is described very briefly as introducing a wavelet at each spike with center frequency at the auditory nerve fiber that elicited the spike. A Gaussion function is modulating the wavelet amplitude. To create the synthesized sound, both wavelets across the whole frequency range and for each simulated auditory nerve fiber are superimposed, see Figure 4.1.

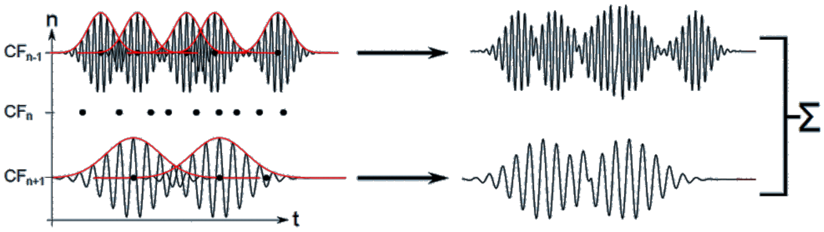


Figure 4.1: Spike-based vocoder sketch.

The output from the vocoder can be subjectively evaluated by simply listening to the synthesized sound. However, to get comparable results with the two previously described measures NSIM and NOPM, an objective measure is needed also in this case.

A proposed objective intelligibility measure which has shown good performance for methods including noisy speech is the Short-Time Objective Intelligibility (STOI) measure, developed partly by support from Oticon A/S. The measure has a relatively simple structure and is based on an intermediate measure analyzing time-frequency regions during short-time segments (40ms) [19].

STOI is a function of clean (x) and processed (y) time-aligned signals. A time-frequency representation is obtained by applying a 50% overlapped Hanning window to both signals where each frame contains 256 samples and is zero-padded with up to 512 samples. Each frame is then Fourier transformed and frequency bins grouped to range over one-third octave bands defined as

$$X_j(m) = \sqrt{\sum_{k=k_1(j)}^{k_2(j)-1} |\hat{x}(k,m)|^2}$$

for the $j$th one-third octave band where $\hat{x}(k,m)$ denote the $k$th DFT-bin of the $m$th frame of the clean signal. The parameters $k_1$ and $k_2$ refer to the one-third octave band rounded end points. The time-frequency representation of the processed speech is defined similarly and denoted $Y_j(m)$. The intermediate intelligibility measure for one time-frequency unit, $d_j(m)$, depends on a region of N consecutive time-frequency units from both $X_j(m)$ and $Y_j(m)$.

CHAPTER 5

# Methods

This chapter presents the methods analyzed to optimize the output from the CI model, initialized with the original configuration for a typical Oticon Medical CI user. It also focuses on the validation and adaptiveness of the objective evaluation measure NSIM.

## 5.1 Original model configuration

In section 3.2.4 it was described how the CI model is configured to match the output from the MAP model so that their respective outputs can be directly compared. The CI model is as mentioned built up by independent platforms each consisting of many variable parameters which can be set to represent the current hearing situation. To limit the set of parameters varied we let the the model be configured for a typical Oticon Medical CI user. This will be considered the original CI reference. Although several common interfaces may be present in the sound processor among the CI users it should be emphasized that each CI patient is individually fitted.

The typical Oticon Medical CI user is fitted with a multichannel 20-channel electrode with a 24 mm insertion depth from the base. If assuming that the cochlea itself has a length of 35 mm when unravelled, an insertion depth of 24 mm means that the electrode array does not reach to the lower frequencies located in the apex [6]. Among the 20 inserted electrodes a subset of 12 are maximally activated in each time frame defining the number of spectral bands of stimulation along the tonotopical map of the cochlear region.

One side-effect of different insertion depth of the implanted electrode array is the degree of frequency mismatch it is causing. Frequency mismatches can occur when the input signal frequencies fail to map to the corresponding characteristic frequency of the neurons at the electrode locations. Clinical practice today is to give most CI users a standard frequency-to-electrode allocation table where basal electrodes are assigned high frequencies and more apical electrodes are assigned lower frequencies. Even if the human brain is highly adaptable and the discomfort of frequency mismatch can be reduced by training, for some patient this hinders their speech-perception [8].

Another restriction among CI users is the suppressed hearing dynamic range. The dynamic range of the auditory system is defined as the difference between the smallest intensity threshold the ear can perceive and the largest intensity the ear can tolerate (pain threshold). In the typical Oticon Medical user the dynamic range is set between 23- 95 dB SPL.

### 5.1.1 Frequency-electrode mapping

From the STFT transform and the envelope detection in the BTE signal processor a number of frequency bands are generated representing the theoretically available frequency range for the CI user. Something that has become more and more important for commercial implants to consider is the mapping of frequency bands to electrodes in the implant using a frequency-to-electrode allocation table. Studying the literature, most mapping strategies have been selected to preserve speech perception optimally not taking pitch or harmonic structure into account as an example [9].

The signal encoding strategy used in Oticon Medical devices encodes signals between 130 Hz and 8333 Hz onto maximally 20 electrodes. The frequency range is typically similar in all today's commercial cochlear implants. The uniformly spaced FFT bins are combined by summing the powers to provide the required number of channels $M$ including the envelopes in each spectral band. The frequency range up to 1 kHz follows a linear mapping whereas for frequencies above 1kHz frequency bins are mapped logarithmically. The number of FFT bins per frequency band with corresponding center frequencies for a 20-channel electrode device can be studied in Table 5.1.

| **Channel No.** | **No. of bins** | **Bandwidth [Hz]** | **Center freq. [Hz]** | **Channel No.** | **No. of bins** | **Bandwidth [Hz]** | **Center freq. [Hz]** |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 130-260 | 195 | 11 | 1 | 1432-1693 | 1562 |
| 2 | 1 | 260-391 | 326 | 12 | 2 | 1693-1953 | 1823 |
| 3 | 1 | 391-521 | 456 | 13 | 2 | 1953-2344 | 2148 |
| 4 | 1 | 521-651 | 586 | 14 | 3 | 2344-2734 | 2539 |
| 5 | 1 | 651-781 | 716 | 15 | 3 | 2734-3385 | 3060 |
| 6 | 1 | 781-911 | 846.3 | 16 | 5 | 3385-4036 | 3711 |
| 7 | 1 | 911-1042 | 977 | 17 | 5 | 4036-4817 | 4427 |
| 8 | 1 | 1042-1172 | 1107 | 18 | 6 | 4817-5729 | 5273 |
| 9 | 1 | 1172-1302 | 1237 | 19 | 7 | 5729-6770 | 6250 |
| 10 | 1 | 1302-1432 | 1367 | 20 | 8 | 6770-8203 | 7487 |

Table 5.1: Number of FFT bins with corresponding center frequencies for M=20. Using a 128-point FFT and an input sampling frequency of 16666 Hz provides a minimum resolution of 130.2 Hz.

### 5.1.2 Max selection strategy

Following the envelope detection is the selection block where a subset $N$ $(N < M)$ channels are selected sequentially for stimulation in each frame of the audio signal. The value of $N$ can be set individually and affects the spectral resolution of the signal. For the general patient the active channel number is set to $N = 12$ and selected as the channels with the largest amplitudes among $M = 20$.

### 5.1.3 XDP compression

In the final stage of the signal processor the $N$ largest amplitudes are mapped to the corresponding electrodes and the acoustic amplitudes are compressed into the CI user's dynamic range defined by the measured threshold and the maximum comfortable loudness level.

The Oticon Medical CI BTE Saphyr processing strategy aims at improving speech understanding in noise, provide the largest possible input dynamic range and ensure comfort for patient in all hearing situations. The compressor takes the signal energy in each frequency band as input and maps it directly to the electrical stimulation levels. The mapping procedure can be seen as a transfer function that maps 95% of the signal input energy (dB SPL) below a knee point (KP) which represents 75% of the electrical stimulation range ($\mu s$), see Figure 5.1. For input levels above the KP the signal is heavily compressed [17].

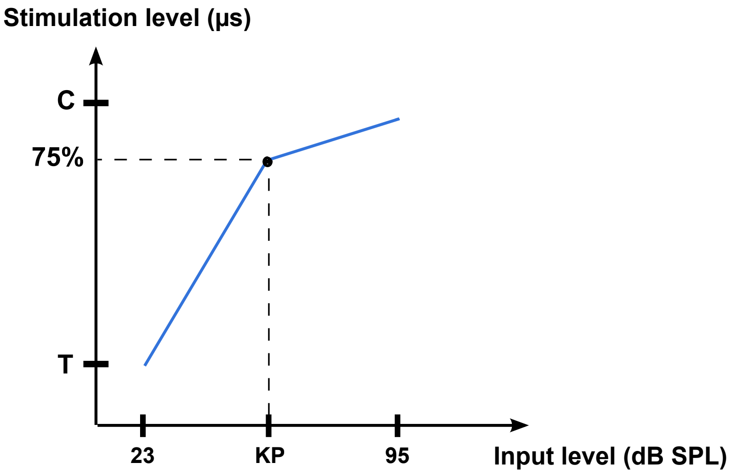


Figure 5.1: XDP compression function.

The XDP strategy allows the compressor to operate on each electrode independently. However, to reduce fitting complexity four frequency bands are grouped together by hierarchical clustering. Within these four ranges the frequency bands have similar power spectral density (PSD) and thus identical compression functions are used.

To optimally preserve speech information and provide comfort for the patient pre-sets have been designed to identify three different hearing environments; quiet 50 dB SPL, medium 60 dB SPL and loud 70 dB SPL. The KPs for each of the pre-sets and frequency bands can be observed in Table 5.2. For further testing, the input signal level will be fixed to 50 dB SPL which have shown to be compatible with the output from the MAP model.

| Frequency range [Hz] | $\mathrm{KP}_{\mathrm{QUIET}}$ [dB SPL] | $\mathrm{KP}_{\mathrm{MED}}$ [dB SPL] | $\mathrm{KP}_{\mathrm{LOUD}}$ [dB SPL] |
|---|---|---|---|
| 130 - 780 | 52 | 60 | 71 |
| 780 - 1430 | 52 | 60 | 71 |
| 1430 - 3380 | 47 | 57 | 65 |
| 3380 - 8333 | 41 | 50 | 48 |

Table 5.2: Frequency ranges with corresponding knee points for three different environments; quiet (50 dB SPL), medium (60 dB SPL) and loud (70 dB SPL).

## 5.2 Music configuration

As previously mentioned the current CI encoding strategies are optimized for speech characteristics whereas music enjoyment is significantly below normal hearing performance. A number of changes will be implemented and evaluated within the blocks of the signal processor with the aim to better capture music characteristics and improve frequency resolution for different music test signals.

### 5.2.1 Octave band distribution

To better preserve the harmonic structure of music signals a frequency-to-electrode mapping that involves tone mapping will be used. With the idea first proposed in [9] this approach includes frequency band mapping based on octave bands or parts of an octave. The octave mapping is as previously restricted by the frequency resolution for each channel.

### 5.2.2 Alternative N-of-M channel selection

Signal encoding strategies are essential for the user's ability to perceive and understand sound. It can be questioned whether the simple max strategy approach mentioned in section 5.1.2 performs optimally. The aim of an alternative method is to increase temporal resolution by neglecting less important spectral components while keeping the ones with more important features.

#### No neighbour selection

A first alternative approach is motivated by the fact that the stimulated $N$ frequency bands are relatively wide to accurately encode tonal components of audio signals. Even though they are non-overlapping it will show in the results that the impact of current spread from each electrode is causing signal interaction. By adding the condition that no neighbouring electrodes are allowed stimulating within each frame such a distortion could be limited. In practice it means that maximally $N = 10$ electrodes can be stimulated each per cycle, hence less envelope information is used to generate the electrical pulses.

#### Psychoacoustic channel selection

Another approach, anchored in simulating the behaviour of a healthy auditory system is referred to as the psychoacoustic masking model and based on psychoacoustic measurements of the masking threshold. The aim of this method is to describe auditory masking effects that occur when the perception of one signal is affected by the presence of another sound. The model will here be used to select the $N$ most significant bands in terms of perception by first calculating the individual masking components for each frequency band and then using non-linear superposition to create the overall masking threshold [1].

For each band, the individual masking threshold modelling the masking effect of the respective band upon the others, $L_i$, is determined using a triangular spreading function. The representation of the spreading function belonging to a frequency band $z_i$ with amplitude $A(z_i)$ is given by

$$L_i(z) = \begin{cases} A(z_i) - a_v - s_l \cdot (z_i - z), & z < z_i \\ A(z_i) - a_v - s_r \cdot (z - z_i), & z \geq z_i \end{cases}$$

where $z$ denotes the frequency band number at the critical band interval, $1 \leq z \leq M$, $a_v$ represents the attenuation level defined as the difference between the amplitude $A(z_i)$ (dB SPL) and the maximum of the spreading function, $s_l$ and $s_r$ correspond to the left and right slopes of the spreading function in dB/band unit, see Figure 5.2b.

A "power-law" model as described in [1] is used for non-linear superposition of different masking thresholds to calculate the overall masking threshold. The sum is formed as

$$I_T(z) = \left[ \sum_i [I_i(z)]^\alpha \right]^\alpha$$

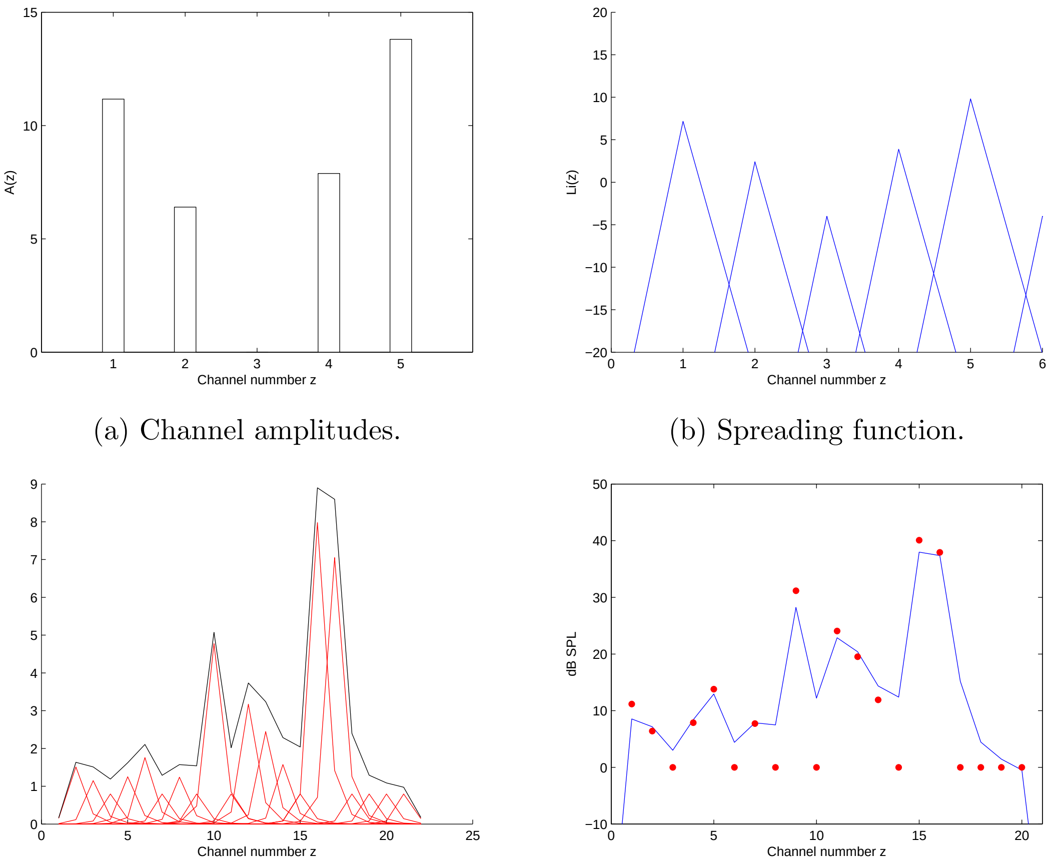
where $I_i$ denotes the sound intensities calculated from the sound levels as

$$I_i(z) = 10^{L_i(z)/10}$$

The parameter $\alpha$, $0 < \alpha \leq 1$, allows the superposition to be carried on in a non-linear mode for $\alpha < 1$. The overall masking threshold $I_T(z)$ can be seen in Figure

5.2c plotted together with $I_i(z)$ for $1 < i \leq M$.

The selection of $N$ bands is performed in a straightforward way as the number of amplitudes $A(z_i)$ reaching above the overall masking threshold, see Figure 5.2d. If the number of peaks over the overall masking threshold is larger than a predefined limit on $N$, the $N$ largest amplitudes are selected, hence the max strategy is utilized.



(a) Channel amplitudes.

(b) Spreading function.

(c) Nonlinear superposition.

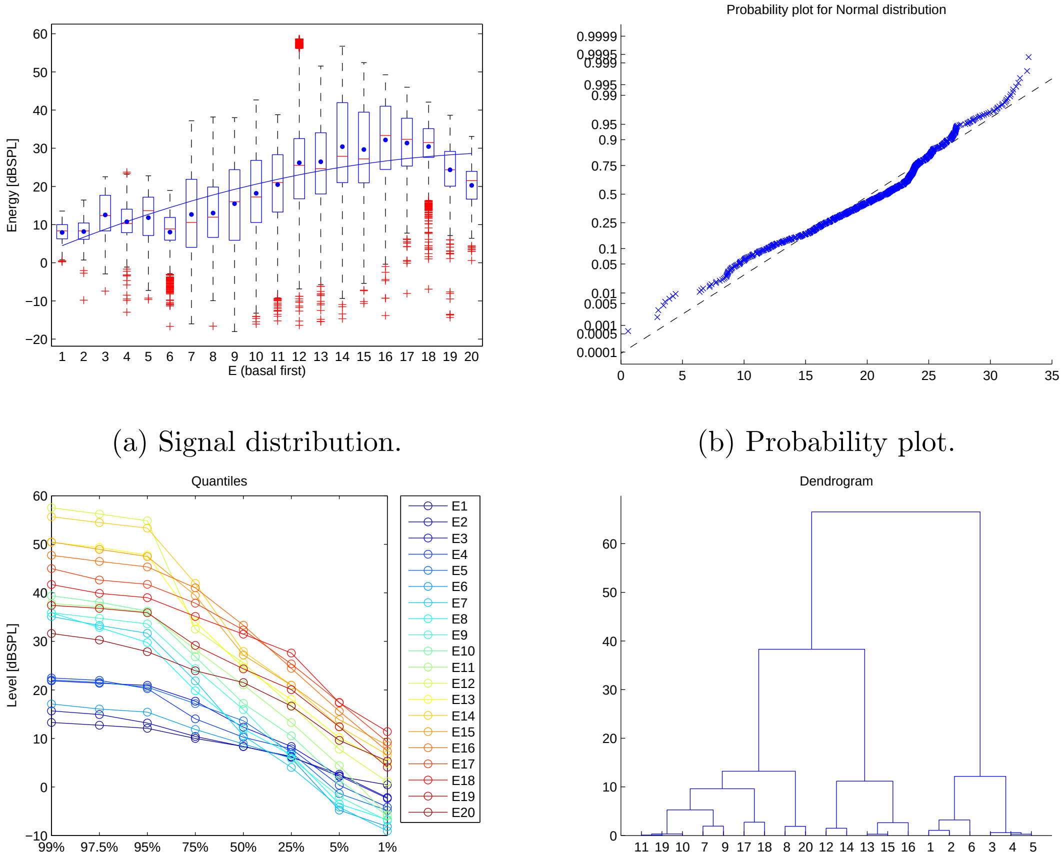
(d) Selection algorithm.

Figure 5.2: Associated levels over frequency band number z. The spreading functions $L_i(z)$ in Figure 5.2b is calculated for every masker component $A(z_i)$ at the band $z_i$, see Figure 5.2a. The left and right masking slopes are chosen to 40 dB/band and 30 dB/band respectively as proposed in [15]. The attenuation level $a_v$ is highly varible and here set to 4 dB to fit the input signals. The parameter $\alpha$ controlling the non-linear superposition in Figure 5.2c is set to 0.25 in accordance with [15].

### 5.2.3 Music compression

The KPs determined in the XDP compression strategy are based on a database of speech signals and then clustered to avoid fitting complexity. To investigate how many frequency bands are optimal for music input and the dynamic spread for different types of music, a new set of KPs will be determined using a statistical approach. The analysis is based on determining the 95-percentile of the signal acoustic energy for each frequency band, see Figure 5.3c, and then applying hierarchical clustering to group several frequency ranges together, see Figure 5.3d.

Depending on the type of music one could expect a significant change of stimulation output when applying new KPs, e.g. pop music which if often compressed and uses a small dynamic range could experience an improved temporal resolution.



(a) Signal distribution.

(b) Probability plot.

(c) Quantile plot.

(d) Dendrogram.

Figure 5.3: Steps of the statistical analysis performed to find new compression parameters. Figure 5.3a shows an example of a boxblot of how the signal energy is spread for each channel. In Figure 5.3b the probability of a selected channel following a Gaussian distribution is shown. Figure 5.3c illustrates how well the channels are separated for different quantiles and a visualisation of the hierarchical clustering algroithm can be seen in the dendrogram in Figure 5.3d.

## 5.3 Resolution of NSIM

NSIM will be used as the primary measure of similarity between the normal hearing and electrical hearing neurogram. It is therefore important that it can track changes made in the CI model in order to be able to indicate improvements made in the signal processor. To quantify the performance of NSIM the measure will be evaluated using a number of test signals and controlled changes implemented in the simulation model. The aim is to see how predictable the value of NSIM is and thus if it is a good similarity measure in this application. The implemented model configurations are listed in Table 5.3.

| **Model config.** | **Org. setting** | **Description** |
|---|---|---|
| Insertion depth | 24 mm | Increase the insertion depth of the electrode array to 29 mm to match the Greenwood function used in the MAP model to correlate the position of IHCs to the frequencies that stimulate the corresponding auditory neurons. |
| Current spread | Exp. function | The intracochlear electric spread is modelled as an exponential function with decreasing slopes and peak value at each electrode position. Vary the amount of current spread by changing the exp. decay parameters. |
| Elec-to-Freq map | linear $\leq$ 1 kHz<br>log $>$ 1 kHz | Octave band distribution. |
| Filtering & SR cleaning | - | Apply a frequency smoothing filter and threshold on the normal hearing neurogram to increase resemblance with the time frames of the electrical hearing neurogram and remove the spontaneous activity. |
| FFT resolution | 128-point | Increase the frequency resolution by applying a 512-point Hamming window. |
| N-of-M strategy | N max amp. | Alternative channel selection strategies. |
| Compression | XDP compression<br>incl. 4 freq. clusters | Vary the clustering and KPs of the XDP compression to find optimal settings for music signals. |

Table 5.3: Table describing the controlled changes made in the signal processing part of the CI block to predict the performance of NSIM.

## 5.4 Test signals

To evaluate the performance of the model chains, the same audio stimulus is fed to the MAP and CI model respectively. Instead of forming the conclusions around a quantitative analysis only a few test signals with required characteristics are selected and used throughout the validation.

The first test signal chosen is a simple chirp. The logarithmic chirp is a sweeping sinus with increasing frequency over time can be loosely referred to as "*gliding tones*". It is typically used as a measurement signal in audio applications.

The other measurements signals are short pieces of music input provided EBU to serve as sound quality assessment material [3]. The music input comprises two signals; a single instrument flute playing a 10 s rising scale and a 3 s pop music selection with ABBA - The Visitors. The flute input can be seen as representative for harmonic music signals with a wide bandwidth (wider than speech) whereas pop music is usually very compressed and uses a smaller dynamic range. The pop music is chosen to include an increase in amplitude after approximately 1.5 seconds.

CHAPTER 6

# Analysis and results

This chapter presents the main results of the NSIM evaluation for the chosen test signals, for different configurations of the CI model, and combination of those. A comparison with the alternative proposed objective measures, NSIM and STOI, is also presented.

## 6.1 Test signal: chirp

The spike pattern response from the acoustic model chain to a chirp input can be viewed in Figure 6.2. Figures 6.2a and 6.2c show spike patterns generated only by the original MAP model for two different input levels. As discussed in section 3.1 the MAP model implements all the stages of the auditory periphery and can thus generate spike patterns without the extension of the Goldwyn model, by including a probabilistic or quantal model simulating neurotransmitter release. By instead cutting the MAP model (before the auditory nerve firings) and extending it with a probabilistic spike generator from the Goldwyn model one get the output presented in Figures 6.2b and 6.2d for the chirp signal.

For input level 50 dB SPL the MAP and Goldwyn responses are similar with the main differences lying in the generated spontaneous activity. Increasing the input level to 70 dB SPL seems to cause an over saturation in the MAP response as the signal is hardly visible. The Goldwyn model on the other hand stays intact showing more spread in the signal response from the spike pattern.

The saturation of the MAP model for 70 dB SPL may seem strange but can in fact be explained by the behaviour of HSR auditory nerve fibers. The auditory nerve firing rates reach saturation at different levels depending on the type of spontaneous rate. Most HSR auditory nerve fibers are saturated by 60 dB SPL and will not increase their firing rate significantly above that level. The minority LSR nerve fibers on the other hand are less sensitive to level differences and saturate at higher intensities. The saturation level of HSR and LSR fibers as modelled by the MAP model can be seen in Figure 6.1.

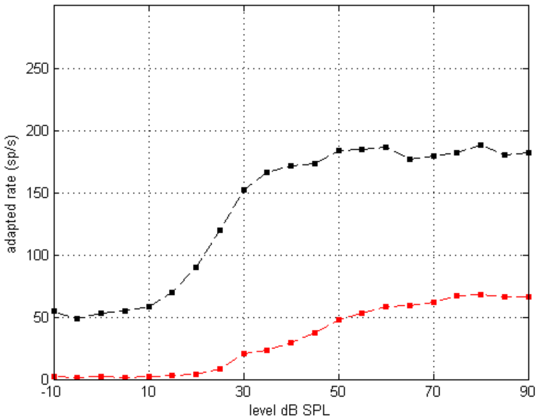
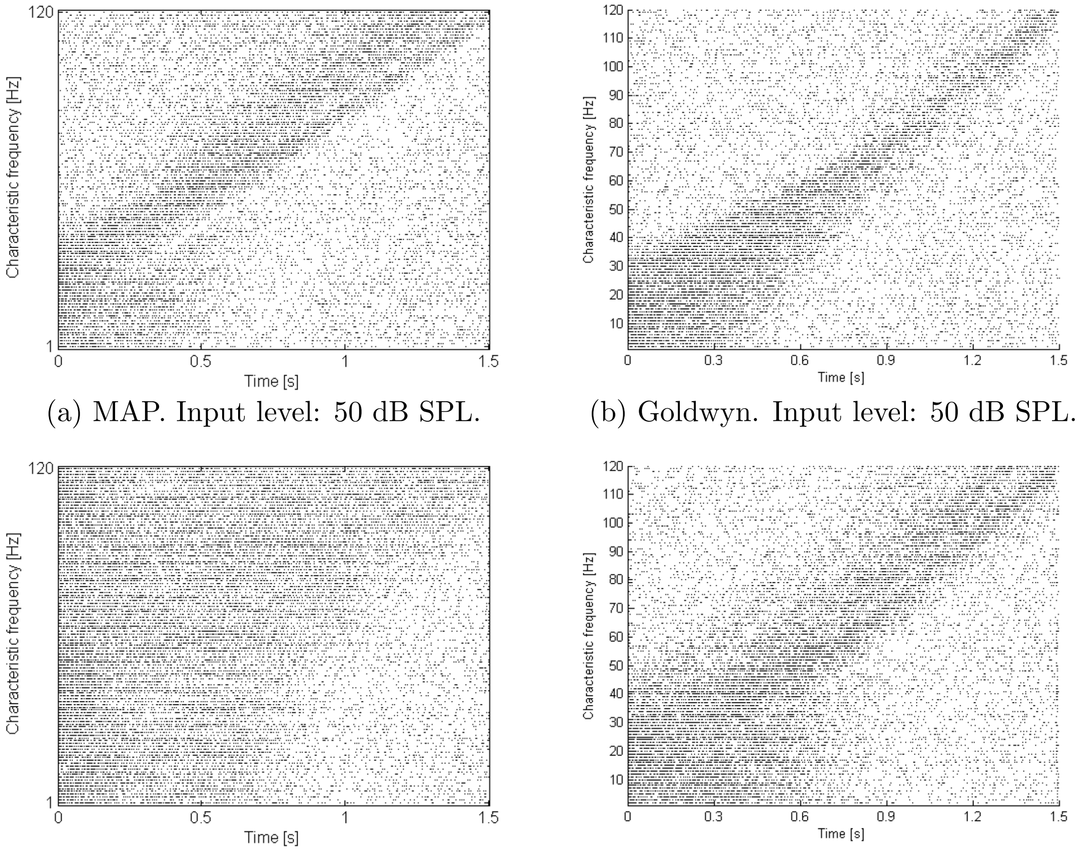


Figure 6.1: Saturation rates for two different types of spontaneous rate fibers as a function of input level. LSR fibers are represented by the red curve and the HSR fibers are shown in the black.

The saturation effect of the spontaneous rate nerve firings are not modelled in the

Goldwyn model and it will therefore give reasonable spike patterns even for higher intensities. To go around this problem and generate comparable results, all further simulation will be done at the same input level at 50 dB SPL. To further ensure that the auditory nerve firings are modelled in the same manner, the Goldwyn model is selected to generate spike patterns both for the MAP and CI model. Recall that one main difference between the two since the MAP model is based on human data whereas Goldwyn is fitted to measurement data from cats.



(a) MAP. Input level: 50 dB SPL.

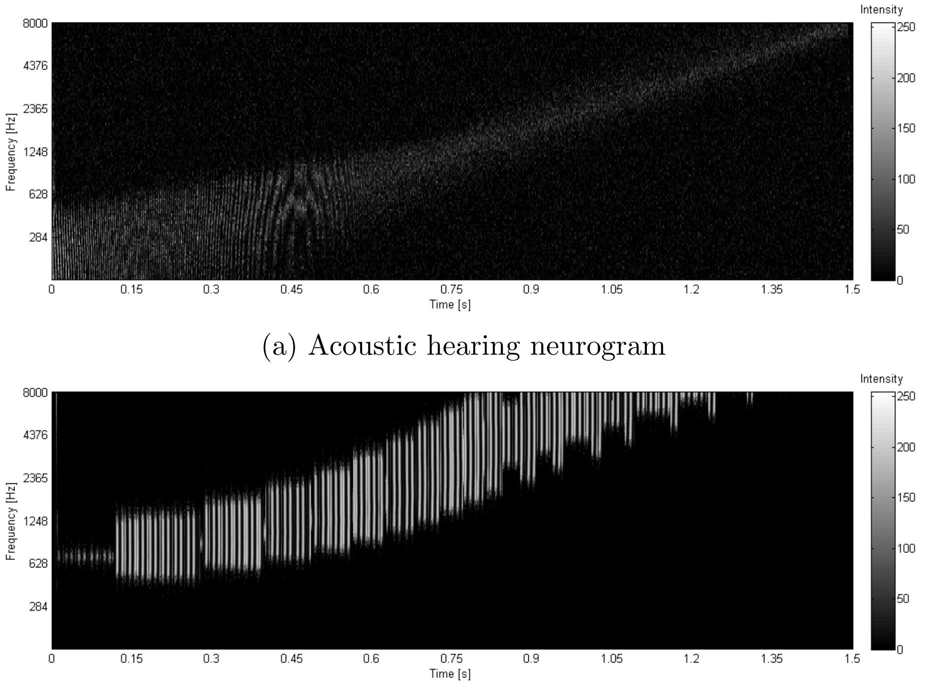
(b) Goldwyn. Input level: 50 dB SPL.

(c) MAP. Input level: 70 dB SPL.

(d) Goldwyn. Input level: 70 dB SPL.

Figure 6.2: Spike patterns generated from the acoustic hearing MAP model and Goldwyn model respectively, for two different input levels; 50 dB SPL and 70 dB SPL.

Figure 6.3 shows the output neurograms from the normal hearing model and electric hearing model respectively, for the same input chirp signal at stimulation rate 520 Hz. On the y-axis we have 120 Greenwood distributed characteristic frequencies and the x-axis represents a time scale in seconds. In the acoustic neurogram, the 120 characteristic frequencies correspond to an equal number of discrete locations along the basilar membrane. For the electric hearing path we have 20 fixed electrodes represented by 20 characteristic frequencies. Modelling the ICEF, 120 frequencies are computed to match the output from the acoustic path.



(a) Acoustic hearing neurogram

(b) Electric hearing neurogram

Figure 6.3: Neurograms generated from the MAP and CI model chain respectively in response to a chirp input stimulus at 50 dB SPL. $\mathrm{NSIM}_1 = 0.1457$.

Both neurograms are scaled equally in time, frequency and amplitude. There are several remarks that need to be made before moving further:

- **Stochastic effects** For simulating normal hearing conditions the MAP model includes three different types of spontaneous rate auditory nerve fibers

(HSR/MSR/LSR). The spontaneous firing rate is clearly visible in Figure 6.3a where the neurogram background is covered in pixels indicating nerve firings. The stochastic effect is however not modelled in the CI model, see Figure 6.3b, where the only firing activity is due to input stimulus. This is the first important difference to consider between the models.

- **Resolution** In the normal hearing neurogram the dynamics in the signal response can be seen on pixel level. In the CI model on the other hand we see a more on/off behaviour in the intensity which is visualised by the vertical "stripes" running along the whole chirp.

Each time frame has reserved time to at maximum 20 active electrodes. Typically this number is smaller for most user fittings. Depending on the audio input level and the number of active electrodes in each 2 ms time frame, there is a period of time with no pulse activity between the last pulse in a time frame and the first following pulse in the next time frame. Thus, the absence in pulse activity in each time frame causes the black vertical lines in the neurogram. The zoomed electrodogram in Figure 6.4 visualizes this behaviour for a broadband signal.

One restriction of the CI model is to not consider the pulse decay effect going from one time frame to another, hence no historic effects are modelled to reduce complexity of the simulation chain. One could think that in reality it is more likely that the CI user experience a continuous sound intensity between time frames rather than a "cut-off" period of silence between the last pulse in each time frame ends and the first one in the next time frame.

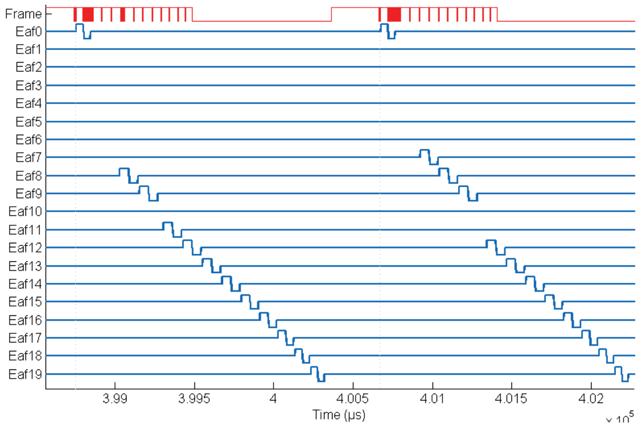


Figure 6.4: Electrodogram showing pulse activity for a white noise input signal. The pulse at electrode Eaf19 in time frame one ends at $4.003 \cdot 10^5$ $\mu s$ and Eaf0 in the next frame starts at $4.007 \cdot 10^5$ $\mu s$, leaving a small gap with no pulse activity measuring 400 $\mu s$.

- **Iterations**

  To produce the neurograms in Figure 6.3 the simulation chain is iterated 10 times. In each trial 120 characteristic frequencies are simulated, related to the tonotopic map of the cochlea. The reason for iterating the paths several times is due to the large spontaneous activity in the normal hearing neurogram affecting the intensity of the signal response. For only one iteration the signal response is hardly visible with a large part of black (no firings) in the neurogram, but superimposing the neurograms of several iterations results in a reduced spontaneous activity effect which is seen as a greyish background over the whole neurogram, and a more prominent signal response. One could think of this as a type of averaging where we are interested in the signal response only, to make the acoustic and electric neurogram comparable.

  The similarity measure NSIM will be dependent on the number of iterations as it is sensitive to the spontaneous firings in the normal hearing (i.e the black region of no pulse activity). Figure 6.5 shows NSIM similarity as a function of number of iterations for a chirp signal. NSIM convergence will depend on the type of input signal and input level. $N = 10$ is chosen to limit the run-time of simulation and shows to produce good enough results.

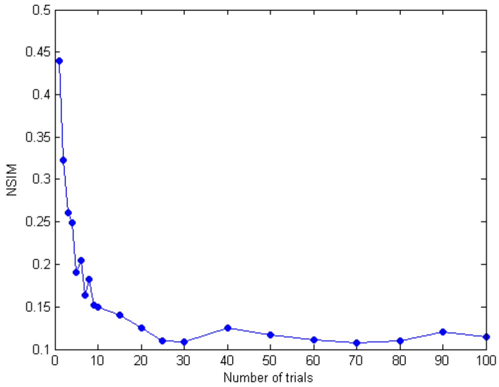


Figure 6.5: Calculated NSIM value for a 50 ms chirp signal at 50 dB SPL input as a function of the number of iterations.

- **Frequency up-shifting** It is clear from Figure 6.3 that the electric hearing neurogram is shifted up in frequency compared to the normal hearing neurogram. This is explained by looking at the original positioning of the electrode array (blue curve) relative the Greenwood function (green curve) in Figure 6.6. The less insertion depth we have of the electrode array (below the Greenwood function) the larger frequency shift we get, since all the characteristic frequencies are mapped to the Greenwood function. Thus, a position along the cochlea corresponding to a small frequency according to the electrode array will in fact correspond to a larger frequency in the Greenwood function which is exactly what we see in the neurogram plots.

As NSIM performs its similarity analysis it assumes that the two compared neurograms are aligned in all dimensions. Since we want NSIM to give an indication of the relative similarity improvement of the signal responses it makes sense to align those to get relevant results. For this reason, in all further testing, the electrode array will be inserted 29 mm and frequency-to-electrode mapped according to octave band distribution, to match the Greenwood function optimally.

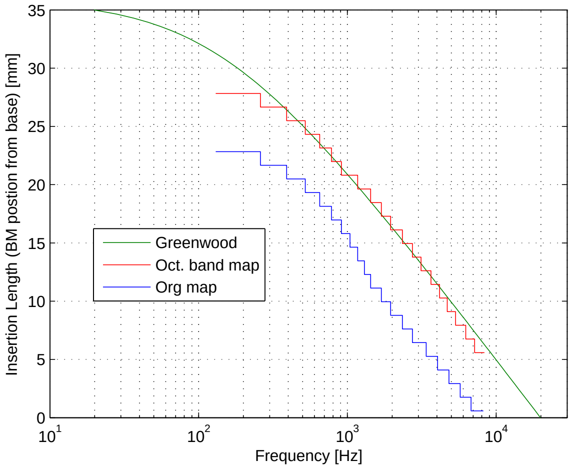


Figure 6.6: Insertion depth [mm] from cochlear base plotted as a function of frequency. The step functions illustrates each channel's allocated frequency range. The blue curve represents the typical patient configuration with 24 mm insertion depth of the electrode array. The red curve shows the electrode array with deeper insertion and an octave band frequency-electrode distribution which matches the Greenwood function (green curve).

- **Current spread** Observing the neurogram obtained from the CI model in Figure 6.3b we see that the even though the electrodes are positioned on discrete locations along the cochlea, the modelling of the intra-cochlear electric spread results in a wide stimulation of frequencies along the whole signal duration. In contrast to the acoustic neurogram the intensity of the auditory nerve firings are mostly either black or white, hardly any gray scale pixels are visible. This means that the firing intensity is almost equally large within the whole signal, even though the electric spread is modelled as an exponentially decaying curve.

Modelling the current spread, the exponential function will in fact never reach zero, meaning that we always have a positive probability of firing a spike along the 120 characteristic frequencies. The Goldwyn model is as familiar used to generate firing spike patterns in response to electrical stimulation. The firing probability is dependent on the length and amplitude of the electric pulse. Here we assume that the amplitude is fixed to 1 mA at the electrode position with a varying pulse length. Figure 6.7 shows the simulated spiking probabilities for a pair of pulses from the Goldwyn model for three different interpulse intervals (time since the last spike). In our case we have a interpulse interval of 1900 $\mu s$ (observing each channel independently) which matches best with the blue curve. On the x-axis current is scaled to dB with the reference of a neural response threshold set to 0.852 mA [5].

For 1 mA current, transformed to $20 \cdot \log \left( \frac{1}{0.852} \right) = 1.371$ dB, we see that we reach the maximum of the blue curve, i.e. probability one for generating a spike. When the current to the auditory nerve decrease we move along the slope of the blue curve to the left where the range for firing a neuron (prob 1) to case fire (prob 0) is relatively small. In Figure 6.3 this effect is visualised through the vertical lines which are dashed at the ends.

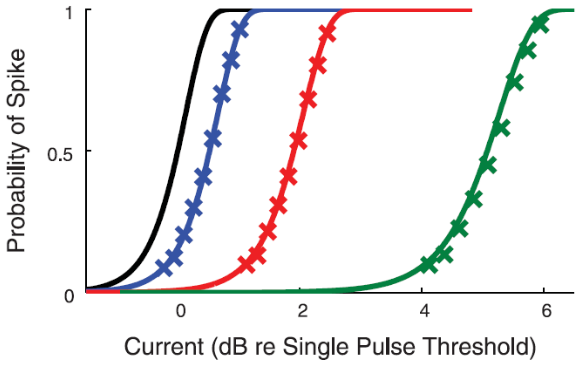
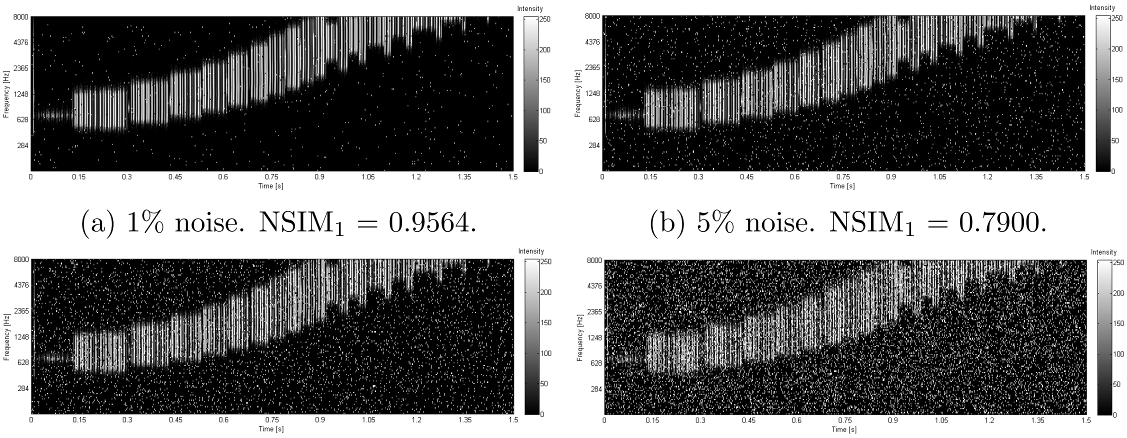


Figure 6.7: Simulated spiking probabilities for three different interpulse intervals of a paired pulse stimuli: 667 $\mu s$ (green), 1000 $\mu s$ (red), 1500 $\mu s$ (blue) [5], with current in dB unit using a reference of 0.852 mA. The simulated ICEF of the CI model with interpulse lengths of 1900 $\mu s$ is best matched to the blue curve. The shifting of firing efficiency curves to the right as the interpulse intervals decrease is related to the refractory period of the excitable neurons. As neurons are less excitable immediately after a spike, shorter interpulse lengths result an in increased current to obtain the same spiking probability. Also, the amount of relative spread increases when more spikes are generated resulting in less flat slopes.

- **NSIM** The similarity between the neurograms in Figure 6.3 is calculated to $\text{NSIM}_1 = 0.1457$. It has already been shown that the NSIM value is dependent on the number of simulated trials. The reduction of NSIM when increasing the number of iterations, as shown in Figure 6.5, is due to the superimposed spontaneous activity in the acoustic neurogram. The sensitivity of NSIM can be observed in Figure 6.8 where random noise is introduced systematically. It is clear that NSIM decrease rapidly with increased amount of in noise. One can also look at it as NSIM decrease as the amount of black regions in the neugrams decrease. This is, as it will also show later, an important restriction of NSIM. Two neurograms where the signal responses are significantly different but both have an equally large black background will in fact be given a high similarity score by NSIM. It turns out that NSIM is more sensitive to changes in the background activity that in the signal response making it difficult to draw conclusions on what type of similarity among the neurograms is actually calculated.



(a) 1% noise. NSIM$_1$ = 0.9564.

(b) 5% noise. NSIM$_1$ = 0.7900.

(c) 10% noise. NSIM$_1$ = 0.6246.

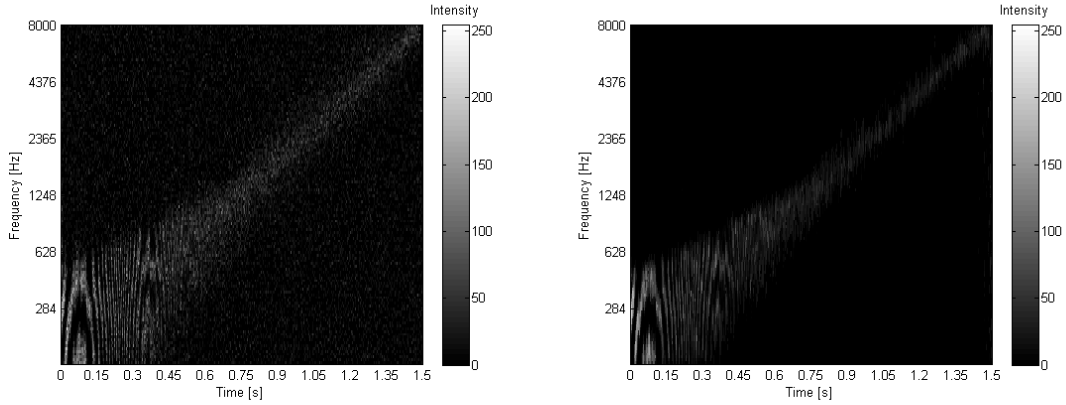
(d) 25% noise. NSIM$_1$ = 0.2881 .

Figure 6.8: Cacluated NSIM for four different percentage levels of introduced noise for a chirp signal; 1%, 5%, 10%, 25%. All noise induced neurograms are compared with a reference with 0% noise.

## 6.1.1 Model validation

### Reference neurogram

With the aim to increase NSIM resolution a median filter is applied on the normal hearing reference including removal of spontaneous neural activity (i.e. removing noise) by applying a density threshold on the spike pattern, see Figure 6.9. However, with the discussion on NSIM sensitiveness from the previous section, a filtering of the acoustic neurogram in practice means that more black regions are exposed. This will most likely result in higher similarity scores as the CI model does not model for any stochastic effects. In reality it is thus not likely that the relative similarity changes, i.e. the resolution, of NSIM increases. Both acoustic neurogram references will however be used when presenting the neurogram results.

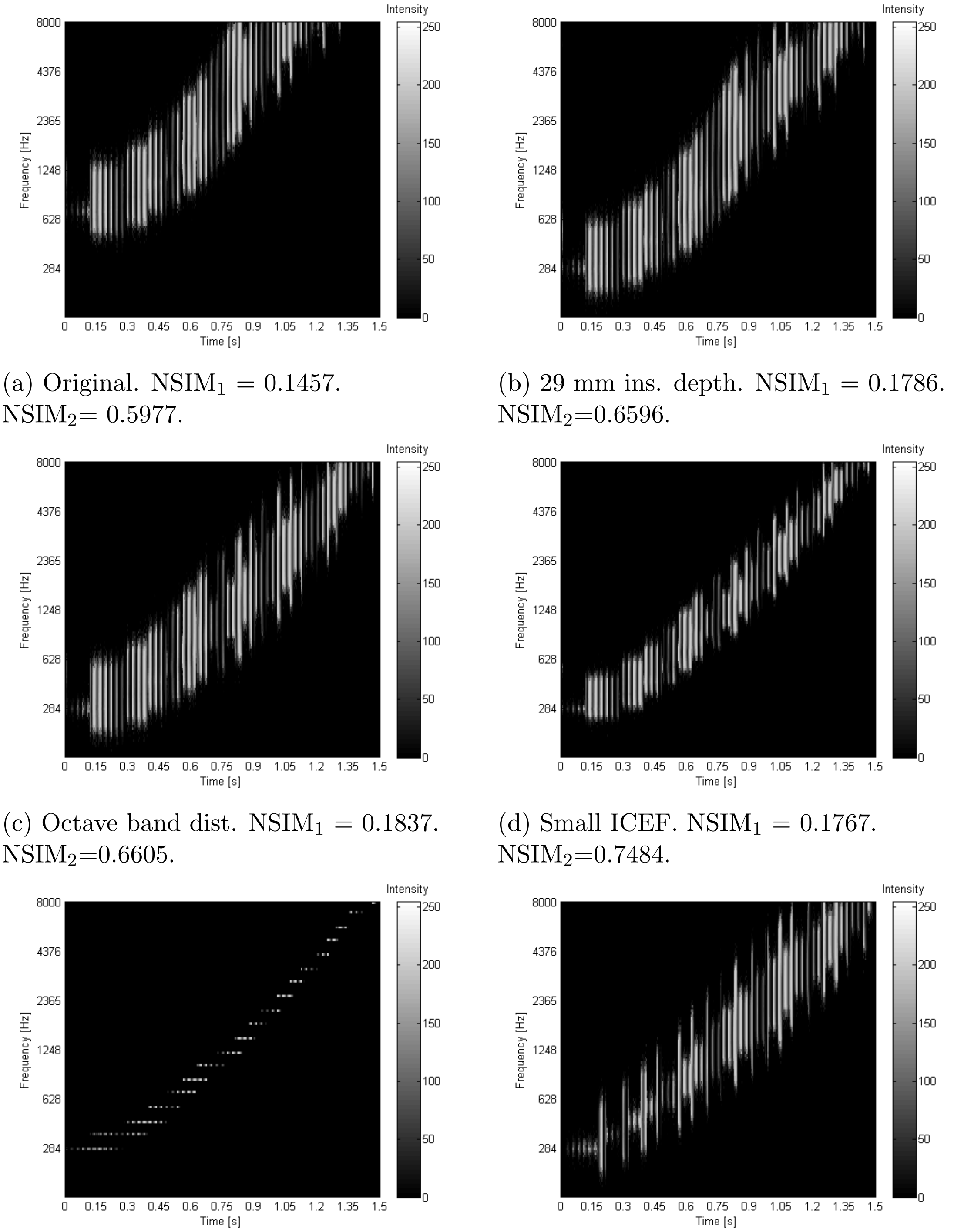


(a) Original reference.

(b) Filtered reference.

Figure 6.9: Reference neurograms with and without included spontaneous activity and smoothing in frequency. In the filitered reference almost all spontaneous neural activity is removed resulting in a narrower signal response.

## CI model configurations



(a) Original. $\mathrm{NSIM}_1 = 0.1457$. $\mathrm{NSIM}_2 = 0.5977$.

(b) 29 mm ins. depth. $\mathrm{NSIM}_1 = 0.1786$. $\mathrm{NSIM}_2 = 0.6596$.

(c) Octave band dist. $\mathrm{NSIM}_1 = 0.1837$. $\mathrm{NSIM}_2 = 0.6605$.

(d) Small ICEF. $\mathrm{NSIM}_1 = 0.1767$. $\mathrm{NSIM}_2 = 0.7484$.

(e) No ICEF. $\mathrm{NSIM}_1 = 0.1609$. $\mathrm{NSIM}_2 = 0.7145$

(f) 512-point FFT. $\mathrm{NSIM}_1 = 0.1809$. $\mathrm{NSIM}_2 = 0.6410$.

Figure 6.10: Superimposed model configurations: insertion depth 29 mm and octave band frequency-to-electrode mapping from Figure 6.10d-6.10f. $\mathrm{NSIM}_1$ refers to the original reference and $\mathrm{NSIM}_2$ to the filtered reference.

### Similarity analysis

An overview of the similarity scores of the CI model configurations for the three different objective similarity measures are found in Figure 6.11, using the unfiltered acoustic reference. For the original configuration of the typical patient NSIM scores 14.57 % while the other image analysis method NOPM results in i similarity score of 0.106 %. Changing the insertion depth to 29 mm and aligning the frequency-electrode distribution with Greenwood is expected to increase the similarity with the acoustic neurogram significantly. Observing the bar diagram NSIM scores 18.37 % and NOPM 22.05 %. To be able to get a sense of the magnitude of change in NSIM one could calculate the standard deviation of the acoustic neurogram. Measurements show that NSIM differ in order of $10^{-3}$ from different simulation of spontaneous activity in the normal hearing for the chirp signal. For the NSIM changes of the CI configuration to be seen as significant it is reasonable to think that the magnitude of change at least have to exceed the standard deviation.

A smaller amount of current spread is expected to give increased similarity due to a more accurate representation of best frequencies at the electrodes. However, a decrease of current spread results in a very small decrease in similarity for both NSIM and NOPM compared with the previous configuration. The result of removing the current spread completely can be observed in Figure 6.10e. Even if current spread generally is seen as a negative effect it is not necessary that an optimal electrical output only has neural activity at the discrete locations of the electrodes. Both NSIM and NOPM has a lower similarity for no CS compared to the third configuration.

Increasing the frequency resolution from a 128-point to 512-point FFT Hamming window length is expected to show more signal dynamics and thus a higher likeness. From Figure 6.10f it can be seen that the effect of the FFT side lobes is reduced with a decrease in current spread. This is however not captured in the NSIM or NOPM which scores 18.08 % and 17.68 respectively.

The STOI scores represented by the blue bars are significantly higher for all configurations than the other objective measures but does not follow the bars of NSIM and NOPM. This approach is as familiar based on the vocoding of the spike pattern to generate an audio signal which is compared with the original input signal.

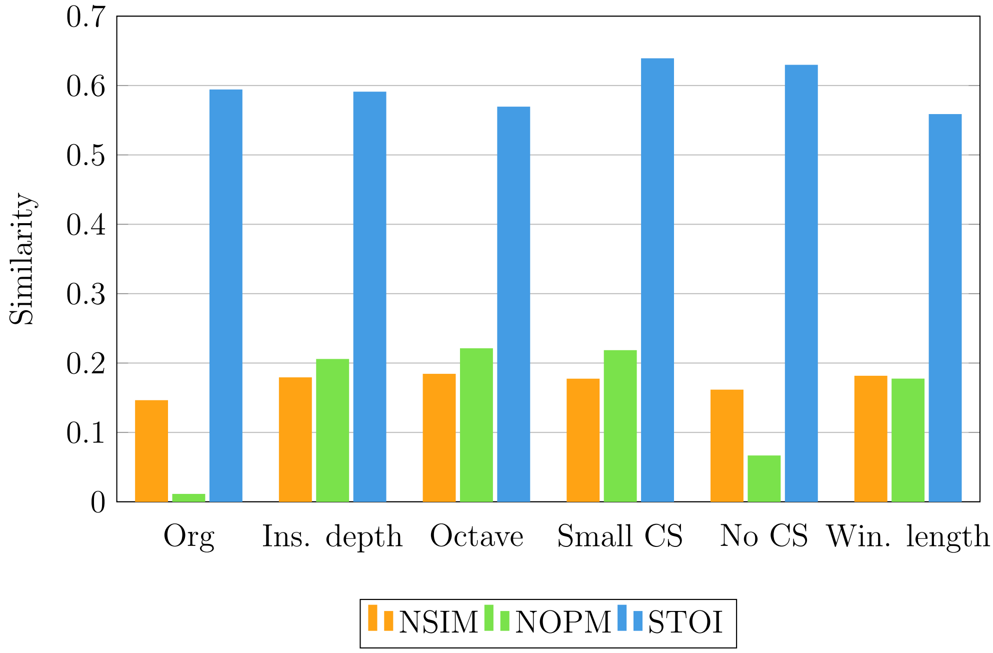
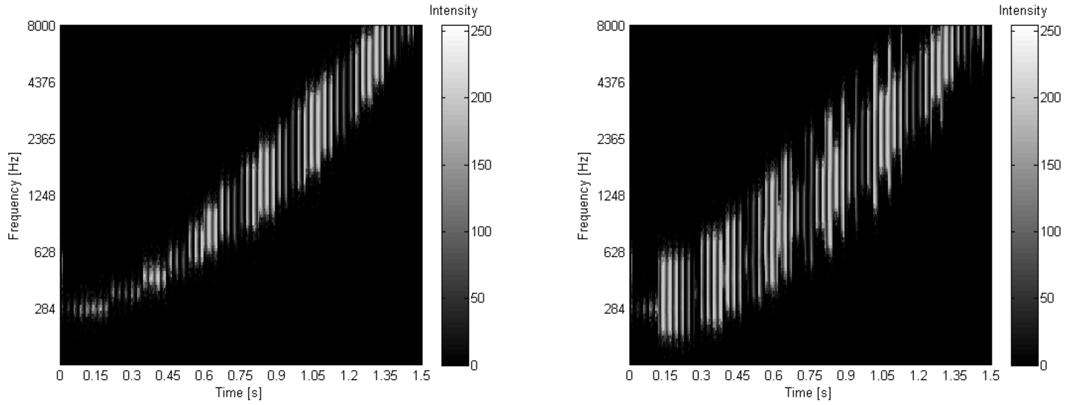


Figure 6.11: Bar diagram showing similarity between CI and normal hearing outputs for different CI model configurations with an input chirp signal. Three different similarity measures are included to observe the relative changes; NSIM(orange), NOPM(green), STOI(blue).

### 6.1.2 N-of-M strategy

A simple alternative to the max channel selection strategy is to allow no adjacent neighbour channels to be stimulated in the same time frame. Another more sophisticated method takes psychoacoustic masking effects into account which are present in an healthy auditory system. The resulting neurograms for each model are found in Figure 6.12.



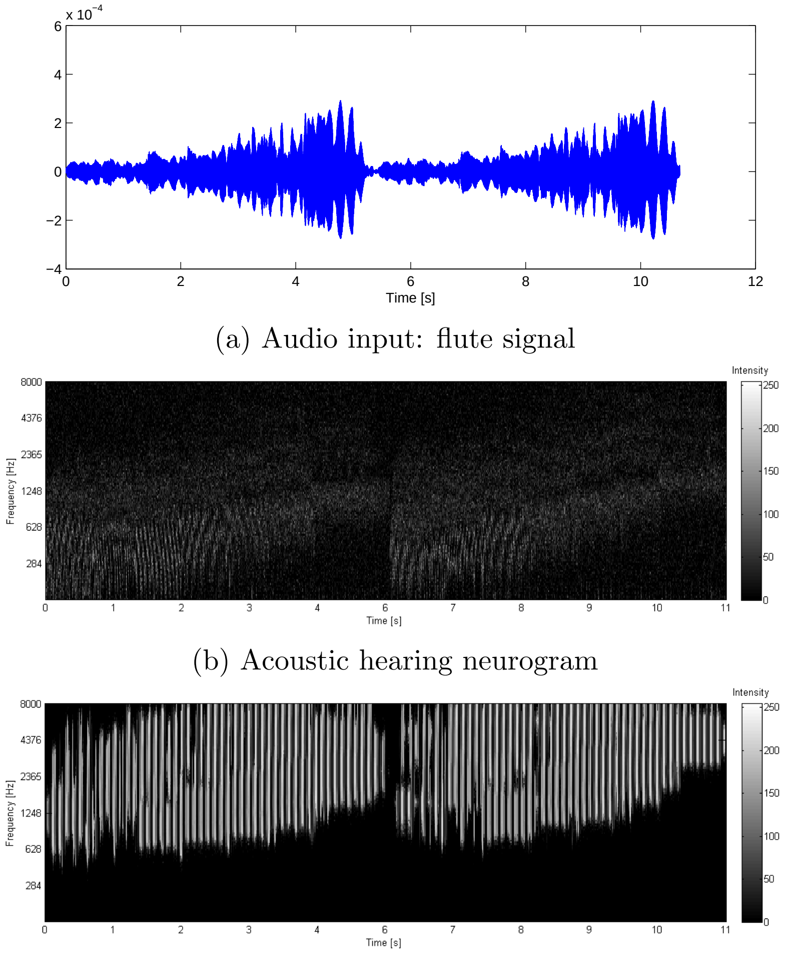
(a) No neighbour. $\text{NSIM}_1$=0.1840. $\text{NSIM}_2$=0.6701.

(b) Psycoacoustic. $\text{NSIM}_1$ = 0.1839. $\text{NSIM}_2$=0.6243.

Figure 6.12: Resulting neurograms for alternative N-of-M channel selection for the chirp signal. In Figure 6.12a no adjacent neighbours are allowed stimulation and 6.12b shows the result of psychoacoustic masking selection Both neurograms are configured with electrode insertion depth of 29 mm and octave band frequency-electrode distribution.

The alternative N-of-M channel selection strategies are expected to generate a more sparse electrodogram and hence a decrease in current spread. For the chirp signal the no neighbour strategy in Figure 6.12a seems to result in an visual improves compared to the reference but this effect fails to be captures in NSIM. Applying a psychoacoustic masking threshold does not give any additional information about the signal in this case.

## 6.2 Test signal: Flute



(a) Audio input: flute signal

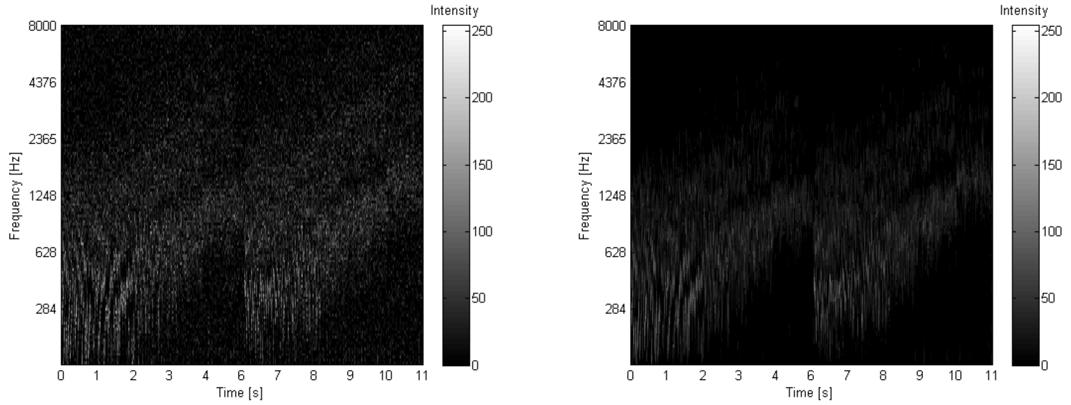
(b) Acoustic hearing neurogram

(c) Electric hearing neurogram

Figure 6.13: $\text{NSIM}_1 = 0.0780$.

### 6.2.1 Model validation

#### Reference neurogram

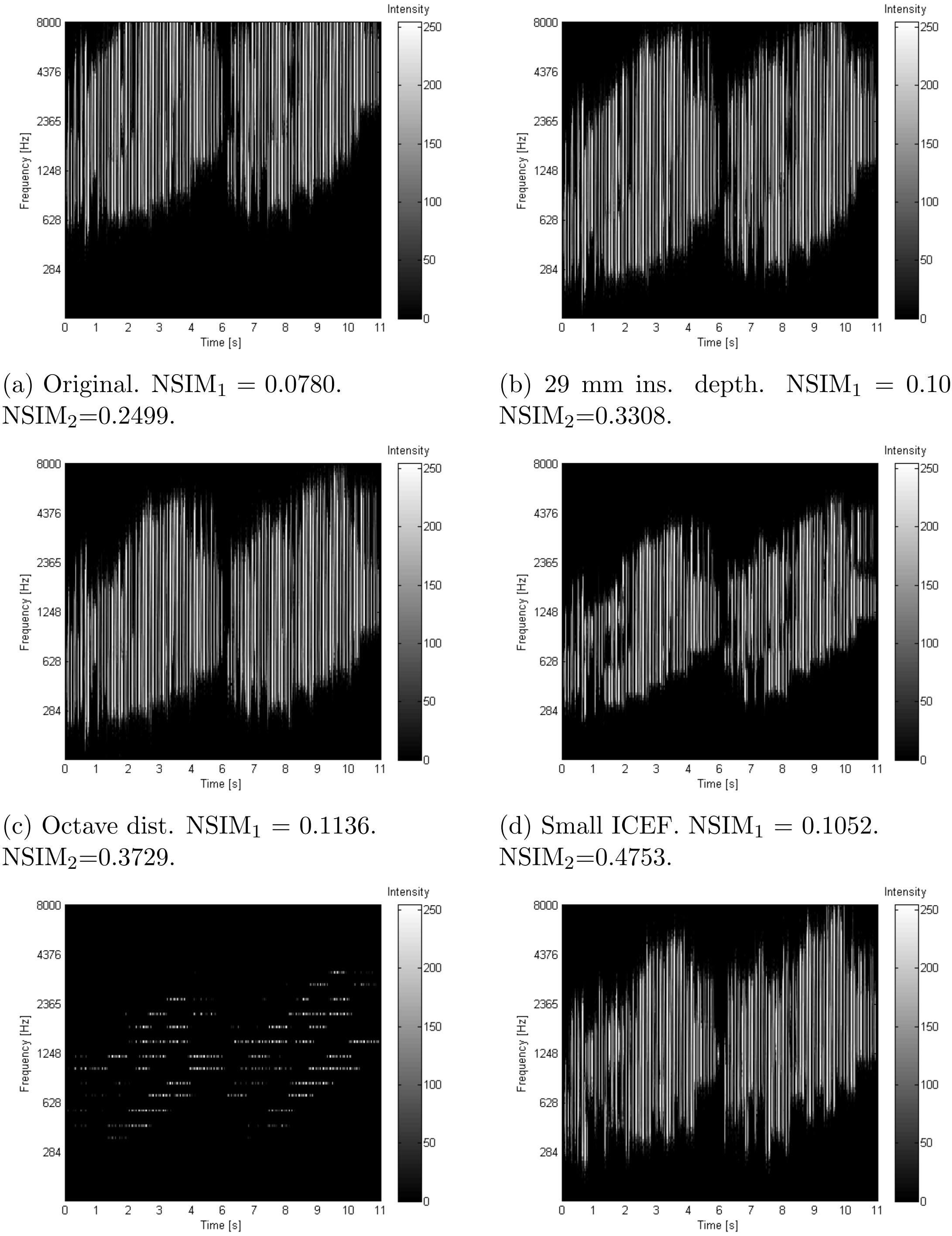


(a) Original reference.

(b) Filtered reference.

Figure 6.14: Reference neurogram with and without included spontaneous activity and smoothed frequency.

### CI model configurations



(a) Original. $\mathrm{NSIM}_1 = 0.0780$. $\mathrm{NSIM}_2$=0.2499.

(b) 29 mm ins. depth. $\mathrm{NSIM}_1 = 0.1090$. $\mathrm{NSIM}_2$=0.3308.

(c) Octave dist. $\mathrm{NSIM}_1 = 0.1136$. $\mathrm{NSIM}_2$=0.3729.

(d) Small ICEF. $\mathrm{NSIM}_1 = 0.1052$. $\mathrm{NSIM}_2$=0.4753.

(e) No ICEF. $\mathrm{NSIM}_1 = 0.0849$. $\mathrm{NSIM}_2$=0.5229.

(f) 512-point FFT. $\mathrm{NSIM}_1 = 0.1123$. $\mathrm{NSIM}_2$=0.4035.

Figure 6.15: Superimposed model configurations: insertion depth 29 mm and octave band frequency-to-electrode mapping from Figure 6.10d-6.10f. $\mathrm{NSIM}_1$ refers to the original reference and $\mathrm{NSIM}_2$ to the filtered reference.

### Similarity analysis

The flute signal is more complex in its structure compared to the chirp which results in generally lower similarity scores for NSIM and NOPM. Also the STOI measurements are in the same range. However, calculating the standard deviation of the stochastic effects it is also of a lower order $10^{-4}$.

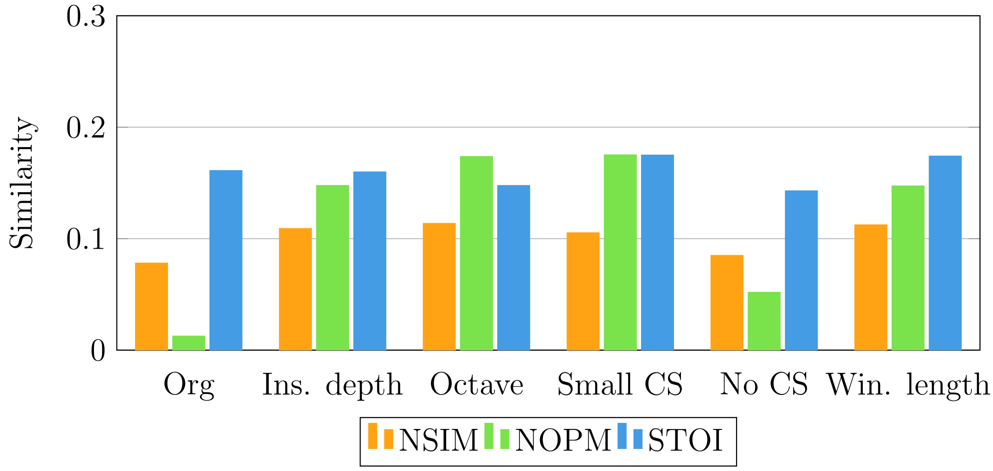
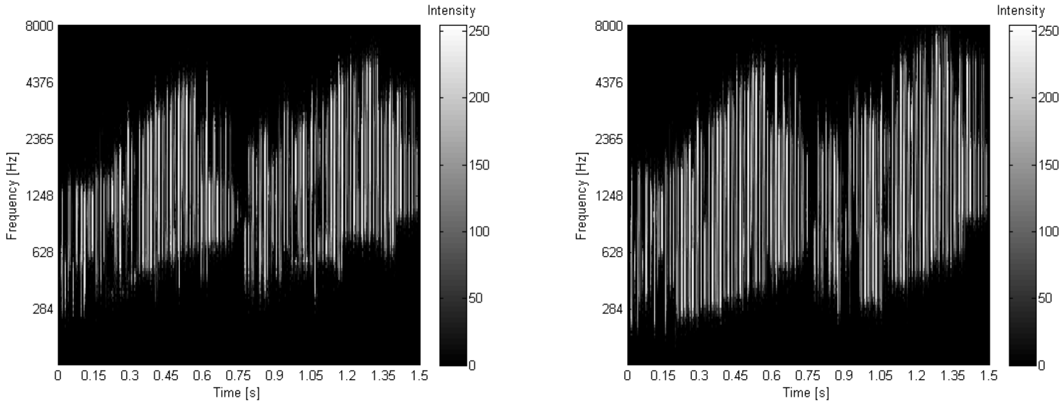


Figure 6.16: Bar diagram showing similarity between CI and normal hearing outputs for different CI model configurations with an input flute signal. Three different similarity measures are included to observe the relative changes; NSIM(orange), NOPM(green), STOI(blue).

## 6.2.2 N-of-M strategy



(a) No neighbour. $\text{NSIM}_1$=0.1136. $\text{NSIM}_2$=0.4531.

(b) Psycoacoustic. $\text{NSIM}_1$ =0.1142 . $\text{NSIM}_2$=0.4004.

Figure 6.17: Resulting neurograms for alternative N-of-M channel selection for the flute signal. In 6.17a no adjacent neighbours are allowed stimulation and 6.17b shows the result of psychoacoustic masking selection Both neurograms are configured with ins. 29 mm and octave band frequency-electrode distribution.

### 6.2.3 Compression

The XDP compression strategy aims at maximizing the user's dynamic range per frequency band and plays an important role in the frequency tuning. Each of the 20 available channels has an independent transfer function but are clustered to reduce the complexity of fitting. Figure 6.18 shows each channel's compression slopes and how the signal activity is mapped depending on the input level. A mapping above the KP results in a strongly compressed signal to eliminate the risk of overshooting. The neurogram response for a varying number of predefined clusters and KP's can be observed in Figure 6.19. The compression changes are expected to be seen in the intensity of the neural firings which are also affecting the current spread. However, when comparing the three clustering approaches in Figures 6.19b- 6.19d there are only very small (hardly visible) differences between the compression strategies.

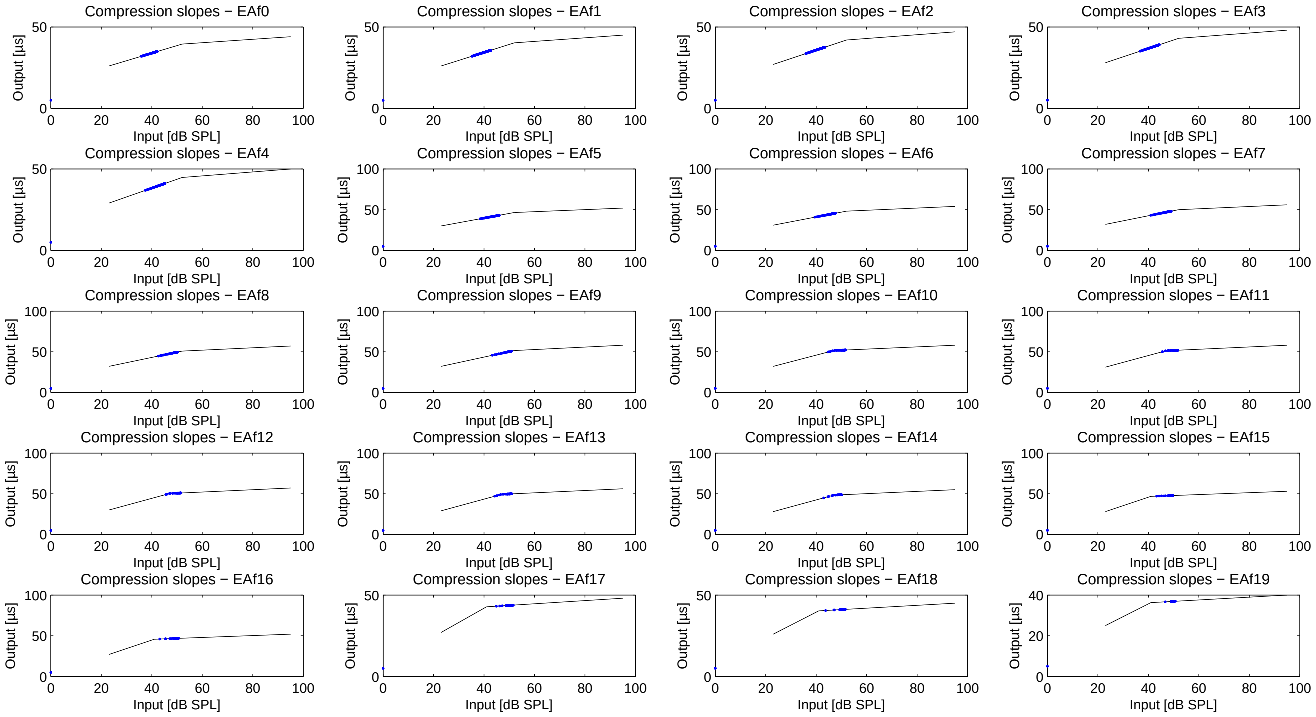
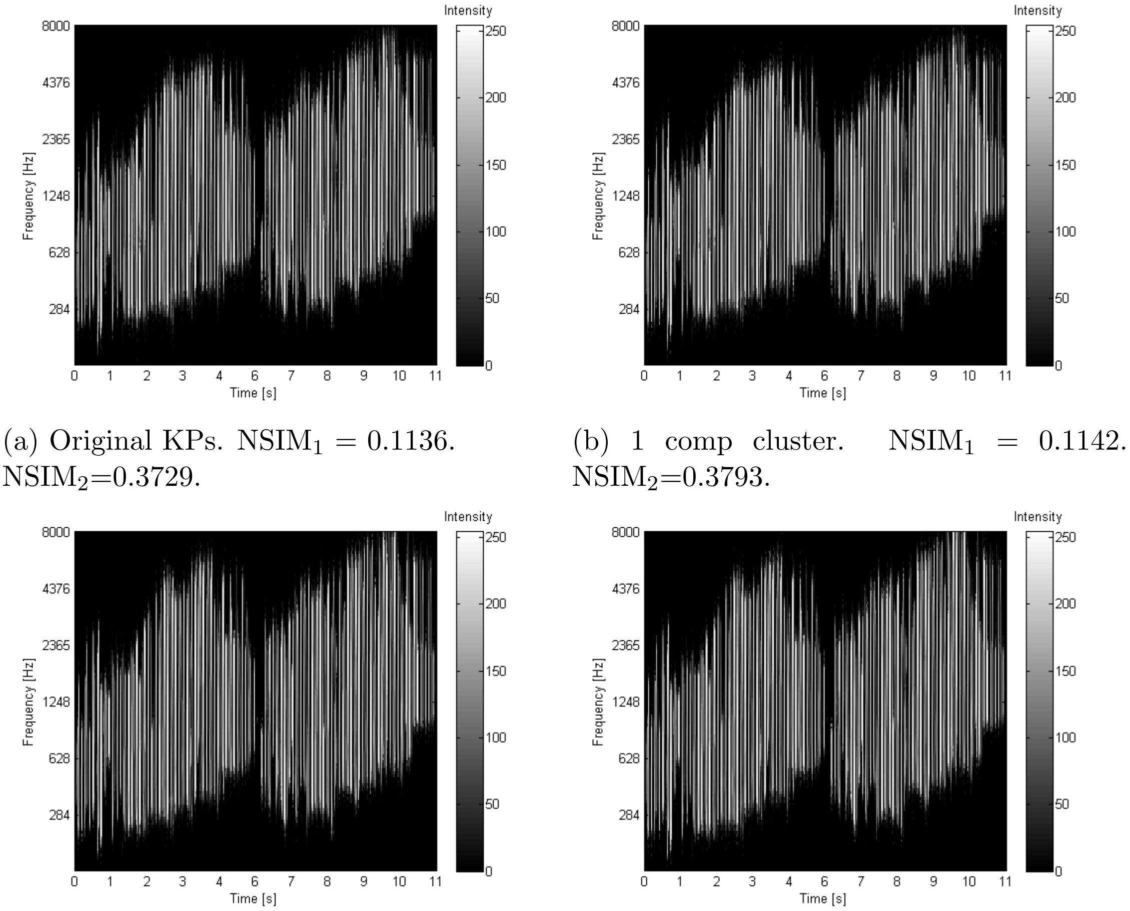


Figure 6.18: Compression slopes for each channel showing the transfer function mapping energy level to electrical stimulation. All four neurograms are configured with electrode insertion depth 29 mm and octave band frequency-to-electrode mapping.



(a) Original KPs. $\text{NSIM}_1 = 0.1136$. $\text{NSIM}_2$=0.3729.

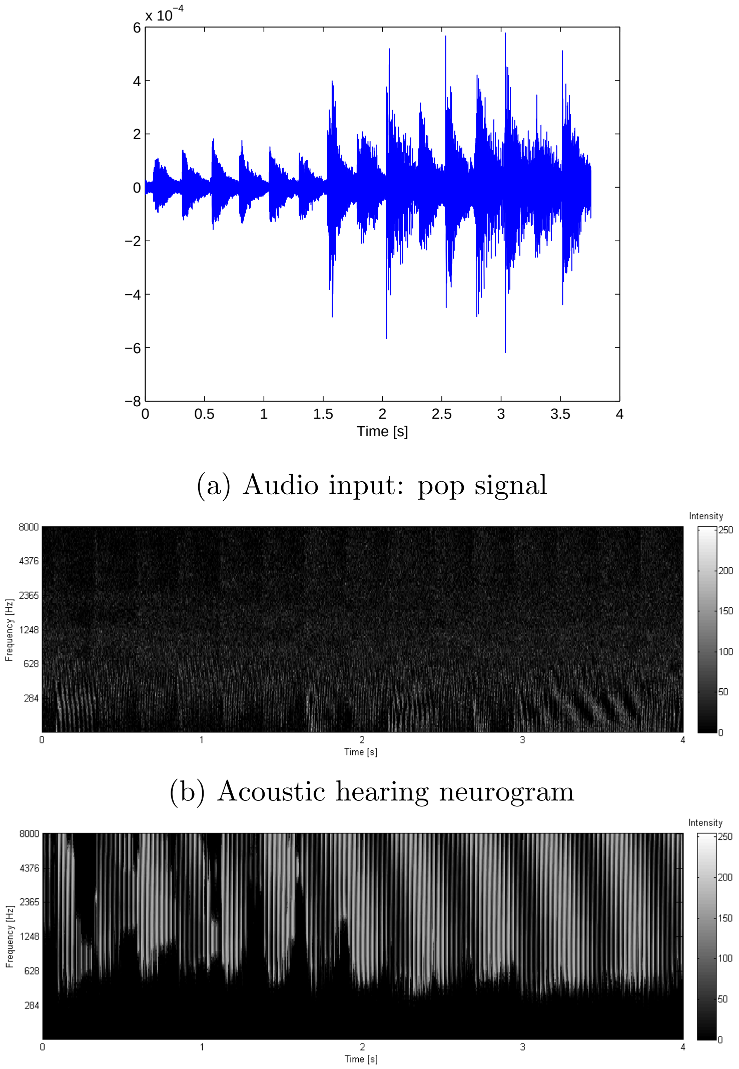
(b) 1 comp cluster. $\text{NSIM}_1 = 0.1142$. $\text{NSIM}_2$=0.3793.

(c) 4 comp. clusters. $\text{NSIM}_1 = 0.1126$ . $\text{NSIM}_2 = 0.3630$.

(d) 8 comp. clusters. $\text{NSIM}_1 = 0.1124$ . $\text{NSIM}_2 = 0.3607$.

Figure 6.19: Varying compression clustering and KPs for flute music input. Figure 6.19a illustrates the response for the original set of KPs.

## 6.3 Test signal: Pop



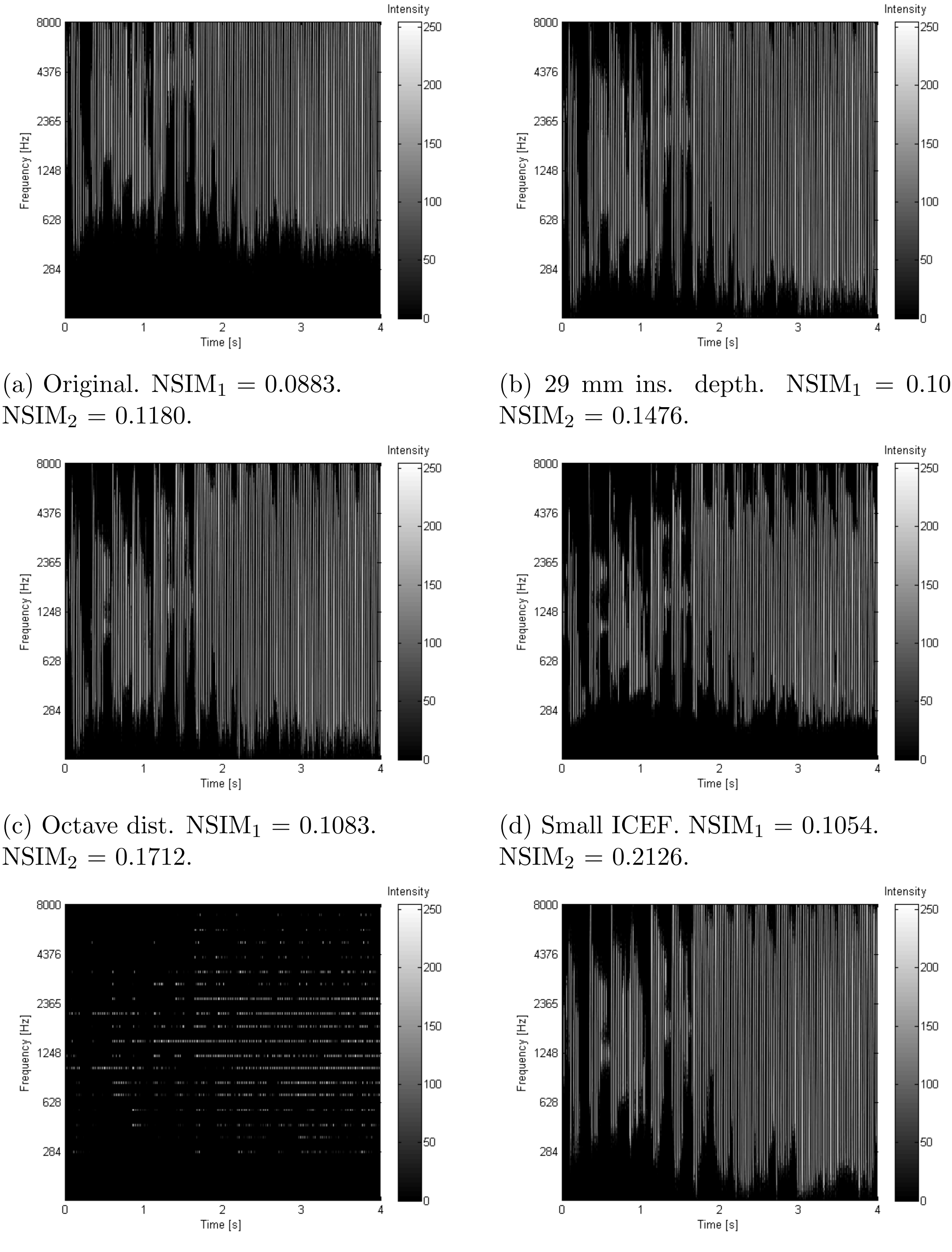
(a) Audio input: pop signal

(b) Acoustic hearing neurogram

(c) Electric hearing neurogram

Figure 6.20: $\text{NSIM}_1 = 0.0883$.

### 6.3.1 CI model configurations



(a) Original. NSIM$_1$ = 0.0883. NSIM$_2$ = 0.1180.

(b) 29 mm ins. depth. NSIM$_1$ = 0.1035. NSIM$_2$ = 0.1476.

(c) Octave dist. NSIM$_1$ = 0.1083. NSIM$_2$ = 0.1712.

(d) Small ICEF. NSIM$_1$ = 0.1054. NSIM$_2$ = 0.2126.

(e) No ICEF. NSIM$_1$ = 0.0835 . NSIM$_2$ = 0.2696.

(f) 512-point FFT. NSIM$_1$ = 0.1031. NSIM$_2$ = 0.1628.

Figure 6.21: Superimposed model configurations: insertion depth 29 mm and octave band frequency-to-electrode mapping from Figure 6.21d-6.21f. NSIM$_1$ refers to the original reference and NSIM$_2$ to the filtered reference.

### 6.3.2 Similarity analysis

Implementing the CI model configurations it can be seen in Figure 6.22 that each of the three objective measures calculate a very small similarity score. For NSIM and NOPM it holds that increasing the insertion depth together with changing the frequency-to-electrode distribution to octave band always gives the largest positive similarity. In this case the STOI measure results in unrealistic values.

Comparing similarity with the psychoacoustic selection strategy, NSIM actually shows an improvement of the order $10^{-3}$ from configuration three in Figure 6.22 where $\text{NSIM}_1$=0.1083 to $\text{NSIM}_1$=0.1142 in Figure 6.23b. Visually there can also be seen a change in Figure 6.23b as the neurogram is more sparse between 2 s and 4 s.

As NSIM was not able to capture the relatively large changes in 6.22 it is reasonable to assume that neither the changes in the compressor are traceable. Studying Figure 6.24 NSIM reveals no significant differences among the neurograms.

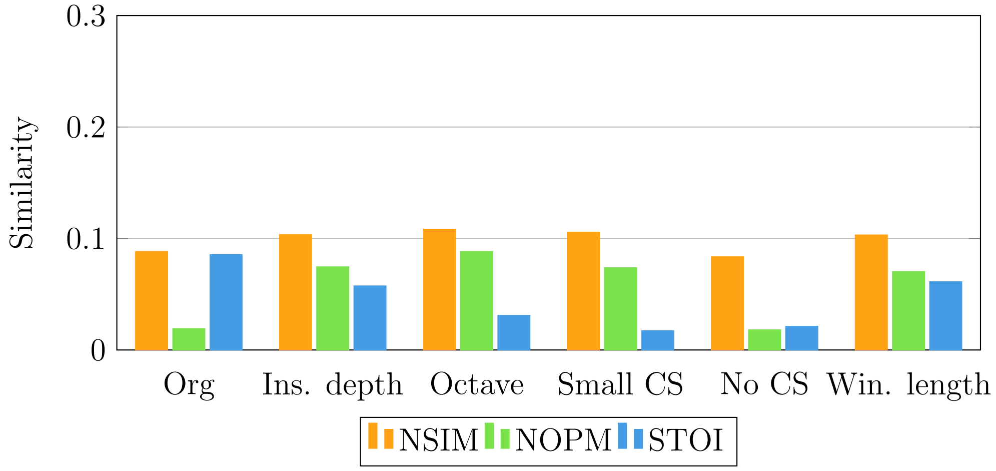
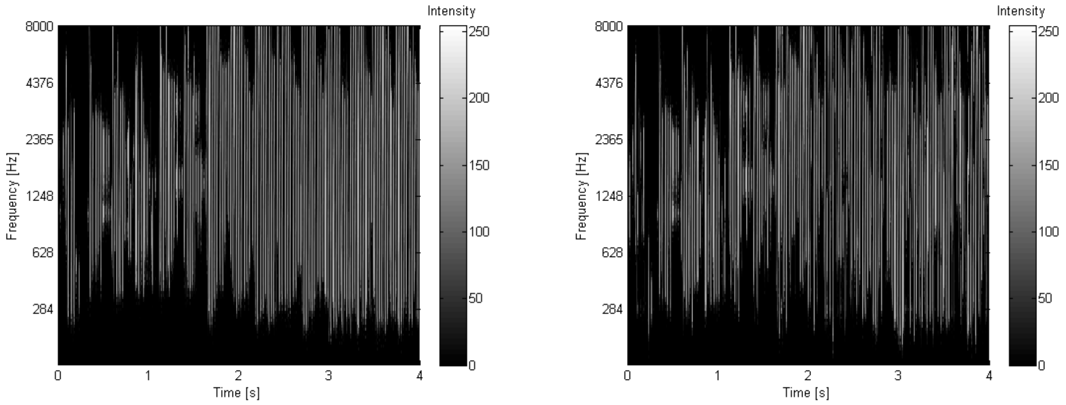


Figure 6.22: Bar diagram showing similarity between CI and normal hearing outputs for different CI model configurations with an input pop signal. Three different similarity measures are are included to observe the relative changes; NSIM(orange), NOPM(green), STOI(blue).

### 6.3.3 N-of-M strategy

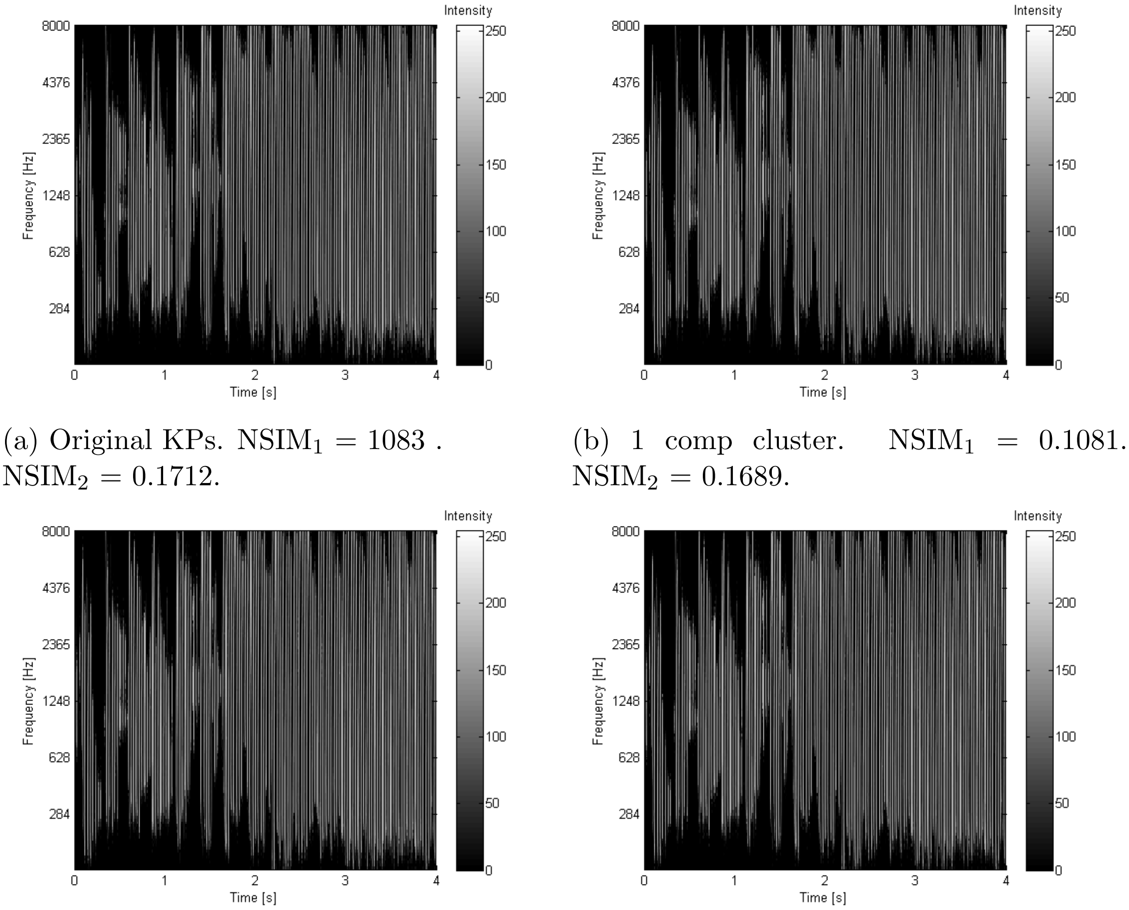


(a) No neighbour. $\text{NSIM}_1 = 0.1090$. $\text{NSIM}_2 = 0.1897$.

(b) Psycoacoustic. $\text{NSIM}_1 = 0.1142$. $\text{NSIM}_2 = 0.2000$.

Figure 6.23: Resulting neurograms for alternative N-of-M channel selection for the pop signal. In Figure 6.23a no adjacent neighbours are allowed stimulation and Figure 6.23b shows the result of psychoacoustic masking selection. Both neurograms are configured with ins 29 mm and octave band frequency-electrode distribution.

### 6.3.4 Compression



(a) Original KPs. NSIM$_1$ = 1083 . NSIM$_2$ = 0.1712.

(b) 1 comp cluster. NSIM$_1$ = 0.1081. NSIM$_2$ = 0.1689.

(c) 4 comp. clusters. NSIM$_1$ = 1079. NSIM$_2$ = 0.1664.

(d) 8 comp. clusters. NSIM$_1$ = 0.1075. NSIM$_2$ = 0.1659.

Figure 6.24: Varying compression clustering and KPs for pop music input. Figure 6.24a illustartes the response for the original set of KPs.

CHAPTER 7

# Conclusions and future work

The MAP model, simulating the steps of the auditory periphery, and the CI model, coding for direct electrical stimulation of the auditory nerves, are very different to their nature. In this project it has been assumed that the two model outputs can be aligned and compared by extending each of the paths with a third model simulating auditory nerve firings. When evaluating the normal hearing and electrical hearing neurograms respectively using NSIM the absolute similarity score shows to be highly dependable on the complexity of the input signal.

NSIM was originally developed to measure speech degradation for hearing impairment. To the author's knowledge it has never been used to measure likeness between normal hearing conditions and CI hearing before. In this work several limitations on NSIM has been documented. The controlled changes in the CI model were set up to see if we were able to predict the behaviour of the objective measure and thus if it was able to register the type of changes we want to evaluate in the CI chain. For example we expected a significant increase in NSIM when the electrode array was inserted 29 mm and frequency-to-electrode mapping was performed according to an octave distribution. Even if the NSIM likeness increases of order $10^{-2}$ for all three test signals, the similarity change is very small to what was expected. Further analysis show that no significant change are calculated in NSIM when decreasing the current spread or increasing the frequency resolution by applying a longer FFT window.

It has been shown that NSIM is very sensitive to noise and thus the spontaneous activity in the normal hearing reference. When using the original reference neurogram NSIM calculated a much lower similarity than when using the filtered reference. The main reason for this is that the two backgrounds compared are completely different, i.e. comparing black and greyish (spontaneous activity) background NISM scores low likeness while comparing two neurograms with both a large amount of black regions generates high likeness. This restriction makes it very difficult to evaluate the signal response in the neurograms.

As an alternative to NSIM two other objective measures have been tested; NOPM and STOI. NOPM which aims at capturing smaller intensity changes efficiently, shows larger positive and negative changes than NSIM for the chirp and flute signal. STOI is a different evaluation approach completely bypassing the normal hearing reference and instead using a spike based vocoder to generate back an audio signal which is then compared to the unprocessed signal. STOI has been shown to successfully predict speech intelligibility in noise. In this application the STOI results are difficult to draw a conclusion around since it there is nothing saying that the processing CI signal has the input stimuli as optimal reference. The output from the vocoder tells us how correlated the processed signal from the CI is with the input stimulus, but does not necessarily give information about how the sound is perceived in the brain.

The large insecurity of evaluating NSIM on only a few audio signals makes it difficult to measure the performance of the music compression and the alternative N-of-M selection. Visually, the small changes can be discriminated in the neurograms as intensity changes.

Finally, here we have assumed that no relevant results will be generated unless the CI output is aligned in both time and frequency with the Greenwood function. It is important to keep in mind that using this assumption we are moving away from the configurations of the typical patient and thus no longer simulating the up-shifting in frequency that is most likely occurring for many CI patients, before they can learn to adapt for the basal shifts to some degree through training.

## 7.1 Future work

This thesis has focused on the validation of the MAP and CI chain respectively as well as simulating strategies optimal for music input. For future recommendations both models can be improved. In the results it can be observed that Goldwyn generates a different spike pattern than the MAP model. One could consider using the MAP model for simulating the complete normal hearing chain, without the attachment of Goldwyn. This would assure simulation of all steps of the auditory pathway (including the afferent attenuation).

As for the CI model it also needs to be tested further. One suggestion for improvement is to include the modelling of the ICEF historic current potential effects by using data from previous time frames. To avoid the large variance in using two models, a different reference can be implemented where an optimal CI configuration can be established. There is certainly not one true electric reference but with further testing some optimal parameters can be establishing.

The objective evaluation measure NSIM has shown to give inconclusive results. To further evaluate its performance it needs to be applied on a larger amount of test signals with different characteristics. Alternative objective measurement could be further analyzed in parallel and adapted to the type of change in the CI model that are to be tested.

There are several improvements in the CI implementation that could be done for music. Extracting the temporal fine structure (TFS) from the amplitude signal is one of the current research areas to improve pitch perception. Further evaluation of the proposed compression analysis is also needed on genre classified music signals with longer duration. Although NSIM did not indicate an improvement when using the alternative psychoacoustic N-of-M strategy the method aims at describing the natural psychoacoustic selection and should be investigated further, both for speech and music signals.

# Bibliography

[1] F. Baumgarte, C. Ferekidis, and H Fuchs. A nonlinear psychoacoustic model applied to the ISO MPEG layer 3 coder. *Audio Eng. Soc Convention*, 4087, 1995.

[2] G. Clark. *Cochlear implants: Fundamentals and applications*. Springer-Verlag New York, Inc, 2003.

[3] EBU. Sound quality assessment material recordings for subjective tests. *EBU – TECH 3253*, 2008. Geneva.

[4] A. Frater. Development and objective evaluation of EAS and cochlear implant model. Master's thesis, Aalborg Universitet, Ins. of Electronic Sys., Electronics and Information Tech., June 2014.

[5] J. H. Goldwyn, J. T. Rubinstein, and E. Shea.Brown. A point process framwork for modeling electrical stimulation of the auditory nerve. *Journal of Neurophysiol*, 108:1430–1452, June 2012.

[6] D. D. Greenwood. A cochlear frequency-position function for several species-29 years later. *J. Acoust. Soc. Am.*, 87(6), June 1990.

[7] A. Hines and N. Harte. Speech intelligibility prediction using a Neurogram Similarity Index Measure. *Elsevier Speech communication*, 54:306–320, 2012.

[8] D. Jethanamest, C-T Tan, M. B. Fitzgerald, and M. A. Svirsky. A new software tool to optimize frequency table selection for cochlear implants. *Otol Neurotol.*, 31(8):1242–1247, 2010.

[9] K. Kasturi and P. C. Loizou. Effect of filter spacing melody recognition: Acoustic and electrical hearing. *JASA Express Letters, Acoustical Society of America*, 122(2), August 2007.

[10] P. C. Loizou. Mimicking the human ear. *IEEE Signal Processing Magazine*, September 1998.

[11] N. Mamun, W. A. Jassim, and M. S. A. Zilany. Prediction of speech intelligibility using a neurogram orthogonal polynomial measure (nopm). *IEEE/ACM Trans. Speech and Audio Processing*, 23(4), April 2015.

[12] H. J. McDermott. Music perception with cochlear implants: A review. *Trends in Amplification*, 8(5), 2004.

[13] R. Meddis. MATLAB model of the auditory periphery (MAP). URL `http://www.essexpsychology.macmate.me/HearingLab/modelling.html`. Essex Hearing Research Laboratory.

[14] B. C. J. Moore. *An introduction to the psychology of hearing.* Emerald, sixth edition edition, 2012.

[15] W. Nogueira, A. Büchner, T. Lenarz, and B. Edler. A psychoacoustic ”NofM” -type speech coding strategy for cochlear implants. *EURASIP J. Appl. Signal Processing*, 18:3044–3059, 2005.

[16] W. Noguiera, M. Haro, P. Herrera, and X. Serra. Music perception with current signal processing strategies for cochlear implants. *Music Technology Group, Pompeu Fabra Univeristy*, 2011.

[17] M. Segovia-Martinez and D. Gnansia. Design and effects of post-spectral output compression in cochlear implant coding strategy. Oticon Medical, 2013.

[18] O. Stakhovskaya, D. Sridhar, B. H. Bonham, and P. A. Leake. Frequency map for the human cochlear spiral ganglion: Implications for cochlear implants. *J. Assoc. Res. Otolaryngology*, February 2007.

[19] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen. A short-time objective intelligibility measure for time-frequency weighted noisy speech. *IEEE/I-CASSP 2010*, 2010.

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P Simoncelli. Image quality assessment: From error visibility in structural similarity. *IEEE Trans. on Image Processing*, 13(4), April 2004.

[21] J. Wolfe. Speech and music: acoustics, signals and the relation between them. *Proceedings of ICoMCS, Sydney, Australia*, December 2007.

[22] W. A. Yost. *Fundamentals of hearing.* Academic Press, fourth edition, 2000.

[23] F-G. Zeng. Trends in cochlear implants. *Trends in Amplification*, 8(1), 2004.