Master thesis

Estimation of the Phase Response of Auditory Filters

by

Katharina Zenke

(0773083)

Institute for Electronic Music and Acoustics

University of Performing Arts and Music Graz

Acoustic Research Institute

Austrian Academy of Sciences

Master's Programme "Elektrotechnik-Toningenieur" (V 066 413)

Supervisor: Dr. Bernhard Laback

Assessor: o. Univ. Prof. Dr. Robert Höldrich

Graz, December 2014

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

....

Date **Katharina Zenke**

Acknowledgements

First and foremost, I want to thank my supervisor Bernhard Laback for all the productive and motivating discussions and the great advice and also for the opportunity to be part of this project.

I also want to thank the colleague from the Acoustic Research Institute, in particular from the Audiology Group, for their warm welcome and their help during my stay in Vienna.

I specially want to thank my parents for all the support and encouragement throughout the past years. I wish you all the best for your new independence - in every sense!

Finally I want to thank all my friends and colleague in Graz for the wonderful time.

Abstract

It has often been assumed that the human auditory system is insensitive to differences in the relative phases of spectral components of a multicomponent sound. Recent studies have proven this wrong. Especially at higher frequencies and for signals with a spectral range covering a single auditory filter, strong phase effects can be observed. The project BiPhase at the Acoustics Research Institute in Vienna involves a series of studies aiming to better understand human phase sensitivity. In particular, different hypotheses about the mechanisms underlying phase effects are studied by testing both normal-hearing and hearing impaired subjects. A new approach for studying phase effects, based on measures of the sensitivity to interaural time differences, is tested in addition to the standard approach based on a masking task. As part of the BiPhase project, this master thesis developed and evaluated a method to measure the cochlear phase response that is applicable in cases of nonuniform phase curvatures of underlying auditory filters, such as occurring in hearing-impaired listeners. The underlying assumption was that stimuli with phase relations causing a more peaky internal temporal representation, after passing the phase response of auditory filtering, cause a stronger ITD cue and are thus more lateralized than flat waveforms. Variation of the phase relations of the stimulus should thus enable to infer the phase response of the auditory filters. The experiment used multicomponent harmonic stimuli with variable phases based on Schroeder phase harmonic complexes. The stimuli were presented with a large envelope ITD. In two consecutive stages the most lateralized stimuli were determined individually for each subject. Four out of eight subjects appeared to show some systematic differences between phase configurations. While the results show a large amount of inter-subject variability, the curvature of the estimated phase response appears to be positive for most subjects, which is consistent with other studies.

Kurzfassung

Lange Zeit wurde angenommen, dass das menschliche Hörsystem Unterschiede in der relativen Phasenlage spektraler Komponenten eines Klangkomplexes nicht wahrnimmt. Neue Studien haben dies wiederlegt. Besonders im hochfrequenten Bereich und für Signale im spektralen Bereich eines auditiven Filters können starke Phaseneffekte beobachtet werden. Das Projekt BiPhase am Institut für Schallforschung in Wien umfasst eine Reihe von Studien zur Untersuchung der menschlichen Phasensensitivität. Speziell werden verschiedene Hypothesen über den zugrundeliegenden Mechanismus der Phaseneffekte geprüft. Dazu werden Experimente sowohl mit normalhörenden als auch mit hörgeschädigten Testpersonen durchgeführt. Zusätzlich zum standardmäßigen Verfahren durch Maskierungsexperimente, wird ein neuer Ansatz für die Untersuchung der Phaseneffekte, der auf der Ermittlung der Sensitivität zu interauralen Zeitdifferenzen basiert, getestet. Als Teil dieses Projektes wurde in dieser Masterarbeit eine Methode zur Messung der Phasenantwort der Cochlea entwickelt und ausgewertet. Diese Methode ist auch bei ungleichmäßigen Phasenkrümmungen der zu Grunde liegenden auditiven Filter, so wie es bei hörgeschädigten Personen der Fall ist, anwendbar. Die zugrundeliegende Annahme war, dass Stimuli mit Phasenlagen, welche nach der Phasenfilterung des auditiven Systems eine zeitlich spitzenreiche interne Repräsentation erzeugen, einen stärkeren ITD Cue erzeugen und dadurch stärker lateralisiert werden als flache Wellenformen. Das Variieren der Phasenlagen sollte es somit erlauben, Rückschlüsse über die Phasenantwort des auditiven Filters zu ziehen. Dafür wurden im Experiment harmonische Stimuli mit veränderbaren Phasen verwendet, die auf sogenannten Schroeder-Phasen-Signalen basieren. Die Stimuli wurden mit einer großen Einhüllenden-ITD versehen. In zwei aufeinander aufbauenden Experimentstufen wurde der weitest lateralisierte Stimulus für jede Testperson ermittelt. Vier der acht Testpersonen schienen systematische Unterschiede in der Lateralisation zwischen den Variationen der Phase aufzuweisen. Obwohl die Ergebnisse starke Unterschiede zwischen den Testpersonen zeigen, scheint die Phasenkrümmung der geschätzten Phasenantwort für die meisten Testpersonen positiv zu sein, was mit den Ergebnissen anderer Studien übereinstimmt.

Contents

1 Introduction

The human auditory system is amazing. It is able to receive various sound signals at the same time, process them and gain complex information about each of them like the physical properties of the signal or the location of the sound source in space. The processes responsible for the detection of these pieces of information are distributed over different stages of the auditory system. In this work we will focus on the perception of one signal property: the phase.

1.1 Field of study

For a long time researchers assumed that the cochlea of the auditory system is filtering sound signals by amplitude but not by phase. Thereby the mechanism causing amplitude filtering and the perceptional consequences of amplitude changes have been studied extensively. The fact that the auditory system is also able to perceive and process phase relations of sound signals wasn't proven until the middle of the 20th century. The human hearing system cannot perceive absolute phase values but is able to resolve relations in phase between different components of one sound within the spectral range of one auditory filter of the cochlea. This phase dependency has been studied in various experiments, which discovered that the human cochlea has a specific phase response for each auditory filter that filters the incoming sound and leads to a phase-changed internal signal. These phase responses were found to be nonlinear for normal hearing (NH) listeners. Each auditory filter seems to have a certain constant curvature of the phase response, which is the same for all listeners.

However, these findings don't apply for listeners that are hearing impaired (HI). Hearing impairment and in particular cochlear hearing impairment changes many perceptional processes including the phase filtering of the cochlea. Methods used to measure the phase response for normal hearing listeners, for example masking tasks, are less effective or not applicable for hearing impaired listeners due to their complex hearing limitations. Thereby new methods have to be developed to examine the phase responses for hearing impaired listeners.

1.2 Project background

The BiPhase project is a research project currently conducted at the Acoustic Research Institute in Vienna that attempts to determine the phase responses of normal hearing and hearing impaired listeners by means of a new paradigm based on the perception of interaural time differences (ITDs). ITDs enable listeners to localize sound sources in the horizontal plane, particularly in the left versus right dimension. ITD perception is divided into fine-structure ITD for low frequencies and envelope ITD for modulated high frequencies. The ITD sensitivity of HI listeners is reduced compared to NH listeners but may, especially in the case of envelope ITD, still be an important cue for localization that can help to determine the individual phase response of a listener.

This thesis, as part of the BiPhase project, aims to develop an efficient method to estimate the individual phase response of listeners and particularly allow for measuring phase curvatures that are not constant, as expected to be the case in HI listeners and for some configurations in NH listeners. The method shall thereby be equally suitable for NH and HI listeners. Since the experiment is meant to provide just a rough estimation, it has to be performable in a relatively short time and not require extensive training beforehand. The experiment will be using ITD cues to distinguish between complex stimuli that are consisting of equal components with varied phase relations. The experiment may be conducted in advance to a longer experiment for a quick estimation of the phase curvature of an auditory filter.

1.3 Aim of the project

The general aim of phase response studies is to get a better understanding about the processing of sound in the auditory system and thereby be able to model this complex system in a better way. Many spectro-temporal models of auditory perception have been developed so far. Those models yield a good representation of the amplitude filtering but still have difficulties with representing the phase filtering. Often their simulation data don't agree with experimental results. These models would profit from a more detailed knowledge about the phase filtering in the cochlea.

Especially in the case of hearing impaired listeners it would be beneficial to gain knowledge about their phase responses, since they are individually different and can't be

estimated by general experimental results. This knowledge could be used to better simulate their individual hearing impairment and thereby to better understand their individual perception and also to contribute to the improved performance of hearing devices such as hearing aids or cochlear implants. Hearing devices so far only consider amplitude enhancement but could also incorporate individual phase filtering to be more effective in restoring an internal signal representation such as occurring in NH listeners.

1.4 Structure of the thesis

In the first chapters the physiological and psychoacoustic mechanisms that are necessary to detect phase relations will be presented. Chapter 2 will describe the structure of the human auditory system and the functionality of each part. The focus will lie on the cochlear processing of sound, since this is the stage in which we expect the phase filtering. Hearing impairment often originates inside the cochlea. This type of damage and its consequences on perception will be introduced in Chapter 2.3. The following Chapter 3 will give an overview over localization cues and especially treat ITD detection, which will be used in the experiment. Chapter 4 presents previous studies on the human phase response.

The next two chapters describe the actual experiment. In Chapter 5 the experimental planning is presented, consisting of the general approach towards the topic, the used method and stimuli and the process of conducting the experiment. Chapter 6 provides the results of the experiment. In Chapter 7 the experimental results are discussed.

The last Chapter 8 summarizes the experiment and provides some conclusions.

2 The human auditory system

This chapter describes the physical, physiological and psychological aspects of sound perception by the human auditory system [Moore 2012, Laback 2010]. It will explain the filtering effects of the cochlea, focusing on the phase response of auditory filters and the consequences of hearing impairment on these effects.

2.1 Structure of the auditory system

The auditory system consists of the peripheral and the central auditory system. Outer, middle and inner ear (see Figure 2.1) form the peripheral part of the auditory system. They transform the acoustic waves arriving at the listener's ear positions into neuronal potentials that are led to the central nerve system. The outer and middle ear are often referred to as the conductive system, since they conduct the sound signal towards the inner ear. The inner ear and the connected vestibulocochlear nerve likewise are called the sensorineural system [Gelfand 1997].

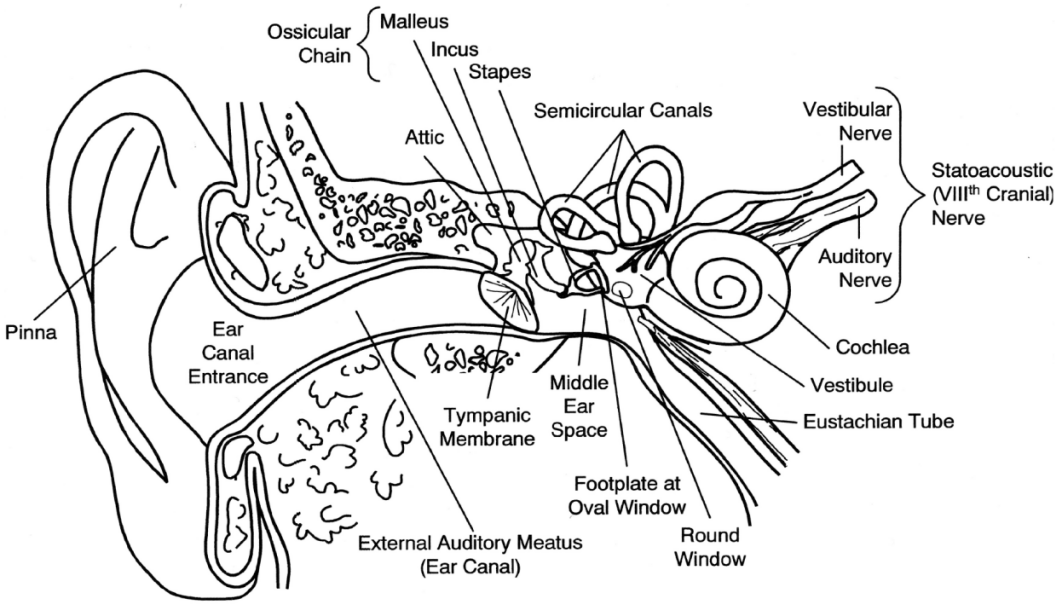


Figure 2.1: Peripheral part of the human auditory system [Gelfand 1997].

Outer Ear

The outer ear is consisting of the pinna and the ear channel (formally known as external auditory meatus).

The incoming sound is bundled at the pinna, the visible part of the hearing system, which is mostly made out of elastic cartilage. The pinna operates like an acoustic funnel. The sound waves are reflected and attenuated. The overlapping of direct and reflected diffuse sound based on the inhomogeneous structure of the pinna generates a directional filtering of the signal, by which individuals are able to localize sound sources in the vertical plane. Only sound sources with a signal period smaller than the pinna dimensions are filtered directionally (beyond 4-5 kHz). For lower frequencies, the form of head and upper body are important for the localization. The pinna has a resonance at about 3 kHz, which can be seen in Figure 2.2.

The entrance to the ear channel is on the base of the concha, a resonant cavity of the pinna. The ear channel transmits the sound signal to the ear drum of the middle ear. It is slightly bent for protection of mechanical damages and has a directional resonance at 2-3 kHz. In combination pinna and ear channel lead to a resonance peak of up to 12 dB at around 2.5 kHz.

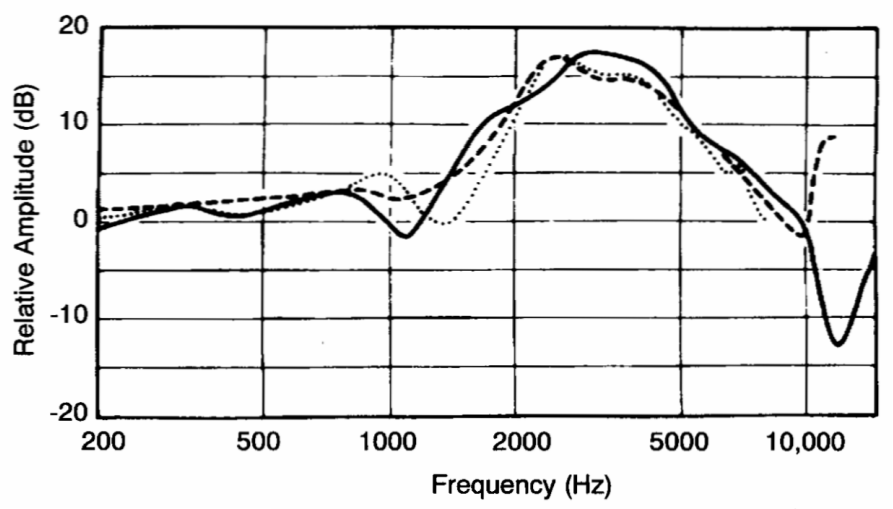


Figure 2.2: Head related transfer functions for a frontal sound source for three different subjects [Gelfand 1997].

The micro structure of pinna shapes can differ strongly across individuals. The frequency and directional filtering of pinna and ear channel (displayed in Figure 2.2), known as head related transfer function (HRTF), thereby also differ across individuals.

Middle Ear

The middle ear is located inside the tympanic cavity. It consists of the ear drum (tympanic membrane), the ossicular chain (malleus, incus and stapes) and the oval window (see Figure 2.1).

The middle ear connects the air-filled ear channel of the outer ear with the fluid-filled cochlea of the inner ear. Therefore its task is to achieve the impedance matching for acoustic waves between these two conditions. In air a small force leads to a big displacement. In liquid much higher forces, i.e. vibrations with higher pressure, are needed to achieve the same displacement.

Over the chain of ossicules the ear drum is connected to the oval window. The sound waves that travel through the ear channel of the outer ear oscillate the ear drum. These vibrations are transmitted over malleus (also called "hammer"), incus ("anvil") and stapes ("stirrup") towards the oval window. The impedance matching is achieved mostly by the proportion of the big ear drum membrane to the smaller membrane of the oval window. To a smaller amount also the leverage of the long hammer grip towards the short anvil membrane and the curvature of the eardrum add to this matching. In total a pressure gain by the factor 50 is reached in the most efficient range of 500 to 4000 Hz[1] [Gelfand 1997].

Inner Ear

The inner ear contains the oval window and the cochlea. The cochlea is 33-35 mm long, tube-like and rolled-up like a snail. It consists of three fluid-filled chambers: scala vestibuli, scala media and scala tympani (see Figure 2.3).

[1] That change in pressure would correspond to an acoustic level difference of 33 dB [Gelfand 1997].

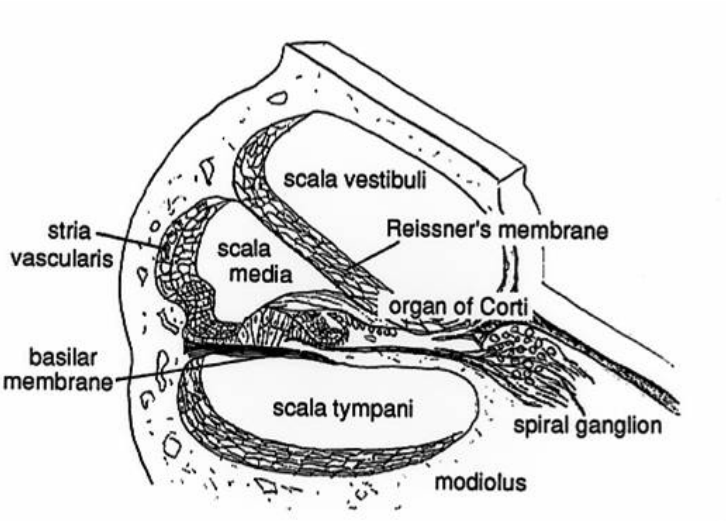


Figure 2.3: Cross section of the cochlea [University Minnesota Duluth].

The scala vestibuli is connected with the stirrup of the middle ear via the membrane at the oval window. Scala vestibuli and scala media are separated by the membrane of Reissner. The scala media is located in the middle of the cochlea and contains the organ of Corti. Scala media and scala tympani are separated by the basilar membrane. On the base of the cochlea the scala tympani is connected to the round window for pressure compensation. Scala tympani is also connected to scala vestibuli over the apex at the end of the cochlea. They are both filled with perilymph whereas scala media is filled with endolymph[2]. That leads to an electric potential difference of -40 mV at the basilar membrane[3]. This difference is used to transform the mechanical information of the acoustic wave into electric impulses.

The mechanical oscillation of the oval window leads to a periodically changing pressure difference between scala vestibuli and scala tympani and thereby to an excitation along the basilar membrane. This wave pattern on the membrane is called the traveling wave. As shown in Figure 2.4 the wave starts at the base of the cochlea and gains in amplitude until it reaches its maximum and thereafter declines quickly.

[2]Perilymph contains a low concentration of calium and a high concentration of natrium. In reverse endolymph contains a high calium and low natrium concentration. The electric potential of endolymph is 80-100 mV more positive than perilymph. This difference is called “endocochlear potential”.

[3]The difference at the basilar membrane is called “intracellular potential”.

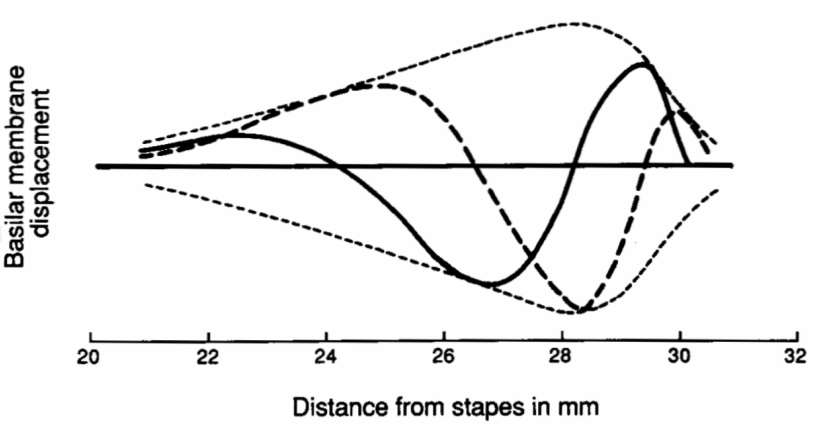


Figure 2.4: Traveling wave [Gelfand 1997].

The maximum of the excitation along the basilar membrane depends on the frequency of the signal. Higher frequencies are mapped near to the base of the membrane, where the membrane is very stiff, low frequencies near the apex, where the basilar membrane is broader and the stiffness lower (see Figure 2.5). Thereby each position at the basilar membrane has its own "characteristic frequency". This frequency-place representation is called tonotopy.

The movement of the basilar membrane can be measured in velocity or displacement as a function of stimulus level. Figure 2.5 shows the traveling wave of the membrane.

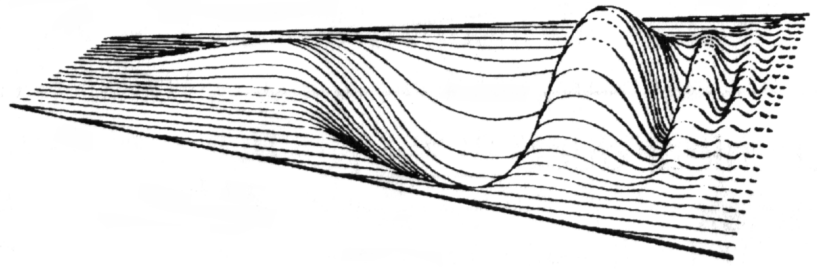


Figure 2.5: Dispersion of the traveling wave on the basilar membrane [Gelfand 1997].

The transformation of the signal into neuronal impulses is realized by hair cells that are part of the organ of Corti and attached to the basilar membrane. There are two types of hair cells that serve different purposes: inner and outer hair cells. In total

there are one row of inner hair cells (approx. 3500 cells) and three rows of outer hair cells (approx. 12000 cells) on the basilar membrane.

The inner hair cells transmit information towards the brain. They transform the signal, i.e. the movement of the basilar membrane, into neuronal action potentials that are transferred to the central nerve system by afferent nerve fibers of the vestibulocochlear nerve. A larger movement of the membrane due to a higher level of the acoustic signal leads to a larger stimulation of the inner hair cells and triggers them to a stronger firing of action potentials.

The outer hair cells on the other hand are steered both by mechanical sensors tracking basilar membrane movement and by action potentials coming from the central nerves. They are electromotil and thereby able to affect the motions of the basilar membrane. They receive neural signals via efferent nerve fibers and shorten and lengthen themselves in response. Outer hair cells are also referred to as "cochlea amplifier" since they are assumed to be responsible for the cochlear compression (more details in Chapter 2.2) and thereby produce high sensitivity and sharp tuning.

Vestibulocochlear nerve

The vestibulocochlear nerve is the 8th cranial nerve. It connects the peripheral auditory system with the central nervous system. It consists of two nerves, the cochlear nerve that transmits the signal information and the vestibular nerve that transmits equilibrium sense information, which is also resolved in the peripheral auditory system.

Central auditory system

The sound information transmitted via the cochlear nerve is processed neuronal at several stages of the central auditory system. The nerve passes through the cochlear nucleus[4], the superior olivary complex[5] of the brainstem and the inferior colliculus of the midbrain, the latter two particularly resolving spatial information of the sound. From there the signal is transmitted to the thalamus and the auditory cortex, where further characteristics of the signal are ascertained. A schematic overview over the

[4]The cochlear nucleus is divided into two parts, the dorsal and the ventral cochlear nucleus.

[5]The superior olivary complex likewise consists of two separate parts, the lateral superior olive and the medial superior olive.

central auditory system can be seen in Figure 3.3 in Chapter 3.2, where the neuronal processes for sound localization are explained in detail.

2.2 Cochlear filtering

A deeper understanding of the cochlea's functionality can give insight into many aspects of sound perception. Since the signal is split into its spectral components on the basilar membrane, the cochlea can be regarded as a sort of frequency analyzer using the hair cells at the basilar membrane as a bank of overlapping bandpass filters. These filters are called auditory filters. Each auditory filter is placed around a center frequency, which is the characteristic frequency for this specific location on the basilar membrane. The bandwidth of an auditory filter is also called "on-frequency" range. Frequencies beyond this range are accordingly called "off-frequencies" for this specific auditory filter. The bandwidth of these filters is equivalent to the spectral distance within which frequencies are able to mask each other. Masking describes the effect of one sound signal reducing the audibility of another signal when presented simultaneously or within a short temporal distance. The signals have to be in the same frequency region or have significance differences in their sound pressure level. In these cases the sensitivity for the second signal is reduced or suppressed completely. One important contributor to the effect of masking is the spread of excitation of a given narrowband signal on the basilar membrane: spectrally close frequencies are also spaced narrowly on the basilar membrane, leading to interference and thereby swamping of each others' maxima. The bandwidth of this effect is rising with frequency as can be seen in Figure 2.6. The factor, by which the width of the auditory filters rises, stays constant over a wide frequency range. The border between on- and off-frequencies is approximately at $0.7 \cdot CF$ and $1.3 \cdot CF$. Each frequency can be regarded as center frequency of an auditory filter centered at the position on the basilar membrane.

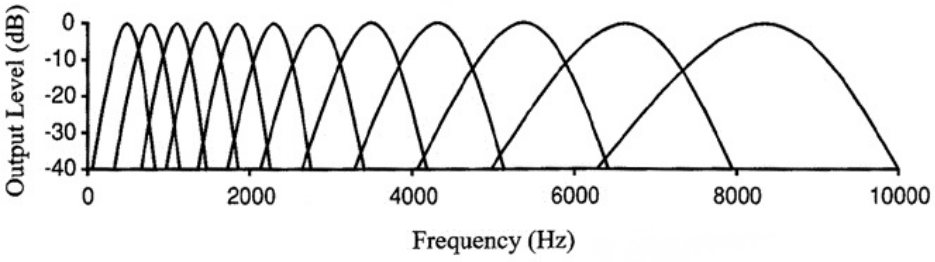


Figure 2.6: Auditory filters with characteristic frequencies from 500 to 8500 Hz [Columbia College Chicago].

Thus, if the spectral components of a complex sound differ largely they are represented at different places along the basilar membrane and each evokes a separate wave pattern. In this case the cochlea acts as a Fourier analyzer. When frequencies are close to each other, their vibration patterns interact and the resultant waveform is more complex. Frequencies even closer lead to a fusion of corresponding maxima of excitation along the cochlea. The frequencies can't be resolved individually and are perceived blurred.

Another form of describing the spread of excitation in the cochlea is by means of tuning curves. Figure 2.7 shows ten exemplarily curves between 0 and 50 kHz for single neurons of anesthetized cats, which are shaped similar to human tuning curves in the lower region up to 20 kHz. The tuning curves characterize the level of sound needed to achieve a constant excitation at a specific location of the basilar membrane and thereby the frequency selectivity of a nerve fiber at that location. The threshold is lowest for the characteristic frequency of this location and rises with increasing spectral distance. High frequency tuning curves are generally steeper than low frequency tuning curves. The frequency tuning is sharpest for low signal levels. The response region on the basilar membrane widens and the tuning flattens at high levels. This can be seen in Figure 2.7 where tuning curves over the whole audible spectrum are compared. The figure shows neuronal tuning curves of cats, which are similar to human neuronal tuning curves. It is not possible to noninvasively measure the excitation of individual nerve fibers in humans.

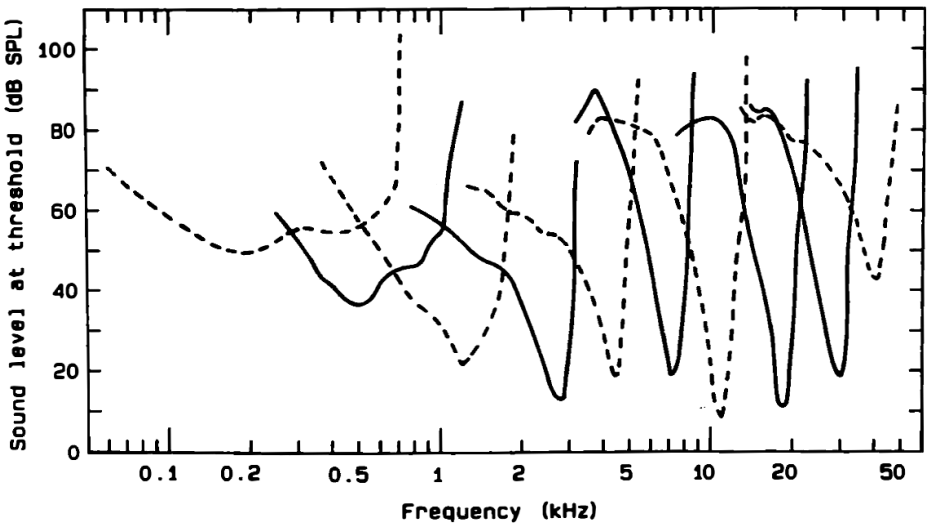


Figure 2.7: Logarithmic representation of cat tuning curves from ten different neurons between 0 and 50 kHz. The tuning curves are displayed alternating as continuous and dashed lines for visual clarity [Moore 2012].

The transfer functions of the auditory filters are characterized by their absolute value function and their phase response, thus the parameters amplitude and phase.

In terms of signal amplitudes the human auditory system is able to perceive a very wide dynamic range. The maximum dynamic range can be observed at mid-frequency tones of 1-4 kHz with approximately 120 dB. That corresponds to an intensity difference of the factor 10^{12}. This very large input range is transformed onto a smaller internal dynamic range, which is physiologically manageable by the central auditory system. This conversion is called cochlear compression. The auditory filters are nonlinearly dependent on the input amplitude, i.e. frequencies within the auditory filter range are amplified nonlinearly. For low levels up to 20 dB and high levels beyond 90 dB they are found to be almost linear whilst in the range in-between they are highly nonlinear. In this level range a change of 50 dB in sound pressure corresponds to a change of about 10 dB in the velocity of the basilar membrane what equals a compression ratio of 5:1. Low frequencies are amplified by 50-80 dB. The nonlinearity occurs mainly at frequencies near the characteristic frequency of the monitored location of the basilar membrane. For frequencies off the auditory filter ("off-frequencies") the transmission is linear, which means that the detection threshold is much higher and the loudness rises faster when the signal level rises (similar to the level perception of hearing impaired

listeners depicted in Figure 2.8a).

For a long time the focus of research was on level effects of cochlear filtering. Recent studies confirmed that each filter also has a nonlinear phase response. For a single auditory filter the change in phase gradient across frequencies seems to be constant, i.e. the phase curvature seems to be uniform. The current state-of-the-art of phase response studies will be described in detail in Chapter 4.

2.3 Consequences of hearing impairment on cochlear filtering

Hearing impairment can have different reasons. The most common cause is a damage of the inner or outer hair cells in the inner ear and is called sensorineural or cochlear hearing impairment. Other forms of hearing impairment are a dysfunction of the auditory nerve or middle ear hearing impairment, the latter produced by a damage of the ear drum, a dislocation of the ossicles or a fixation of the stirrup plate with the oval window.

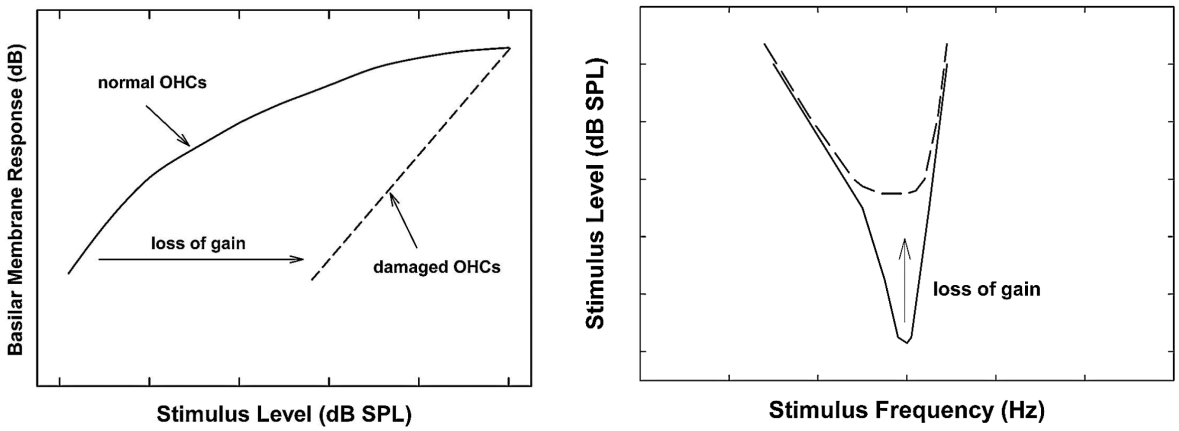
In this work the focus lies on the cochlear hearing impairment[6]. Loosing inner or outer hair cells has different effects on the perception of sound [Moore 1996].

A loss or dysfunction of the inner hair cells reduces the sensitivity of the auditory system. The audiometric thresholds for the frequencies transmitted by these hair cells rise in comparison to normal hearing (NH) listeners (see Figure 2.8a).). Given that the maximum tolerable sound level is about the same as in NH listeners, hearing impaired listeners have a smaller range of levels they can perceive as compared to NH listeners.

Lost or damaged outer hair cells leads to a reduction of the active processes like the basilar membrane's compression and frequency selectivity. In case of a significant damage of these hair cells the input-output function can be less compressive or even linear[7]. A comparison between the level perception of NH and HI listeners is shown in Figure 2.8a. Since the absolute sensitivity is reduced for HI listeners and thereby the hearing threshold is elevated but the maximum tolerable sound level is unchanged, the loudness growth is steeper than for NH listeners ("loudness recruitment").

[6]For simplicity the term hearing impairment (HI) will be used in this thesis. When not stated otherwise, it will always refer to cochlear hearing impairment.

[7]In the study of Ruggero and Rich (1991) on chinchillas the outer hair cells were "deactivated" medically by an intravenous injection of the ototoxic drug quinine, which led to a temporary linearization of the compression.



(a) Cochlear compression (b) Tuning curve

Figure 2.8: Consequences of cochlear hearing impairment on the perception of sound - comparison between NH (continuous line) and HI listeners (dashed line) [Bacon 2006].

The shape of a tuning curve for NH and HI listeners is displayed in Figure 2.8b. Tuning curves are proportionally broader for hearing impaired listeners and have a less sharp notch at the characteristic frequency. Hence the frequency selectivity, i.e. the ability to separate and resolve the components of a sound signal, is poorer for HI listeners. This broadening of the auditory filters results in a greater difficulty to process signals when competing sound sources are present simultaneously.

The auditory system resolves time patterns in each auditory filter and then compares them to each other. This temporal integration of sound signals is also degraded by hearing impairment. HI listeners perform equally well as NH listeners for deterministic signals, but much worse for nondeterministic signals like speech and environmental sounds. Additionally low sensation levels and a restricted audible bandwidth degrade their ability of temporal processing.

Further perceptual changes between normal hearing and hearing impaired listeners affect the perception of pitch[8], the intensity discrimination as well as the sound localization with different effects on binaural and spatial hearing (see Chapter 3.5).

The experiments conducted to measure these effects are extensively described in [Moore 1996].

[8]The pitch may be shifted upwards when the hair cells of low frequencies are missing.

3 Lateralization and interaural time differences

The process of extracting spatial information from sound signals and thereby detecting the positions of sound sources in space is called localization. When the spatial information is presented via binaural signals on headphones, this process is called lateralization. In contrast to signals in space, binaural headphone signals can be either diotic, when the same signal is presented on both ears, or dichotic, when different stimuli are presented on left and right ear [Yost and Hafter 1987].

In this chapter the general functionality and neuronal processing of localization will be presented with a special focus on interaural time delays and their importance for localizing sound sources for normal hearing and hearing impaired listeners [Blauert 1983] .

3.1 Localization

The human auditory system is able to resolve the auditory space very precisely. It resolves differences of an intracranial image with a minimal audible angle of 1° in front of the listener's head and 5-7° off to the sides [Yost and Hafter 1987]. Different mechanisms are responsible for the localization in vertical and horizontal dimension [Grothe et al. 2010].

Vertical localization

Vertical displacement is resolved by pinna and concha of the outer ear in form of a direction-specific attenuation of different frequencies (as explained in Chapter 2.1). Thereby depending on the vertical position of the sound source the transfer function of the outer ear is changing: the central notch shifts towards lower frequencies when the source is presented from below the listener's head and towards higher frequencies when the perceived sound is presented from above. This effect can be seen in Figure 3.1.

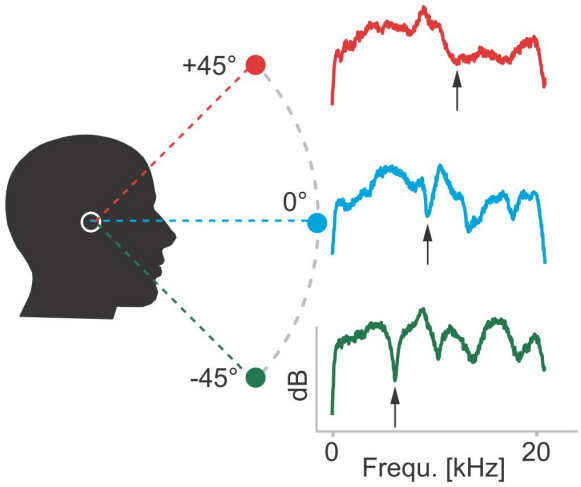


Figure 3.1: Vertical localization: Relation between vertical position of a sound source and the HRTF [Grothe et al. 2010].

The conversion of this spectral information into spatial cues is done by special type IV neurons in the dorsal cochlear nucleus of the central auditory system. These neurons are particularly excited by low intensity signals at their individual characteristic frequency. Their firing rate lowers for a higher intensity or a bigger difference in frequency. The output neurons to which the type IV neurons project are the type O neurons of the interferior colliculus. These neurons have a matching oppositional frequency-versus-intensity response area. Thereby this neurons processes the direction-dependent spectral cues generated by the outer ear.

The elevation of sound sources can be perceived even with monaural signals since it is only dependent on spectral differences [Grothe et al. 2010].

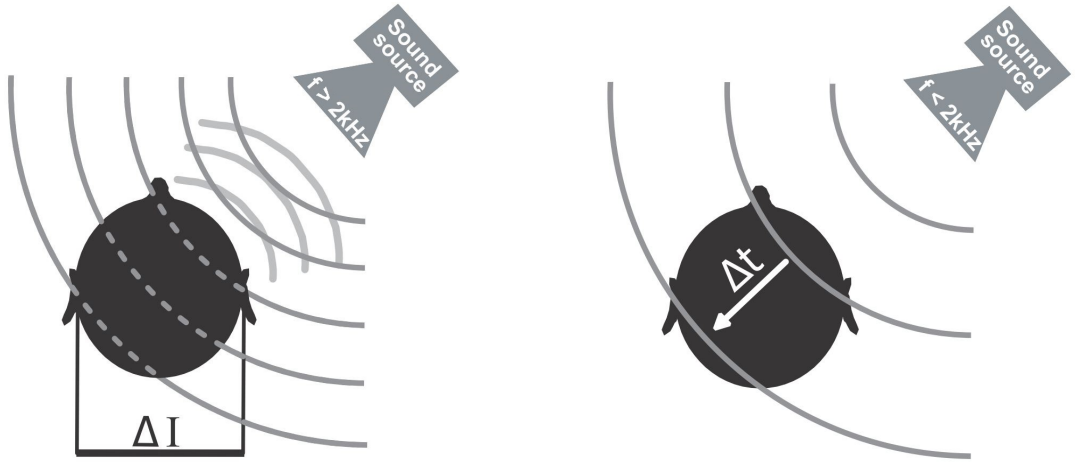
Horizontal localization

Detecting the spatial position of sound sources in the horizontal plane, in contrast, requires signal information from both ears. The cues for horizontal localization are thereby also called binaural cues. They are not resolved by the peripheral auditory system like most other signal information, but solely determined by central neurons in a later stage of the auditory system.

Since the experiment conducted in this thesis will be based on the lateralization of sound sources, the physiology and functionality of binaural hearing will be explained in detail in the following subsection.

3.2 Functionality and physiology of binaural localization

Binaural localization is based on the comparison of the internal signals of left and right ear. The two important cues are the interaural differences in level and time.



(a) Interaural level differences (ILD) (b) Interaural time differences (ITD)

Figure 3.2: Binaural localization cues [Grothe et al. 2010, rearranged by the author].

Interaural level differences (ILDs) arise mainly for high frequencies above 1500 Hz[9]. At these frequencies the head of the listener is large in comparison to the wavelength of the signal and thus causes a shadowing effect at the averted ear (see Figure 3.2a). Thereby emerging level differences can reach up to 30 dB. The lowest perceivable ILD is 1 dB [Yost and Hafter 1987]. ILDs are often also referred to as interaural intensity differences (IIDs).

Interaural time differences (ITDs) can be detected in the fine-structure of low frequency signals up to 1600 Hz and in the envelope of amplitude-modulated signals with high carrier frequencies [Yost and Hafter 1987]. The localization effect arises from the difference in arrival time between the ear facing the sound source and the averted ear (see Figure 3.2b). This time delay depends on the individual head radius and the acoustic velocity. The smallest detectable ITD is 10 µs. The maximal naturally occurring ITD is 600-800 µs, depending on the individual head radius of the listener [Bernstein and Trahiotis 1985]. The use of headphones enables presenting larger ITDs to a listener.

[9] To a low extent ILDs can also be observed in the range of 500 to 1500 Hz.

For frequencies up to 1000 Hz the minimal noticeable change in angular displacement is found to be roughly constant[10]. Thereby the minimal detectable ITD is dependent on frequency due to the signal period. It is high for low frequencies (for example, 30 µs at 250 Hz) and low for higher frequencies (for example 10-15 µs at 1000 Hz) [Zwislocki and Feldman 1956]. For even higher frequencies above 1000 Hz the ITD threshold rises rapidly and leads to a more shallow lateralization [Bernstein and Trahiotis 2002], which may result from frequency-dependent differences in the peripheral processing. ITDs are also described in form of interaural phase differences (IPDs). The human auditory system is able to resolve phase differences of minimal 2° at low frequencies and 5° at 1000 Hz. For a pure tone stimulus, a delay from 0° to 90° degree in phase can be heard as an increasing lateralization towards one side. From 90° to 180° the amount of lateralization becomes smaller and the auditory image becomes more diffuse. At 180° difference the signal is heard on both sides, whilst for delays between 180° and 360° the image appears on the opposite side of the head.

Binaural localization is a highly complex neuronal process. ILDs and ITDs are resolved in different parts of the superior olivary complex (SOC), which is part of the central auditory system (see Figure 3.3). ILDs are detected in the lateral superior olive (LSO) [Grothe et al. 2010]. Over the 8th cranial nerve, the vestibulocochlear nerve, the inner ear is connected to the cochlear nucleus. Spherical bushy cells of the ventral cochlear nucleus respond to the temporal structure of the stimulus, combine the inputs across several nerve fibers and transmit this spike pattern to the nerves of the LSO. Also global bushy cells of the contralateral cochlear nucleus, which combine more auditory nerves and have a better temporal precision, send their information to the LSO. The subtraction of these excitatory inputs from ipsilateral and contralateral cochear nucleus produces the ILD.

[10] The angular displacement is approximately 1.25° for 50 dB SPL at 500 Hz. It is found to be similar for all frequencies from 0 Hz to 1000 Hz [Zwislocki and Feldman 1956].

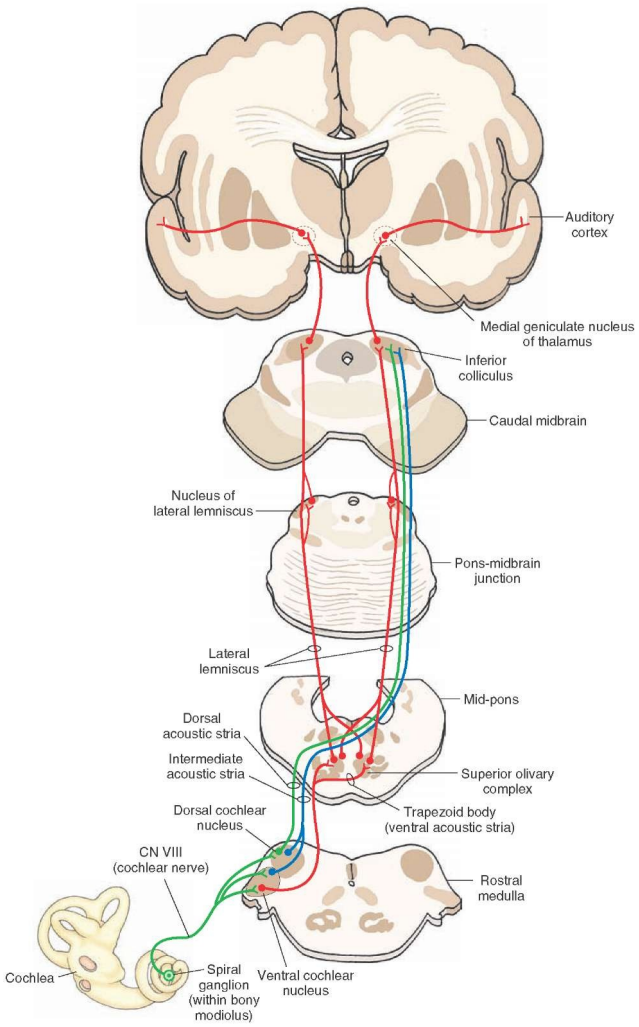


Figure 3.3: Pathways of the central auditory system [Crankshaft Publishing].

ITD detection is a very precise temporal process, since neurons have to resolve differences much shorter than the minimal time interval between action potentials transferring this information [Grothe et al. 2010]. The medial superior olive (MSO) of the SOC is proven to be the main site of ITD processing. To a smaller extent, the LSO also contributes to ITD perception. The principle cells of the MSO receive four segregated inputs: inputs from the spherical bushy cells of the cochlear nucleus and bilateral inhibitory inputs from the medial nucleus of the trapezoid body. The exact interaction of these inputs in order to gain the ITD of the stimulus is not yet scientifically ascertained due to the technical challenge of in vivo recordings of MSO cells. The MSO cells are frequency tuned, which means that they have the highest firing rate at their characteristic frequency. They also have a favorable ITD, at which the phase locking of

the action potentials is maximal for low frequencies. In contrast, the cells of the LSO, also responsible for ITD detection, are frequency independent. They are sensitive to phase inversions of the stimulus[11]. Another effect assumed to contribute to the ITD processing is the cochlear delay arising from the limited speed of the traveling wave on the basilar membrane, which leads to an increasing delay for lower frequencies.

Studies have shown, that the auditory system is able to process ITDs outside the natural range. This could result from the need to process reverberation, which reduces the correlation between the bilateral inputs [Grothe et al. 2010].

3.3 Envelope ITD

As described above, the human auditory system has the ability to resolve ITDs not just through the fine-structure but is also able to extract it from the envelope of a sound signal. Thus also amplitude-modulated signals in a higher frequency range can be resolved, even if these frequencies on their own don't contain ITD information perceivable by humans [Yost and Hafter 1987]. Envelope ITD is based on the perception of the fundamental frequency of the signal complex. It thereby depends on the shape of the waveform over one fundamental period and its properties. Studies systematically manipulated the envelope properties of a stimulus and measured the minimal perceivable ITD differences [Laback et al. 2011, Klein-Henning et al. 2011]. They found four factors that influence the ITD sensitivity: the duration of off-times, the steepness of the envelope slope, the modulation depth and the peak level.

Larger off-times, i.e. silent intervals in each period, enlarge the ITD sensitivity since the affected nerves can recover during these signal portions[12]. A steeper slope of the envelope also improves ITD sensitivity. It decreases the standard deviation of the first spike timing and increases the spike count. These two properties are expected to be the main factors for envelope ITD sensitivity. They depend on the modulation depth of the signal and thereby on the phase relations of the signal components. Phase relations leading to a large modulation depth and therefore to longer off-times and a steeper

[11] Phase ambiguity occures when signals are rotated by one or more full periods and thereby are replica of the original signal.

[12] ITD sensitivity seems to increase for an enlargement of the off-time up to about 12 ms and to stay constant for larger off-times [Laback et al. 2011].

slope will enhance ITD sensitivity. Additionally a rise in peak level improves the ITD sensitivity. This is probably due to the additional recruitment of neuronal fibers with lower spontaneous firing rates and off-frequency neurons at higher peaks. Signals with a "peaky" envelope, having a larger crest factor, like pulse trains can thereby be lateralized more accurately based on ITD than more signals with a shallower envelope.

In general, envelope ITD is a weaker cue than fine-structure ITD. Studies by Bernstein and Trahiotis [2002, 2003] compared the ITD perception of low-frequency stimuli with different types of high-frequency stimuli, including, high-frequency "transposed" stimuli, high-frequency noise, and sinusoidally amplitude-modulated tones (all centered at 4000 Hz). Transposed stimuli have been designed to provide high-frequency channels with similar temporal patterns as low-frequency stimuli[13]. They found greater extents in laterality for the transposed stimuli compared to the other types of high-frequency stimuli. For 125 Hz the laterality of the high-frequency transposed stimuli was almost equivalent to their low-frequency counterpart. The frequency-related differences in the ITD sensitivity for "conventional" stimuli thus appear to result primarily from differences in temporal patterns provided by those stimuli. In subsequent studies [Bernstein and Trahiotis 2009, Bernstein and Trahiotis 2011], they used raised-sine stimuli, in which the peakedness and other waveform parameters can be varied independently. They found that increasing the relative peakedness, the depth of modulation or the rate of modulation (up to 128 Hz) led to a decrease in threshold ITD, i.e. a higher ability to discriminate changes in envelope-ITD. These changes in the peakedness and depth of modulation were later also shown to lead to larger extents of laterality [Bernstein and Trahiotis 2011].

Extents of laterality was measured using a pointing task. Figure 3.4 compares the results of these two studies. The ordinate shows the IID that was required to shift a pointer signal to the intracrainal position of the target. The abscissa shows the threshold ITD. The plot reveals an inverse linear relation between the extent of laterality and the value of the threshold ITD. At least for these high-frequency stimuli, peaky waveforms appear to lead to low ITD thresholds and to larger extents of laterality than flat waveforms.

[13] These transposed stimuli are achieved by the multiplication of a lowpass-filtered half-wave rectified low-frequency noise with a high sinusoidal carrier.

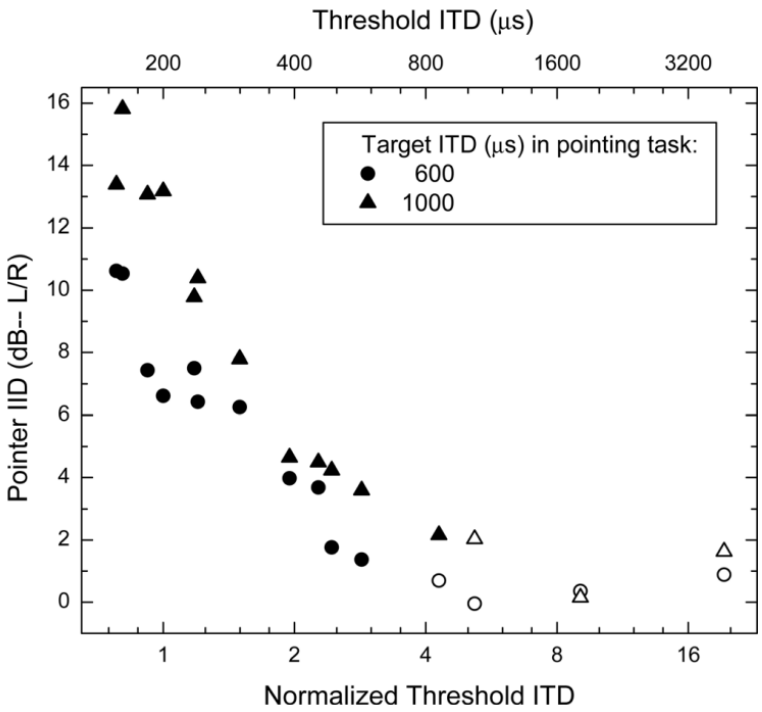


Figure 3.4: Relation between extent of laterality (in form of the IID of a pointer signal adjusted to the target signal) and threshold of ITD for raised-sine stimuli. Open symbols characterize stimuli for which the threshold ITD exceeded the target ITD [Bernstein and Trahiotis 2011].

3.4 Supra-ecological ITDs

The maximal naturally occurring ITD in free-field listening is about 600 to 800 µs. Physiological studies, however, assume that the human hearing system is able to process larger ITDs. Therefore several psychoacoustic studies have been performed to study the perception of these so-called "supra-ecological" ITDs [Mossop and Culling 1998, Noel et al. 2013]. These experiments have been conducted with dichotic signals through headphones. Mossop and Culling measured the just noticeable differences (JNDs) of laterality for two signals. For larger ITDs the laterality of a sound source rises, i.e. the image shifts more towards the side but it also gets more diffuse at the same time. At a delay of 10-30 ms, depending on the stimulus type and experimental set up, the perception of laterality disappears and the image is either heard as separate tones on both sides, spread back towards the midline or it is heard without laterality at all [Mossop and Culling 1998, Noel et al. 2013]. Low frequency sinusoidal signals can be

perceived with a larger ITD than high frequency signals. At high frequencies, which have a short period duration, phase ambiguity occurs already at lower delays. A study by Blodgett et al. [1956] measured a delay threshold of 14 ms for low frequencies and 8 ms for high frequencies. Beyond 15-20 ms the diffuseness of the image begins to rise. For higher delays the diffuseness seems to be the relevant cue for ITD discrimination using a detection task.

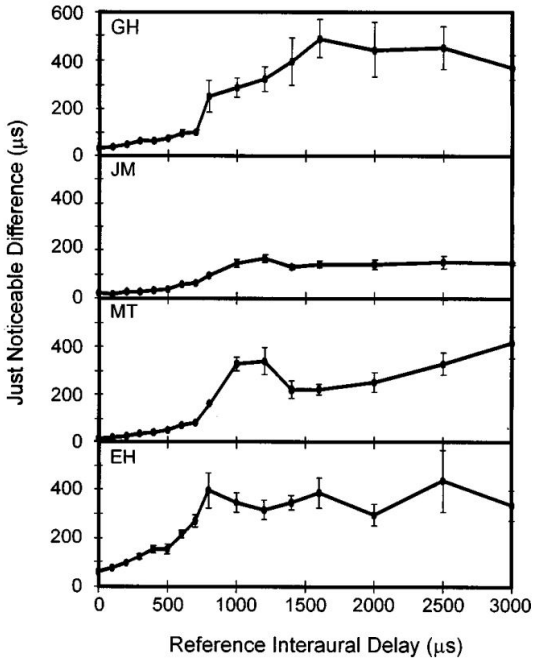


Figure 3.5: JNDs of ITDs as a function of overall ITD for four different subjects [Mossop and Culling 1998].

Figure 3.5 shows the result of an experiment by Mossop and Culling [1998]. Up to an ITD of the reference stimulus of approximately 700 µs the ITD JND rises gradually but is generally low (under 100 µs). Above this value the JND increases sharply for all four subjects. Beyond 1000 µs the JND stays roughly constant but differs among subjects. The abrupt raise of JND over 700 µs could result from the fact, that the subjects were used to smaller ITDs from everyday live, but didn't have enough training to accustom to larger ITDs. An additional experiment using high-pass filtered noise showed that laterality cues are discriminable at much larger ITDs (up to 3000 µs) than are experienced in free-field listening, even in the absence of energy below 3 kHz [Mossop and Culling 1998]. Also physiological studies support this effect [Grothe et al. 2010].

3.5 Binaural perception of hearing impaired listeners

Sensorineural hearing impaired listeners seem to have no reduction in their sensitivity to ILD. Several studies found comparable ILD JNDs for normal hearing and hearing impaired subjects (e.g. Hawkins and Wightman [1980]).

In terms of ITD sensitivity, HI listeners show a significant reduction compared to NH listeners and also a larger inter-individual variability [Moore 1996, Lacher-Fougère and Demany 2005]. Fine-structure ITD sensitivity seems to be more reduced than envelope ITD sensitivity.

Figure 3.6 shows the results of the study by Lacher-Fougère and Demany. The envelope and fine-structure ITD thresholds, described as IPDs in degree, are displayed for signals with different carrier and modulation frequencies. White symbols indicate NH subjects, black symbols HI subjects. In general the thresholds of the HI subjects are poorer than normal for both ITD types but the deficit in fine-structure ITD was significantly stronger than the deficit in envelope ITD[14]. The thresholds for envelope ITDs were lower than the ones for fine-structure ITDs for each of the examined signals. Whilst the ITD sensitivity for narrow-band noise at about 4 kHz was normal in HI listeners, ITD sensitivity for 0.5 to 1 kHz signals was clearly reduced.

[14] With the statistical significance of $P < 0.001$.

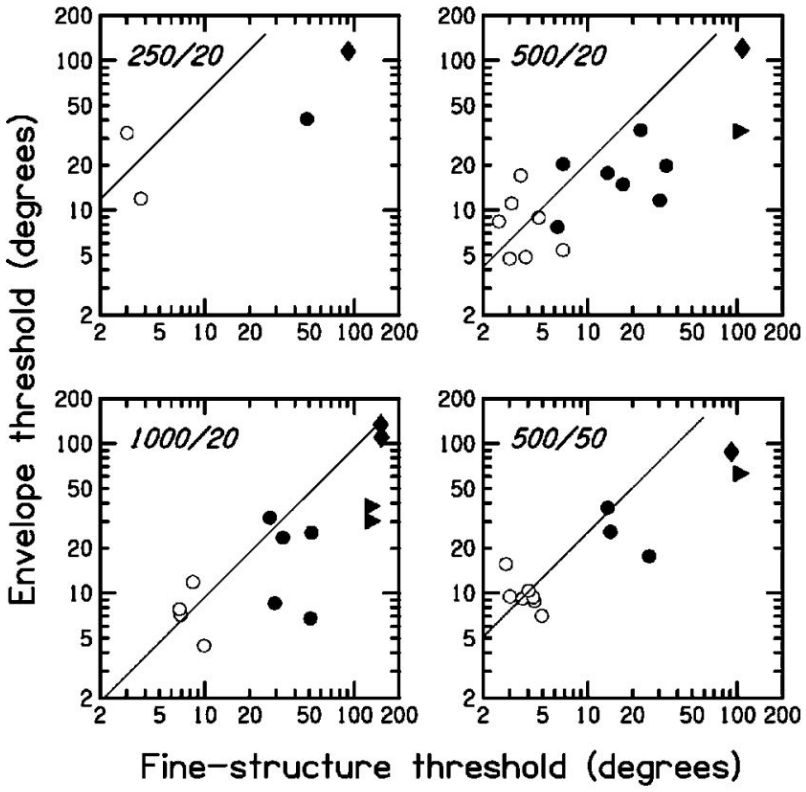


Figure 3.6: Envelope and fine-structure IPD thresholds for different combinations of carrier (250 Hz, 500 Hz, 1000 Hz) and modulation (20 Hz, 50 Hz) frequencies for NH (white) and HI (black) listeners. Diamond and triangle symbols indicate thresholds influenced by ceiling effects [Lacher-Fougère and Demany 2005].

These reported deficit of HI listeners in the sensitivity to fine-structure ITD sensitivity may suggest that those listeners might rely more on envelope ITD than NH listeners. Based on this assumption, the envelope-ITD based paradigm developed in this thesis (see Chapter 5) may be particularly useful for application with HI listeners.

4 Studies on phase response

For a long time researchers assumed that the auditory system is filtering the amplitude of sound signals but not the phase, since most perceptional properties could be explained by the power spectrum of the sound. Whilst much effort has been made to characterize the magnitude response, the phase response hasn't been studied at all before the early 1950s. The ear was assumed to be "phase deaf".

This thesis was falsified in parallel by several studies, i.e. by Mathes and Miller [1943] and Goldstein [1967], who showed that the ear is sensitive to changes in phase within the range of an auditory filter. In 1987 Patterson showed, that to a lesser extent it is also sensitive to phase changes across filters. These results, however, apply merely for relative phase delays: absolute phase and group delays without a fixed time reference have no influence on sound perception [Kohlrausch and Sander 1995].

Basilar membrane models up to the 1980ies (e.g. of Strube in 1985) had not considered the phase response sufficiently and therefore could only be applied for psychoacoustic data where the phase response is not important[15]. In 1995 Patterson developed a first gammatone filter model, called the Auditory Image Model, that attempted to simulate the auditory phase response. This time-domain filter models an impulse response, attempting to reflect the pattern of the auditory nerve firing. However, experiments by Kohlrausch and Sander [1995] and Carlyon and Datta [1997b] showed, that masking data wasn't simulated well by this gammatone model. In 1997 Patterson modified his model with a more realistic gammachirp filter, which includes an onset "chirp" and slope asymmetries.

In 2001 a study by Shera showed that guinea pigs have a phase response with constant curvature. In 2005 this property was also confirmed for the human hearing system, particularly for the main pass-band of the filter [Oxenham and Ewert 2005]. Previous studies with dead cochleae couldn't verify this effect, since the cochleae lost their compression and thereby had a linear phase response.

In the masking experiments of Kohlrausch and Sander and many other phase response studies so called Schroeder phase harmonic complexes have been used. The following

[15] These models have broader filter characteristics than the human cochlea and simulate only passive basilar membrane properties.

subchapter describes the results of these experiments and explains why these signals are well suitable for measuring the phase response.

4.1 Schroeder phase harmonic complexes

Many psychoacoustic studies of the phase response are based on Schroeder phase harmonic complexes (SPHCs). These signal types have been developed by Manfred Schroeder in 1970 and have since been used as maskers for the examination of auditory filter phase responses, for example by Smith et al. [1986] and Kohlrausch and Sander [1995].

Signal properties

Basically a SPHC is a harmonic tone complex with equal-amplitude components and a constant phase curvature [Schroeder 1970]. The phase curvature is the second derivative of the phase and the slope of the group delay: a constant curvature is thereby equivalent to a linear change of the group delay and a monotonous change of the phases across the signal components. The signal consists of N sinusoidal components n in the frequency range of N_1 to N_2 with equal fundamental frequency f_0 that have equal amplitudes A but different phase delays θ_n:

$$m = \frac{1}{N} \cdot \sum_{n=N_1}^{N_2} A \cdot sin(2\pi n f_0 t + \theta_n)$$

The phase delay θ_n of each individual component n is calculated by:

$$\theta_n = C\pi n(n+1)/N$$

Through conversion of the previous equation the curvature of the signal can be computed by:

$$\frac{d^2\theta}{df^2} = C\frac{2\pi}{Nf_0^2}$$

The multiplicative constant C was added by Lentz and Leek [2001]. In the original concept Schroeder used only three different signals: a signal with positive curvature corresponding to $C = 1$, a signal with negative curvature $C = -1$ and a signal with no

curvature $C = 0$. They are called m_+, m_- and m_0. The introduction of the parameter C offers a finer modification of the phase curvature and thereby a suitable tool to approximate the human phase response.

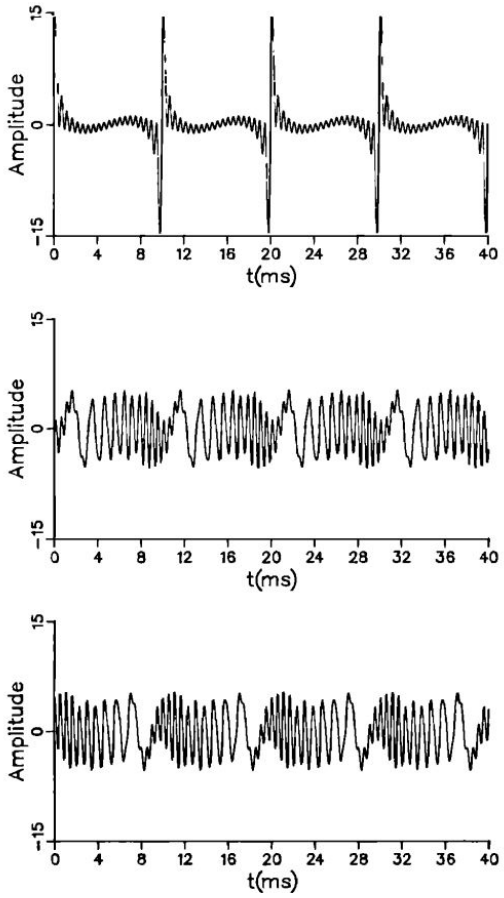


Figure 4.1: SPHCs with zero (m_0), negative (m_-) and positive (m_+) phase curvature (from top to bottom) [Kohlrausch and Sander 1995].

The waveforms of the three original SPHCs are depicted in Figure 4.1. Positive and negative SPHCs have a very flat waveform whilst the signal with curvature $C = 0$ and thereby constant phase of the components has a maximally peaky waveform. The extreme case of m_0, generated by an infinite number of frequency components with infinitely small fundamental frequency, would be a dirac impulse. The envelope of the time-domain SPHC is dependent on the relative phases of the components - the absolute phase value can be ignored since the envelope doesn't change for 2π-phase-shifts.

Positive and negative SPHCs consist of repeating linear frequency glides [Oxenham and Ewert 2005] as can be seen in Figure 4.2. These glides occur over the period T of the

fundamental frequency. The negative curvature leads to a linear rising, the positive to a linear falling sweep in frequency. Positive and negative SPHCs with the same absolute value of C are just temporally reversed and thus have identical long-term power spectra. All three signals of Figure 4.1 have identical long-term amplitude spectra.

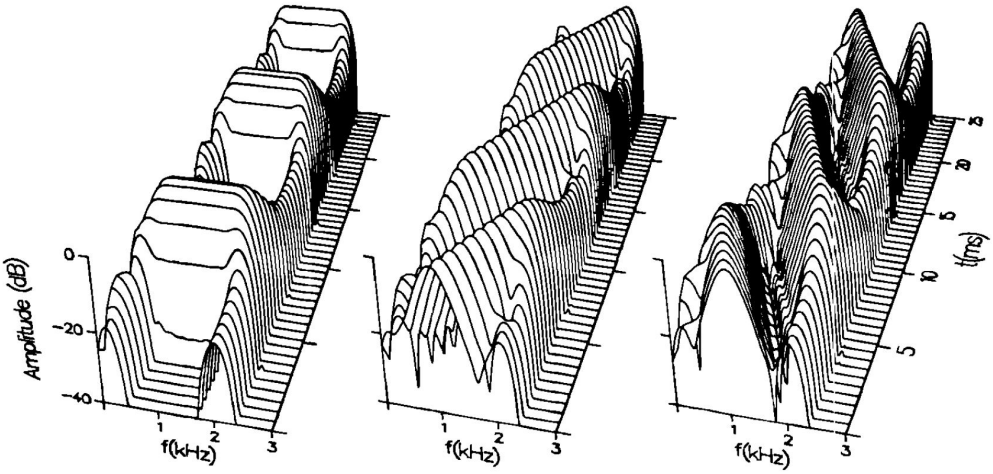


Figure 4.2: Frequency glides of m_0, m_- and m_+ within one fundamental period of the signals [Kohlrausch and Sander 1995].

Studies with Schroeder phase harmonic complexes

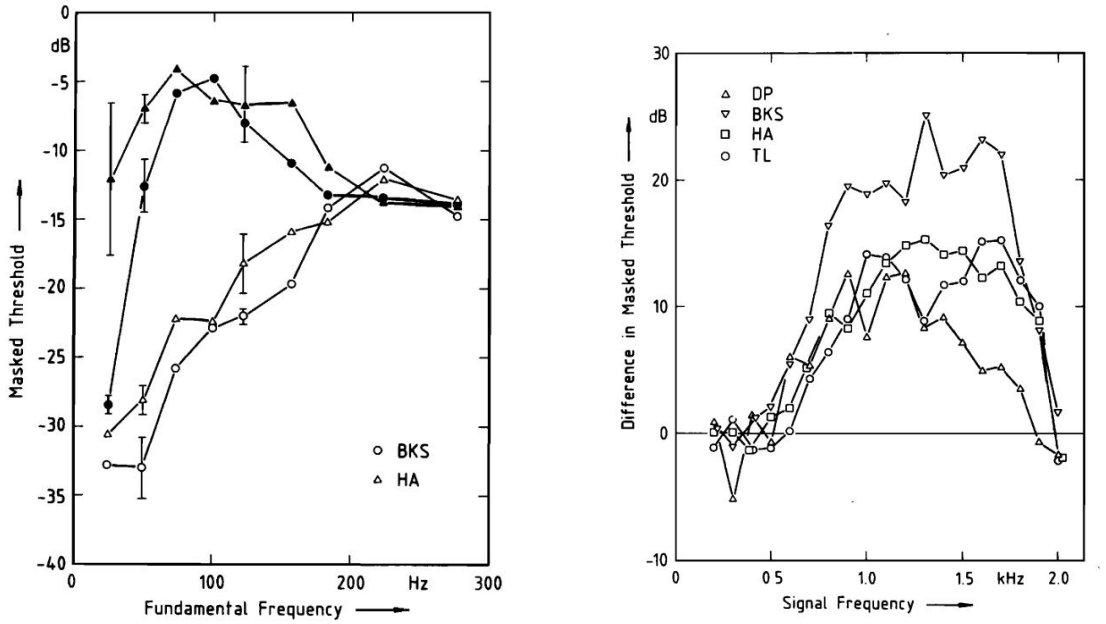
Positive and negative Schroeder phase harmonic complexes have been used as masking signals in experiments by Smith et al. [1986][16] and Kohlrausch and Sander [1995][17]. These studies showed, that the positive SPHC produces significantly less masking than the negative SPHC (see Figure 4.3a). The effect is strongest for fundamental frequencies between 50 and 200 Hz. At these frequencies, differences in masking reach up to 25 dB. This effect is called the masker-phase effect [Smith et al. 1986]. It vanishes for higher modulation frequencies due to the limited temporal resolution of the auditory system and also for lower frequencies, at which the masker signal is perceived as a slowly gliding pure tone. Kohlrausch and Sander [1995] additionally examined the sine-phase signal m_0, which produced a masked threshold that lied in between the thresholds of the other

[16]The experiment by Smith et al. used a three interval forces-choice method in which the subjects had to detect the interval, which included an on-frequency target signal. The level of the signal was reduced with a two-up/one-down adaptive procedure in 1 dB steps until the threshold was reached.

[17]The experiment by Kohlrausch and Sander was conducted similarly to the study by Smith et al. but included additional variations of parameters and the calculation of masking period patterns.

two signals for most conditions. Just for very low fundamental frequencies (under 50 Hz) the masked threshold of m_0 was the lowest.

The masker-phase effect depends on several stimulus properties. One of them is the spectral position of the target that shall be detected. Its consequences are shown in Figure 4.3b. When the target is centered in the spectral range of the SPHC masker, the strongest effect is observable. For target signals towards the spectral edges of the masker signal the masked thresholds for m_+ increases and the difference between m_+ and m_-disappears.



(a) Masked thresholds for m_+ (open symbols) and m_- (solid symbols) for different fundamental frequencies of the SPHC masker for two subjects.

(b) Differences in masking between m_+ and $m_-$for different positions of the target within the spectral range of the SPHC masker for four different subjects.

Figure 4.3: Masker-phase effect in a study by Smith et al. [1986].

The effect also depends on the temporal position of the target within the masking signal. The lowest threshold for the m_+ signal is achieved when the target is presented with a temporal delay at which its frequency coincides with the instantaneous frequency of the masker, which is gliding over its frequency range within one fundamental period [Kohlrausch and Sander 1995, Carlyon and Datta 1997b]. This information can be obtained by conducting the masking experiments for different phase delays between masker and target and display the received data in so-called masking period patterns.

These patterns show a large modulation for the positive SPHC. The masking period pattern of the negative SPHC is flatter. The thresholds for this signal seem to depend less on the temporal position than the threshold for the positive SPHC [Kohlrausch and Sander 1995].

The masker-phase effect is largest for signals at medium hearing level (about 70 dB). Carlyon and Datta [1997b] found the effect to be smaller for lower levels. Summers and Leek [1998] showed that the effect is also reduced for high stimulus levels. They assumed the masker-phase effect to be dependent on the nonlinear input-output function of the cochlea. This function is more linear for very low and high levels [Summers 2000].

Kohlrausch and Sander [1995] suggested that the masker-phase effect arises from the interactions between the phase relations of the auditory filter impulse response and the masker signal. Similar to SPHCs the human hearing system is assumed to have a constant phase curvature. The hypothesis for current experiments on phase response is, that there is one particular phase curvature, which best mirrors the auditory phase response. The phase relations of this signal and the cochlea compensate each other, so that the components of the resulting "internal" signal are approximately in-phase and therefore the signal is maximally peaky with a waveform similar to m_0 in Figure 4.1. Peaky maskers with modulated temporal envelopes cause less masking than signals with unmodulated temporal envelopes ("flat" maskers) due to the ability of humans to "listen in the valleys". They can detect the targets in the low-level parts of the waveform, and thereby temporally enhance the target-to-masker level ratio [Leek and Summers 1993, Oxenham and Dau 2004]. Different studies have shown that cochlear compression is also important for the masker-phase effect to occur [Oxenham and Dau 2004, Carlyon and Datta 1997a].

Since the positive SPHC caused less masking, it is assumed that the human auditory system has a phase response with negative phase curvature. Its phase values mirror the curvature of the positive SPHC better than the negative SPHC and thereby produce a peaky envelope of the internal signal for the m_+ signal.

In order to refine the change in phase gradients, the parameter C was introduced in the calculation of the phase curvature. For signals with $-1 < C < 1$, the change in phase is smaller than for the original SPHCs m_+ and m_- and the frequency sweep is faster. For stimuli outside this range the curvature is faster and the frequency sweep longer.

However, studies mostly concentrated on C-values from -1 to 1.

Lentz and Leek [2000] used signals with C-values from -1 to 2. Their study showed that the SPHC masker had its masking minimum at C-values that depended on its center frequency. The C-values of the masking minimum changed from about $1.3-2$ for 1 kHz to about $0.5-0.8$ for 4 kHz, with a more distinct maximum at large frequencies (see Figure 4.4)[18]. Also Kohlrausch and Sander [1995] concluded, that the phase curvature decreases with the factor two for an increase of the signal frequency of approximately half an octave. Thus, the phase response seems to be frequency dependent.

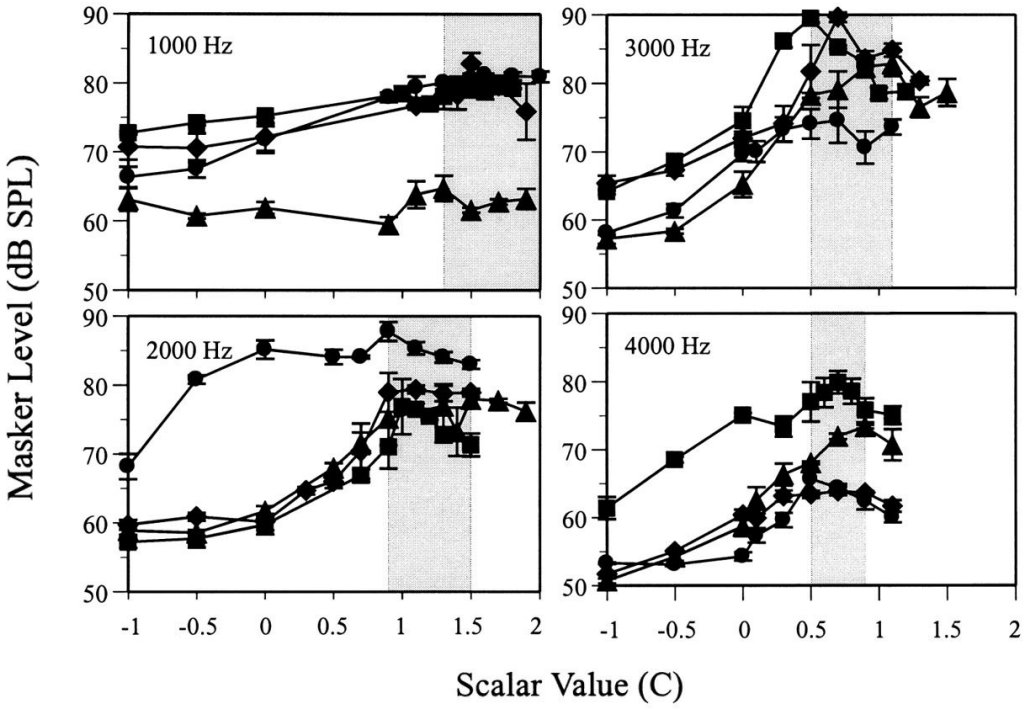


Figure 4.4: Threshold masker level for SPHC masking signals with different C-values (different symbols mark data of different subjects). A maximum of the masker level corresponds to the best detection of the target. The position of the maximum depends on frequency [Lentz and Leek 2001, rearranged by the author].

[18]A detailed examination of the phase response at the auditory filter around 2000 Hz has been conducted in Rasumov [2009]

4.2 Phase response of hearing impaired listeners

The phase response of HI listeners is more difficult to measure than for NH listeners. HI listeners are assumed to have a high inter-subject variability, since the hearing impairment is caused by damage or loss of the inner and outer hair cells. These damages are individually different in extent and in affected spectral range. As described in Chapter 2.3 the damage of hair cells leads to a reduced overall sensitivity and a reduction of the compression of the input-output function [Oxenham and Dau 2004].

Since the hearing impairment affects each person in a different way and therefore each HI listener is assumed to have an specific phase curvature that differs strongly from the phase curvatures of NH listeners, it is not so straight-forward to measure their individual phase responses.

There are rarely any studies about the phase responses of HI listeners. Carlyon and Datta [1997b] conducted a study using SPHC maskers for NH subjects, in which they found the masking-phase effect to be smaller for low level signals. They draw the conclusion that the amount of differences in masking by SPHC maskers is dependent on the shape of the auditory filters. For higher levels the auditory filters broaden and lower the masked thresholds. According to this hypothesis the masker-phase effect would be even higher for HI subjects, who have broader auditory filters over the whole spectrum. Summers and Leek [1998] therefore conducted the same experiment with NH and HI subjects. Their results, however, rejected this hypothesis: The results showed less effect for the HI subjects than for the compared group of NH subjects. When m_- was used as masking signal, the thresholds were very similar among HI and NH subjects for different temporal positions of the target. For m_+, however, the masking period patterns of HI subjects were less modulated than for NH subjects[19]. The masked thresholds were changing less for different temporal positions of the target and resembled the thresholds for the positive SPHC. Summer and Leek thereby assumed the masker-phase effect to be dependent on the nonlinearity of the input-output function of the cochlea. This function is more linear in HI listeners and for very low and high levels in NH listeners, and thus evokes a smaller effect in these cases.

The phase response of NH listeners is assumed to have a constant curvature. The phase

[19] Similar results were also obtained by [Summers 2000] and [Leek and Summers 1993].

response of HI listeners is, however, potentially nonuniform, given the complex pattern of frequency-dependent loss of outer hair cells that are responsible for cochlear compression. Therefore masking experiments with SPHCs are a good way to estimate the phase response of NH listeners but potentially not suitable for HI listeners. Therefore it is necessary to develop a method that takes the individual phase responses of each HI listener into account. This method shall find the phase relations of the components of a harmonic complex that best mirror the phase response of the individual subject.

Another challenge for determining the phase response of HI listeners is the fact that the hearing impairment might be different on the left and right ear. Thereby the method may only be applicable for a bilaterally symmetric hearing loss.

5 Experiment planning

As mentioned in the previous chapter, determining the phase response of hearing impaired listeners is more difficult than for normal hearing listeners. HI listeners have individual differences in the phase response due to their individually different loss of hair cells. Thereby the commonly used Schroeder phase harmonic complexes with constant phase curvatures can't be applied to determine the phase relations.

5.1 Approach

This thesis develops an experimental setup suitable to determine the individual phase response of normal hearing or hearing impaired listeners[20].

The basic idea is to use a multi-harmonic complex similar to the SPHCs, but with separately alterable phase values for all components. Changes in the phase relations result directly in changes in the envelope of the signal waveform and thereby to a more peaky or flat envelope. A signal that evokes a peaky signal waveform[21] after passing the auditory system will evoke a more lateralized auditory image than a signal evoking a shallow waveform (see Chapter 3.3). The center of the distribution of subject's responses is expected to be more lateral for peakier waveforms. Thereby the signal lateralized most is the one which evokes the highest peakedness in the internal envelope and therefore the one which best mirrors the individual auditory phase response of this particular subject. Since envelope ITD seems to be an important cue for distinguishing signals with different phase relations, it seems to be a reliable cue for this purpose.

In this thesis the experiment will merely be conducted with normal hearing subjects due to time limitations. When the paradigm proves to work well, follow-up experiments will be made with hearing impaired listeners.

[20] For the planning of the experiment, methods from [Bech and Zacharov 2006] were used.

[21] A peaky waveform corresponds to long off-times, steep slopes, a high modulation depth and a high crest factor [Laback et al. 2011].

5.2 Expectations for the experiment

We have three main expectations for the experiment with NH subjects:

1. Since normal hearing listeners are assumed to have a roughly constant phase curvature, we expect that they will choose phase values that approximately form a constant phase curvature. The phase values are assumed to converge to a SPHC.

2. The C-values of the SPHCs that agree most with the chosen phase values of each subject are assumed to lie in the positive range between 0 and 1, since previous studies found a maxima at this range of C-values for high frequencies, which are used in this experiment for testing envelope ITD perception.

3. The individual normal hearing subjects are expected to have similar phase responses and thus to choose similar phase values.

5.3 Method

As a basis for the stimuli, Schroeder phase harmonic complexes[22] were used. They were compiled by

$$m = \frac{1}{N} \cdot \sum_{n=N_1}^{N_2} A \cdot cos(2\pi n f_0 t + \theta_n)$$

as explained in Chapter 4.1. Studies so far have used only harmonic complexes with constant curvatures for studying the human phase response [Smith et al. 1986, Kohlrausch and Sander 1995, Carlyon and Datta 1997, Summers and Leek 1998, Oxenham and Dau 2004] but it is not entirely proven that this is the best approximation. Thereby we wanted the subjects to generate their individual optimal stimulus with best fitting phase for each component.

The method of this experiment was to successively build up the harmonic complex by adding components to a simple starting signal and vary the starting phase of the newly added component in each experimental block whilst the phases of all other components stay constant.

[22] In this study we use cosine signals instead of the sinus signals used by Schroeder. Since the phase will be varied over the whole period of the signal this will have no impact on the results.

Therefore in stage one of the experiment we started the experiment with a signal with just three components and varied one component's phase. Three is the lowest number of components for which a change in the phase relations will relate to a change in the envelope. For two components a phase change would just shift the envelope in time. Every change in phase relations of the components leads to a change of the envelope of the signal. We expected to have one phase value where the internal signal gets maximally lateralized and maximally peaky. The phase value that evoked the maximal lateralization was set for this component and another signal component was added.

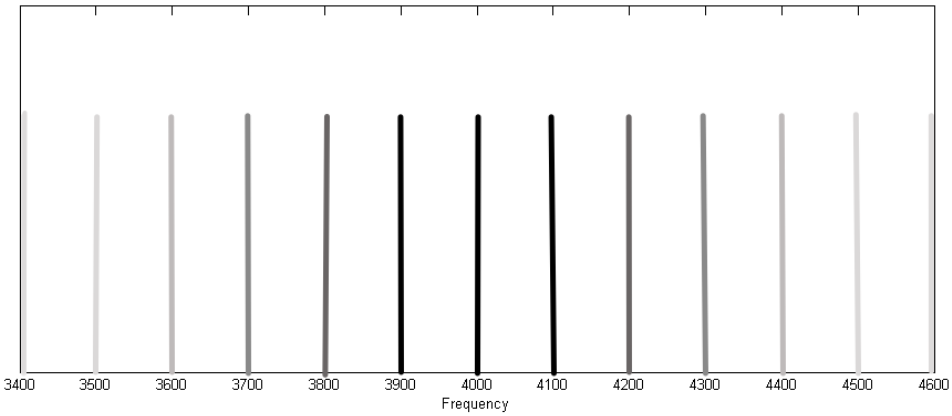


Figure 5.1: Build up process of the signal: the experiment starts with a three-components complex - the center frequency and its two adjacent components. In each further block one component is added, extending the harmonic complex alternately towards lower and higher component frequencies.

This procedure was repeated until the full complex was generated (see Figure 5.1). By alternating between low and high-frequency components, the complex was always maximally centered around the center frequency so that the signal was not spectrally shifted away from the auditory filter centered on the full complex and so that the fundamental frequency could be heard clearly due to the small spacing between components.

After adding all components, the two middle components that have not been varied in phase so far will be adjusted in the same manner within the whole complex. The order of the buildup are the same for all subjects.

This procedure focuses on the comparison of phase variations for each component. To consider interactions that may arise between components of the complex finally obtained in this experimental stage, in a second stage of the experiment the final complex of the

first stage was presented with variable phase of one random chosen component at a time. If interactions between components do not play a role and listeners are perfectly reproducible in their responses, the selected phases should be the same as for the first stage. The final stimulus is assumed to reveal the phase values that evoke the maximal peaky internal signal for this subject.

5.4 Stimuli

5.4.1 Parameters of the Stimuli

Center frequency

The stimuli were centered around 4000 Hz[23]. In this frequency region humans are very sensitive to envelope ITD [Majdak and Laback 2009, Bernstein and Trahiotis 2002]. This center frequency was also used in several previous studies including previous BiPhase experiments and therefore further information can be taken into account and the results can easily be compared.

Fundamental frequency

The fundamental frequency of the stimuli was 100 Hz. At this frequency the human envelope ITD sensitivity is strong [Bernstein and Trahiotis 2002]. The frequency region around 100 Hz implies a trade-off between two oppositional effects: for lower frequencies the envelope ITDs fall due to the deep modulation rates, for frequencies above 150-200 Hz it is reduced due to the binaural adaption effect [Bernstein and Trahiotis 2002].

Frequency range

The stimuli consisted of 13 components in the range of 3400 to 4600 Hz, which corresponds to $0.85 \cdot f_c$ to $1.15 \cdot f_c$ with the center frequency $f_c = 4000$ Hz.

All components fell within the range of one auditory filter centered at the complex, which covers approximately $0.7 \cdot f_c$ to $1.3 \cdot f_c$ [Glasberg and Moore 1990]. The experiment is thereby not be influenced by off-frequency effects on the phase curvature.

[23] For this auditory filter the equivalent rectangular bandwidth is 456 Hz.

Phase relations

The phases were varied within the range of 0 to 2π. Since we added onset and offset ramps to the signal the absolute phase of the signal was considered as negligible and just the relative phase difference between the components was considered important. The amount of phase variations had to be large enough to allow an adequate sampling of the changes in peakedness, so that no potential maxima were missed ("inter-sampling-peaks"). But they were also limited due to the experimental duration, which should not exceed two hours per session. Each phase alteration should be presented with 10 repetitions for each of the 13 blocks in the experiment.

In a preliminary examination of the stimuli, the changes in crest factor for varied phase relations were computed. Since the crest factor is defined as the quotient of maximal value to effective value of the signal amplitude, it is a quick and easy estimate of the peakedness of the signal. Additional factors considered to be important for envelope ITD perception, like the steepness of the rising slope or the duration of the off time in each modulation cycle, are not considered by this metric.

Figure 5.2 shows the crest factors for a SPHC with $C = 0.5$ and one component varied in phase across one signal period. The phase values of this SPHC have been used as default phase values for the first stage of the experiment since for a harmonic complex centered at 4000 Hz NH subjects are assumed to choose phase values that resemble a constant phase curvature in the range of $C = 0.5$ to $C = 0.9$ (see Chapter 4.1, [Lentz and Leek 2001]). In Figure 5.2 three different resolutions were used for sampling the phases over one period: 10, 15 or 20 samples. It can be seen that the crest factor is changing smoothly across the phase cycle. Therefore we decided for using just 10 phase values. This resolution appeared to be sufficient to detect any distinct maxima in peakedness.

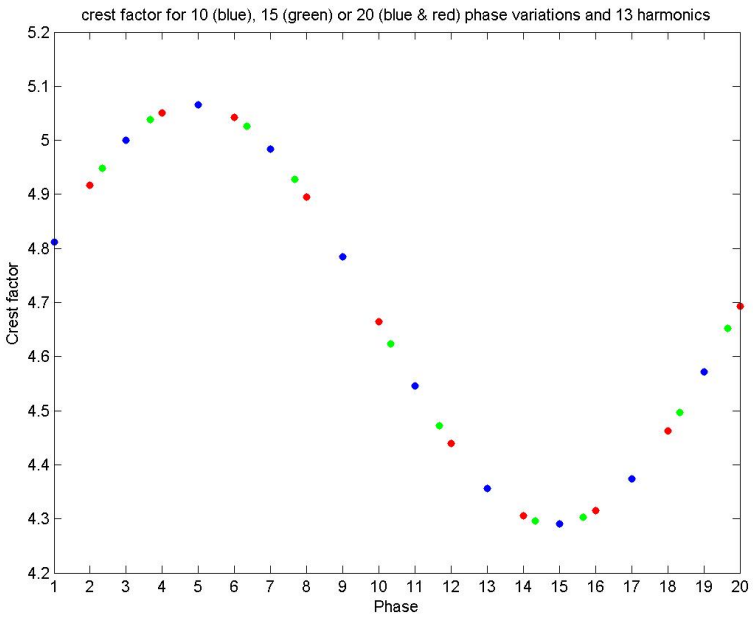


Figure 5.2: Crest factors of a 13 component complex with one phase value changed across the phase cycle in different resolutions: 10 (blue), 15 (green) or 20 (blue and red) samples.

ITD

A ITD of 700 µs was used, which shifted the perceived auditory image of the signal towards the right side[24].

The leading side was not changed during the experiment so that uncontrollable effects or unsymmetrical behavior due to such a change were avoided. As described in Chapter 3.4 700 µs corresponds roughly to the maximal ITD humans experience in their natural environment and are thereby the largest ITD values that humans are used to process[25].

In the second stage the ITD was increased to 2500 µs. This step will be motivated in Chapter 6.1.

[24] We don't use any ILD between left and right channel. Since ILD cues are important for the frequency region around 4000 Hz, the position auditory image will be influenced by both cues - the ILD of an unlateralized signal on the midline and the ITD of a maximal lateralized signal. Due to this time-intensity trading the auditory images of the stimuli will be situated between these two extreme values.

[25] Experimental results from the BiPhase project also showed that subjects achieved a high probability of correct responses for an left-right discrimination test with SPHC stimuli at this ITD value.

Ramp

Raised cosine on- and offset-ramps of 125 ms were added to the stimulus in order to avoid onset and offset effects (so called "spectral splatter") and to avoid uncontrolled effects of the absolute phase of the signal envelope.

Duration

The stimulus duration was 300 ms. Up to a signal duration of about 250-300 ms the ITD sensitivity is increasing, while there is no significant improvement for larger durations [Buell and Hafter 1988].

Level

The level of the full 13 component stimulus was 70 dB SPL. This level was also used in other BiPhase experiments and corresponds to one of the levels used in the studies by Shen and Lentz (2009) and Oxenham and Dau (2001) that investigated the level dependency of the phase response (see Chapter 4). Because the phase perception of a signal depends on the level, changing the levels of the individual components in order to maintain the same overall level over the successive buildup of the stimulus would lead to incorrect phase relations between the components. Therefore the level of each frequency component was kept constant at 58,8 dB SPL. That led to a rise of the overall level from 64 dB SPL for three components to 70 dB SPL for the full complex.

5.4.2 Background noise

An interaurally uncorrelated diffuse background noise was added in order to mask any audible distortion products of the cochlea which could potentially serve as lateralization cues. Therefore independent white Gaussian noises were filtered with a second order low-pass filter (with a cut-off frequency of 1300 Hz) and presented at a spectrum level of 25 dB/Hz at the two ears. The level of the noise was around 15 dB lower than the stimulus level, which is sufficient to mask primary combination products [Greenwood 1971].

5.5 Subjects

Eight normal hearing subjects participated in the experiment (5 female and 3 male). Their ages ranged from 23 to 35 years. An audiogram was made from each subject in advance to ensure that they had normal audiometric thresholds and no hearing damage greater than 20 dB hearing level in the relevant range of 100 to 12000 Hz[26].

Three of the subjects (NH39, NH43, NH144) had participated in other lateralization experiments prior to this experiment. Thereby they might have been better trained on ITD cues and be more familiar with the experimental environment. These subjects were known to have good ITD sensitivity and were chosen for this reason. The other subjects did not participate on previous ITD experiments and thus have not been tested on their ITD sensitivity.

5.6 Procedure

In total, the experiment took 4 to 5 hours for each subject. It was divided into two sessions on separate days, each containing one experimental stage.

The method was intended to work without training the subjects in advance. The experiment should be a relatively fast estimation of the phase response, which should be accomplishable without specific training on ITD-based lateralization.

The only training the subjects received was a short presentation of the signals used in the experiment in order to get familiarized with the differences between the stimuli as well as a practice block in form of the first block of the actual experiment. The subjects have thereby become acquainted with the stimuli and the experimental procedure and were allowed to ask questions before the start of the actual experiment.

First stage

Each block of this stage consisted of two consecutive tasks. The first one was a magnitude estimation task. The goal was to determine the phase relation that evokes the highest lateralization towards the leading (i.e. the right) side. Subjects were asked to

[26]The audiogram was measured with a three-alternative-forced-choice method with a 3down/1up staircase procedure. The audiogram for each subject can be seen in the appendix.

indicate the position of the perceived auditory image on a horizontal scale ranging from the left side of the head (value: -15) to the right side (value: +15). The graphical user interface is shown in Figure 5.3. The scale was divided into 31 response buttons. This resolution was considered to be sufficient considering the variability in lateralization responses. Although the signals were generally lateralized toward the right side due to the right-leading ITD, the full scale ranging from left to right ear was used. This should provide the subjects with a more balanced orientation than just a middle-to-right-ear scale and also account for different response behaviors of the subjects. While some subjects may use only a small part of the scale, others may require more space toward the sides. Presenting stimuli always at the same side may potentially lead to an increasing shift in perception towards the midline by means of adaptation to the presented ITD. As this adaption is expected to occur with a relatively long time constant, it may not affect the results because we were only interested in finding the maximal lateralization value within a block of relatively short duration. By choosing this large scale, the subjects were not restricted to the right side. They were allowed to exceed the midline of the response scale in order to avoid potential compression of their responses.

Figure 5.3: Experiment GUI: Magnitude estimation (Instruction to the subjects: “Press on the position from where you heard the sound”).

The subjects gave their responses on a touchscreen. The touchscreen was preferred to other input devices since the subjects could directly reach towards the position of the stimulus instead of translating the position into the movement of a computer mouse or the buttons on a keyboard or a gamepad. This should provide a more accurate response pattern.

Each individual block consisted of 10 repetitions of each of the 10 phase variations of one component. The stimuli within one block were presented in random order. Thereafter the mean lateralization values for each phase relation were computed and the three maximally lateralized phase values (towards the right side) were selected for the second part of the block. When there were two phases with the lateralization same value, we chose one of them randomly by taking the one played first in the shuffled item list.

The second part of each block consisted of a direct comparison of these three most lateralized candidate stimuli, corresponding to a ranking task. The three stimuli were presented in random order and the subject had to choose which one he or she perceived farthest to the right side (see Figure 5.4).

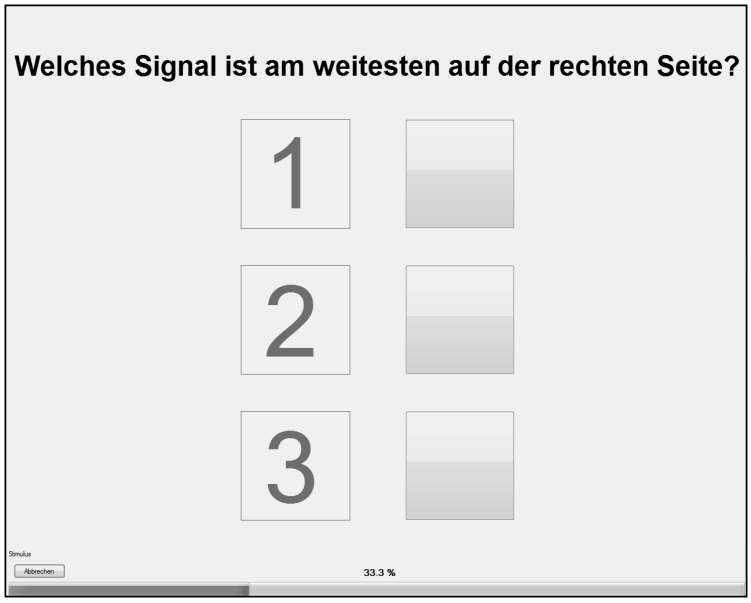


Figure 5.4: Experiment GUI: Ranking (Instruction to the subjects: “Which signal is farthest on the right side?”) .

Each of the six different sequences was presented five times. A ranking was made to determine the best stimulus and the significance of this decision. When there were two stimuli with the same number of selections, we decided to choose the one, which had the larger lateralization value in the first part of the block.

The touchscreen input device was also used for this task to avoid confusion and distraction from the actual experiment from multiple changes between touchscreen and buttons. The three response buttons were arranged vertically on the graphic interface to avoid interactions between auditory spatial (left/right) perception of the three stimuli with visual spatial orientation of the corresponding response buttons. For this purpose a touchscreen was also more suitable than a keyboard or gamepad.

We chose this two-part system to improve the decision making. In the first part we are able to distribute the phase values over the whole phase period whilst in the second stage we can achieve a more reliable final result by the direct comparison of the best performing stimuli, which is an easier task for most subjects and allows them to discriminate smaller differences [Hartmann and Rakerd 1989].

The phase value chosen in this second task was set for this component for all following blocks. Successively the signal component was built up from 3 to 13 components. In the end, as mentioned before, the two middle components, at the beginning set to the phase values of a SPHC with $C = 0.5$, were varied.

Second stage

The second stage of the experiment used the phase values determined in the first stage and allowed to readjust them in the context of the full complex. Since the phases in the first stage have been ascertained with different numbers of components these phase values may differ for the full complex because of interactions between the components. In this stage we always presented the full complex with the phase values of stage 1 and varied one component in phase. The 130 different stimuli (10 phase conditions for 13 components) were presented with 10 repetitions each in random order.

For this stage only the magnitude estimation task was used, since the stimuli should be mixed completely randomly to avoid influences of the order of presentation of the individual conditions. The most lateralized phase value (averaged over all trials) of a

component was chosen for the final stimulus complex.

5.7 Set up

The used software framework ExpSuite[27] was developed at the Acoustic Research Institute in Vienna, where this study was conducted. It is based on Visual Basic and is designed for different types of psychoacoustic experiments including electric stimulation for cochlear implant listeners.

The generation of the stimuli as well as the calculation of results was realized in MATLAB. The audio processing was conducted in Pure Data (Pd) [28]. The Visual-Basic core program of ExpSuite is connected directly to MATLAB and Pd and controls experimental process. The stimuli were output via a 24 bit A/D-D/A converter (ADDA 2402, Digital Audio Denmark). They were processed by a digital audio interface (DIGI 96/8 PRO, RME) and sent through a headphone amplifier (HB6, TDT) and an attenuator (PA4, TDT) to circumaural headphones (HDA 200, Sennheiser). A sampling rate of 96 kHz was used to achieve short sampling intervals of 10 µs, which represented the minimum ITD.

The experiment was conducted in a double-walled sound booth in the laboratory of the Acoustic Research Institute.

Before the experiment started, the level of the signals was exactly calibrated using a sound level meter (2260, Bruel & Kjær) connected to an artificial ear (4153, Bruel & Kjær). Each pad of the headphones was calibrated using linear frequency weighting and a fast integration time of 125 ms (calibration level: 94 dB SPL).

[27] ExpSuite is a free software that can be downloaded at www.sourceforge.net/projects/expsuite. The package includes several different experimental models that were used for different studies at the institute and is constantly enlarged and actualized.

[28] Pd (Pure Data) is an open source visual programming language mostly used in computer music and multimedia applications. Pd is also often used for other applications that require real time audio processing.

6 Results

6.1 Experimental process

Most subjects provided a short feedback about the experiment during or after the sessions. The majority of them reported the tasks to be difficult, since no feedback was given after each trial and there is basically no correct or wrong answer. Especially at the beginning of the experiment many subjects were unsure about the accuracy of their responses.

The alternation of procedures in the first stage was regarded as positive diversification.

Because the first stage has been judged as difficult, the ITD was enhanced in the second stage in order to obtain larger magnitudes of lateralization and thereby hopefully also larger differences between the stimuli. One subject performed a pilot test of the second stage with 700 µs and 2000 µs ITDs. Based on the analysis of the results of this pilot test, she performed better for the larger ITD (see Figure 6.5 in Chapter 6.3). Therefore we decided to enhance the ITD to 2500 µs. This value corresponds to a quarter of the signal period (see Chapter 3.1) and thus is the largest delay not leading to aliasing effects. Even if ITDs beyond 700-800 µs exceed the natural ITD experienced by humans, they have been shown to be perceived alike at even farther lateral positions (see Chapter 3.3).

6.2 Intermediate results - First stage

Lateralization responses

In the first stage the signal was built up successively and varied in phase for each added frequency. Figure 6.1 shows the lateralization for each presented stimulus exemplarily for one of the subjects (NH43). Each plot represents one block in which the phase of one component was varied across one period of the fundamental frequency in 10 steps - from a three-component to a thirteen-component harmonic complex. The frequency component that was added and varied in phase in this block is noted in the title of each block. The plots show the mean values as well as the standard deviation of the 10 repetitions for each phase variation. Mean values, highlighted in red, correspond to

those phase values that obtained the maximal lateralization in the magnitude estimation task and were included in the ranking task for this block. The diamond shaped red value corresponds to that stimulus that achieved the highest count in the ranking task and therefore was chosen as phase value for this frequency component during the further procedure.

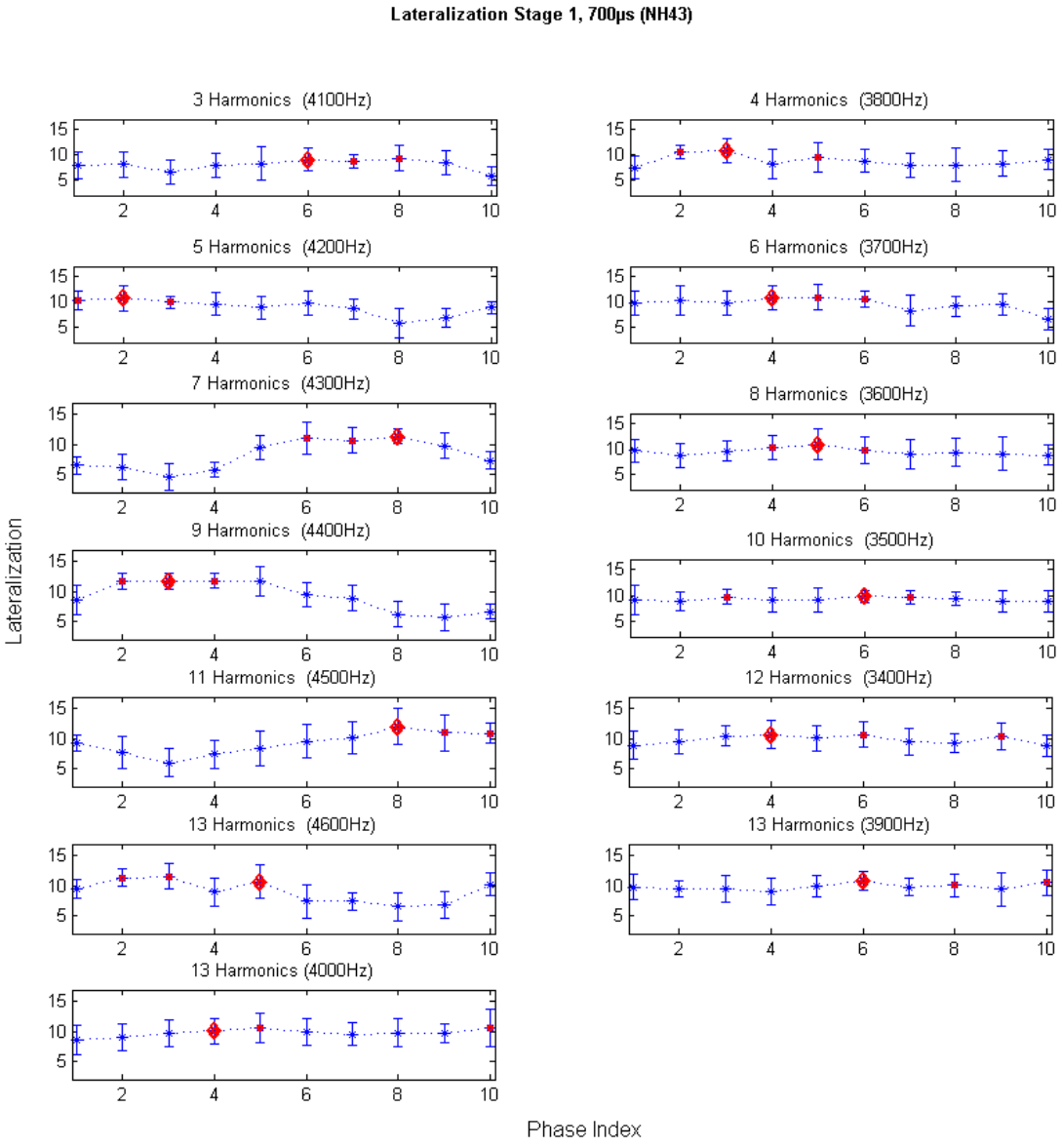


Figure 6.1: Lateralization results of Stage 1, exemplarily for one subject (NH43): Each plot shows the mean lateralization values and the standard deviation of each phase configuration for one block of the build up (newly added frequency component is noted in the title of each block). Positive lateralization values refer to a lateralization towards the right side (with the maximal scale value +15).

It can be seen that the phase values chosen for the ranking task are often adjacent ones. The lateralization is fluctuating smoothly over the phase cycle and in most cases there is only one maximum and one minimum. This basic pattern can be observed for all subjects. For some blocks these maxima seem to be very distinct, while for others they are less pronounced.

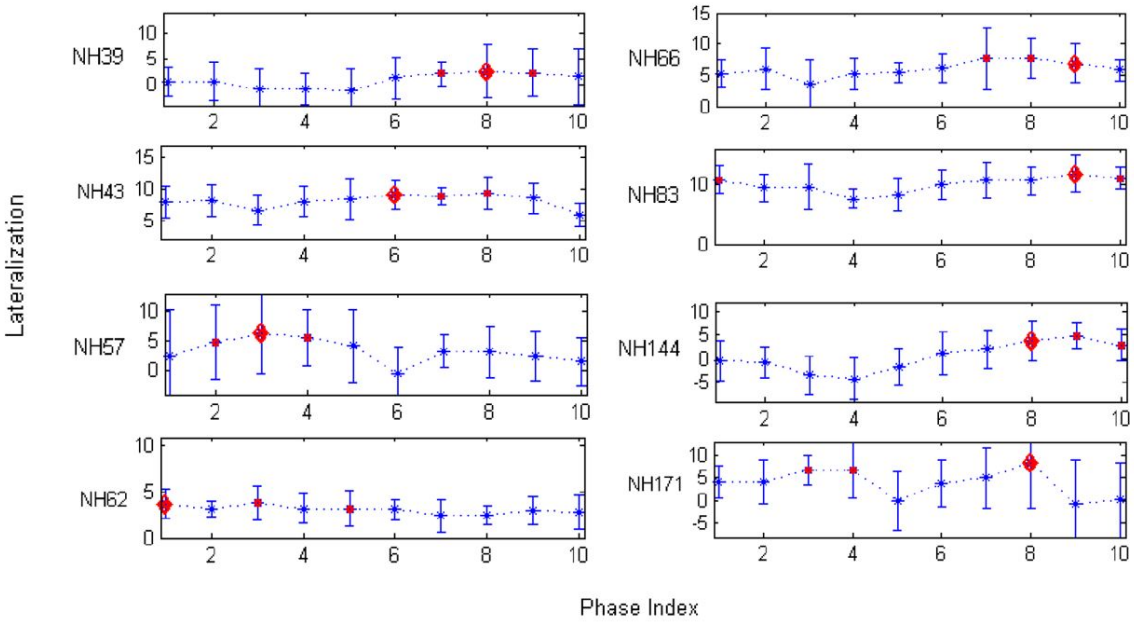


Figure 6.2: Comparison of lateralization results among all subjects for the same three component stimuli (Stage 1).

The absolute lateralization values and the position of maxima and minima, however, differ largely among subjects. Also the standard deviation varies highly between different subjects. As an example, Figure 6.2 shows the lateralization responses for all subjects for the first three-component stimuli. Since no individual phase value has been adjusted before, all subjects received exactly the same stimuli in this block. It can be seen that the shapes of the curves are similar in some cases (5 of 8 subjects chose the phases with number 8 or 9 as maximum) but can also differ strongly (one subject chose phase number 3, which is basically the phase inversion of the maxima for the other subjects). For most of the subjects the three phase conditions yielding the largest lateralization magnitudes correspond to adjacent phase conditions. The absolute magnitudes of lateralization differ largely across subjects, but since we are only interested in the relative differences in lateralization these differences are not important.

Some subjects tend to have a very small standard deviation and seem to stick towards a small range of lateralization for all their responses but partly show clear differences between stimuli. Others have a bigger range but also bigger differences between the repetitions of a single stimulus and between different stimuli. These properties are often consistent across blocks. The significance of these differences in lateralization across phase conditions is evaluated in Chapter 6.3.

The blocks have been presented consecutively in stage 1. As mentioned in the previous chapter, presenting only sound sources at one side of the head might lead to an adaptation in the use of the response scale and thereby to biased response behavior over a longer period of time. Therefore it may be potentially misleading to compare absolute lateralization values across blocks for this stage.

In order to examine the strength of this effect we compared the response behavior by calculating the mean lateralization response across all phase relations for each block. Figure 6.3 shows these values for all subjects. No common change in lateralization range for all subjects can be observed. Some subjects seem to show quite constant average lateralization responses across the whole experimental stage. The responses of the other subjects shift during the experiment - but equally often towards the right side as to the left. The mean value across subjects (black line) stays almost constant during the experiment. Hence, this analysis suggests that our results seem to be not critically influenced by response scale adaptation effects.

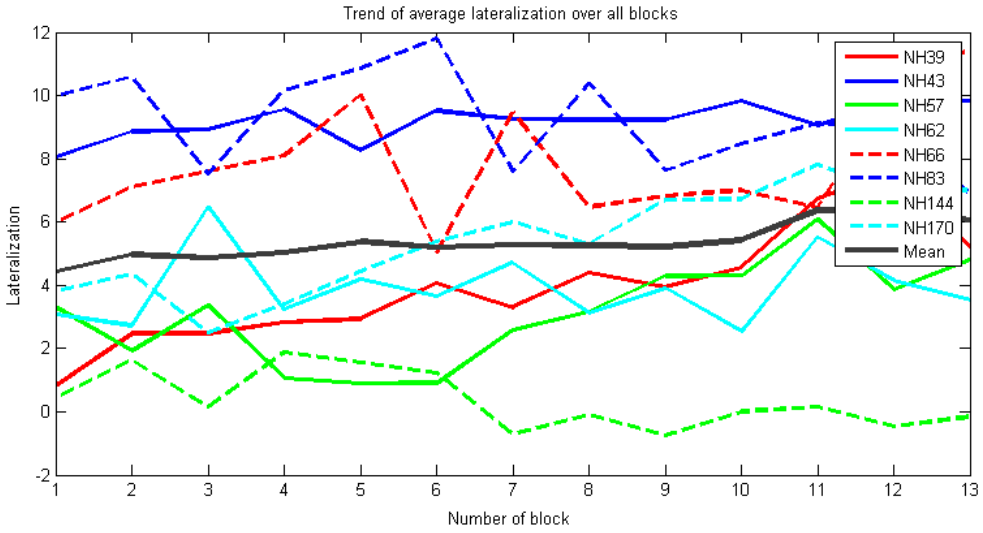


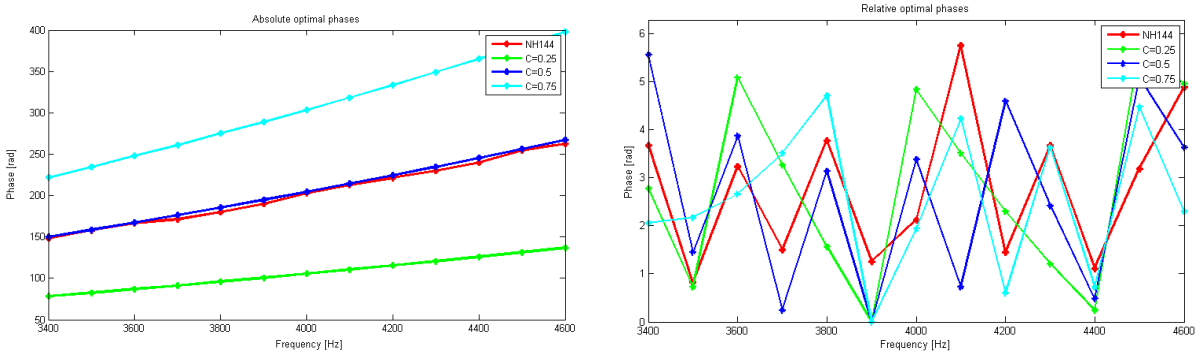
Figure 6.3: Comparison of mean values of each block for all subjects (Stage 1).

Phase response estimation

The phase values determined in this first stage can be shown in absolute or relative values. As an example, the phases of one subject (NH144) are plotted in both ways in Figure 6.4. For comparison, also the values for SPHCs with C values of $C = 0.25$, 0.5 and 0.75 are displayed.

Since we used the phase values of $C = 0.5$ as basis for our phase variations over the range of one signal period, the absolute phase values can be maximal one phase cycle apart from the phase values of this curvature. The plot with absolute phase relations thereby only serves as an illustration of this connection.

When reducing the phases to relative values by subtracting full signal periods, this additional information is disregarded and the values can be compared equally with all three constant phase curvatures. Based on earlier studies using the masking paradigm, which found constant phase curvature around $C = 0.5$ to 0.8 to best match the experimental data for the frequency region tested (see Chapter 4.1, [Lentz and Leek 2000]), we expected the curvatures with $C = 0.5$ and $C = 0.75$ to be closer to the subject's phase values than $C = 0.25$. The phase results of this subject show only little similarity with the phase values of the three SPSCs, however, they show a tendency to be most similar to the signal with $C = 0.5$, which is in agreement with the earlier studies. The relative values change strongly for a small change of the C-value.



(a) Absolute optimal phase values for one subject. (b) Relative optimal phase values for one subject.

Figure 6.4: Phase values of one subject (NH144) in stage 1 in comparison with three SPHCs with phase curvatures of $C = [0.25, 0.5, 0.75]$.

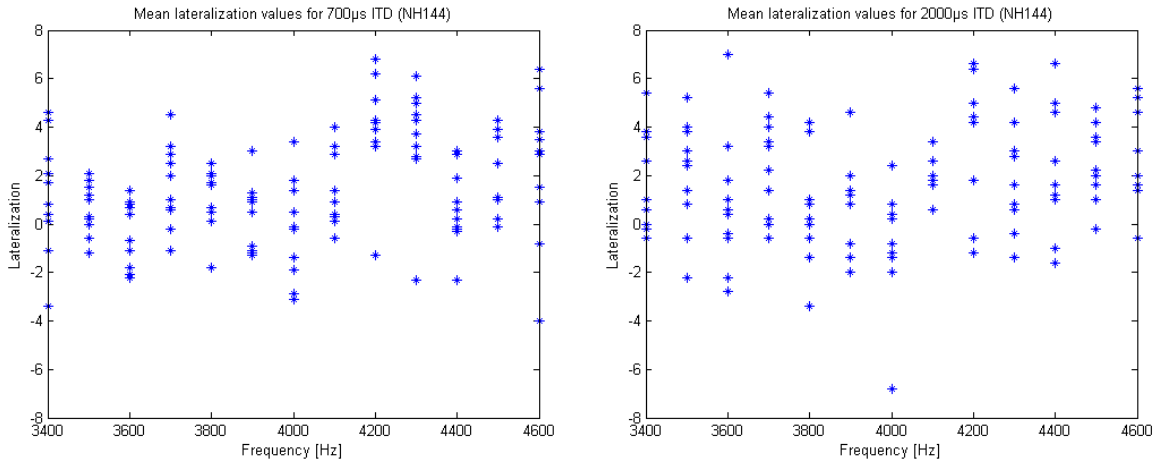
The relative phase values vary strongly amongst subjects and all differ from the phases of the compared SPHCs. The phase results of this first stage represent only a rough approximation, since interactions between the components have not been fully considered in this stage. The next stage will include these effects and is thus supposed to increase the accuracy of the phase response estimation.

6.3 Final results - Second stage

In the second stage, the individually adjusted final signal resulting from the first stage was used as a basis to vary each component's phase within the whole complex. In this stage a comparison across blocks is possible since the stimuli were presented in random order and not consecutively like in the first stage.

Pilot test for determining the best suitable ITD

As mentioned in Chapter 5.1 the ITD of the stimuli was enlarged for the second stage. The reason for this enlargement was a pilot test prior to the second stage of the experiment with one of the subjects, who tested the set up for two different ITD values of 700 µs and 2000 µs.



(a) 700 µs ITD

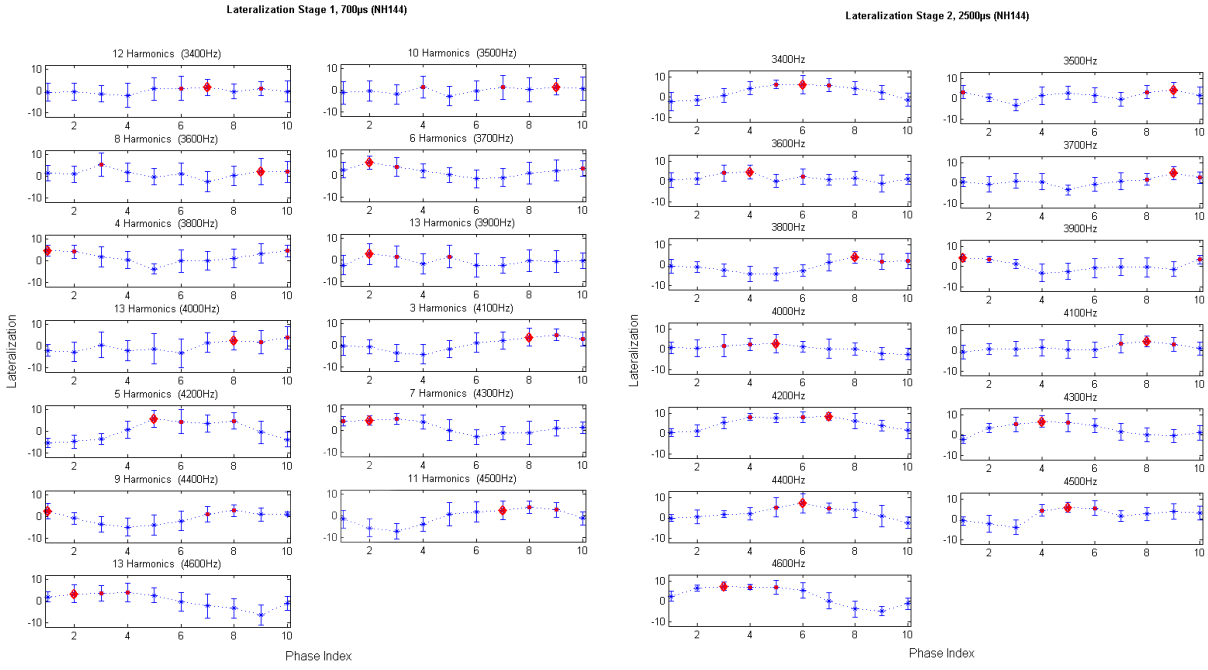
(b) 2000 µs ITD

Figure 6.5: Scatter plots: Comparison of mean lateralization values for all stimuli in stage 2 with different ITDs for one subject (NH144). Each symbol corresponds to the mean response for one phase value.

Figure 6.5 shows scatter plots displaying the mean lateralization value of each stimulus for both ITDs. For the larger ITD (right panel) the mean values within one block seem to be spread wider over the lateralization scale than for the lower value (left panel). Thus the maxima are more distinct. The subject performing the pilot test also showed a large lateralization range in the first stage. Thereby we might expected even larger improvements for the subjects that tended to have a small range of lateralization values, often leading to indistinct maxima, in the first stage. We finally chose the highest possible ITD of 2500 µs (see Chapter 5.1) for the second stage to achieve the maximal enlargement in lateralization.

Lateralization responses

In the second stage, each stimulus contained the whole complex, whose phase values were based upon the results of the first stage. Thus, the second stage takes the interactions between the components into account.

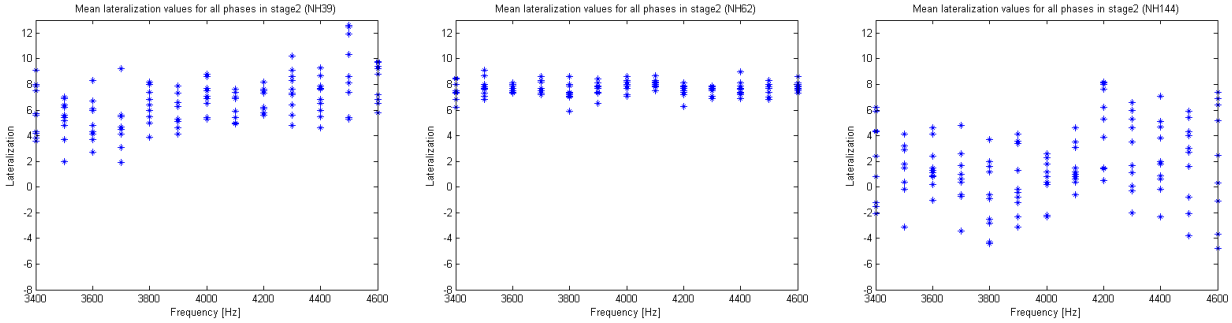


(a) First stage (b) Second stage

Figure 6.6: Lateralization results for the two stages for one subject (NH144). Plots arranged in spectral order.

Figure 6.6 compares the lateralization results of both stages exemplarily for one of the subjects (NH144). The maxima for one component often differ but the overall shape of the curvatures is often similar in both stages. The second stage seems to have generally smaller standard deviations, which could be caused by the simplification of the lateralization task by enlarging the ITD. Since every subject started the second stage with his/her individually determined phase values, the lateralization differs among subjects in most cases.

Figure 6.7 shows scatter plots of the mean values of lateralization for all phase variations of all thirteen components of the complex for three exemplary subjects. Since all stimuli were presented randomly we are able to compare absolute values across components for each subject.



(a) Subject NH39 (b) Subject NH62 (c) Subject NH144

Figure 6.7: Scatter plots for lateralization results in stage 2.

It can be seen that subjects use the lateralization scale very differently. Some of them, like NH62, restrict their responses to a very narrow range of values. Others, like for example NH144, seem to use a larger range. This can also be seen in Table 6.1 that outlines the overall lateralization range for all subjects[29]. The response behavior (range and absolute values) of each subject seems to be similar in both stages.

	NH39	NH43	NH57	NH62	NH66	NH83	NH144	NH171
Range (Stage 1)	[-2 10]	[4 12]	[-3 8]	[2 8]	[3 13]	[4 13]	[-7 7]	[-3 9]
Range (Stage 2)	[2 13]	[6 13]	[4 10]	[6 9]	[5 12]	[6 12]	[-5 8]	[-5 8]

Table 6.1: Ranges of lateralization magnitudes for all subjects.

[29]The scatter plots of all subjects are attached in the appendix.

Statistical Analysis

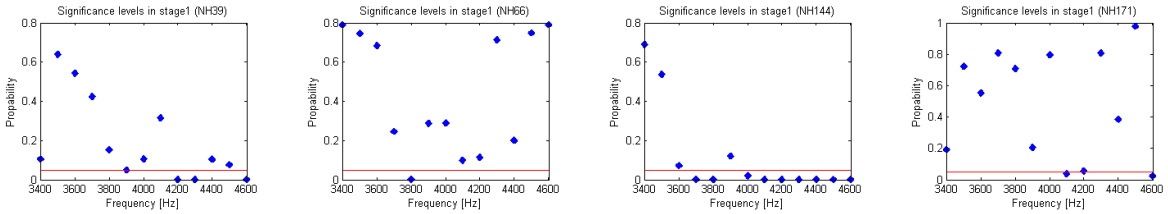
In order to determine the quality of the subjects' responses, we calculated the significance of the differences in lateralization between the phase relations for each component of each subject. Therefore one-way ANOVAs (analysis of variances) were used (extended versions of the t-test, allowing to compare multiple factor levels to each other). Each ANOVA examines whether the variance in lateralization across conditions exceeds the variance among repetitions of one stimulus. The probability value p for the null hypothesis that all groups of values are equally lateralized has to lie below the threshold of $p = 0.05$ in order to achieve significance.

Table 6.2 shows the number of significant blocks out of 13 for each subject and both stages. Also the minimal p-values, i.e. the probabilities of the most significant blocks are shown for each stage. There are big differences among subjects. Some of them seem to perform well at both stages (NH39, NH43, NH144), some seem to be better at either the first (NH83) or second stage (NH66) and others are mostly insignificant in their response behavior in both stages (NH57, NH62, NH171).

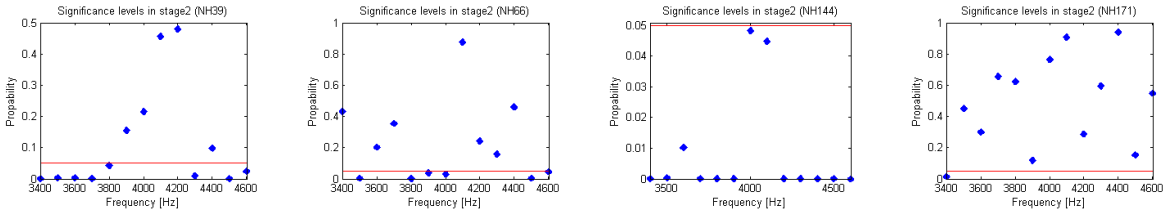
	NH39	NH43	NH57	NH62	NH66	NH83	NH144	NH171
Significant blocks (Stage 1)	4	7	2	1	1	5	9	2
Significant blocks (Stage 2)	7	6	0	0	6	0	11	1
Lowest p-value (Stage 1)	$3.7 \cdot 10^{-5}$	$1.3 \cdot 10^{-15}$	$3.1 \cdot 10^{-3}$	$5.0 \cdot 10^{-2}$	$2.9 \cdot 10^{-3}$	$3.1 \cdot 10^{-4}$	$8.8 \cdot 10^{-13}$	$2.4 \cdot 10^{-2}$
Lowest p-value (Stage 2)	$8.9 \cdot 10^{-14}$	$5.9 \cdot 10^{-12}$	$7.1 \cdot 10^{-2}$	$5.9 \cdot 10^{-2}$	$1.5 \cdot 10^{-3}$	$5.8 \cdot 10^{-2}$	$1.5 \cdot 10^{-21}$	$1.5 \cdot 10^{-2}$

Table 6.2: Number of significant blocks out of 13 and probability of the most significant block for all subjects.

Figure 6.8 shows the p-values for four exemplary subjects. The red line in each plot marks the significance border of $p = 0.05$. The probability values are shown for both stages. As seen above in Table 6.2, big differences occur across subjects. NH39 seems to perform above-average in both stages. NH66 shows a high improvement from first to second stage. NH144 performed great in both stages and achieved the highest significance in this experiment. NH171 had a low amount of significance for both stages.



(a) Stage 1, NH39 (b) Stage 1, NH66 (c) Stage 1, NH144 (d) Stage 1, NH171



(e) Stage 2, NH39 (f) Stage 2, NH66 (g) Stage 2, NH144 (h) Stage 2, NH171

Figure 6.8: Significance of blocks in both stages for four subjects: the red line marks the 0.05% significance boarder, lateralization results below this border are significant.

The subjects seem to reveal big differences in the quality of their performance. Since four of the eight subjects showed more reliable results in the second experimental stage, the analyses in Chapter 6.4 will focus on those subjects[30].

Phase response estimation

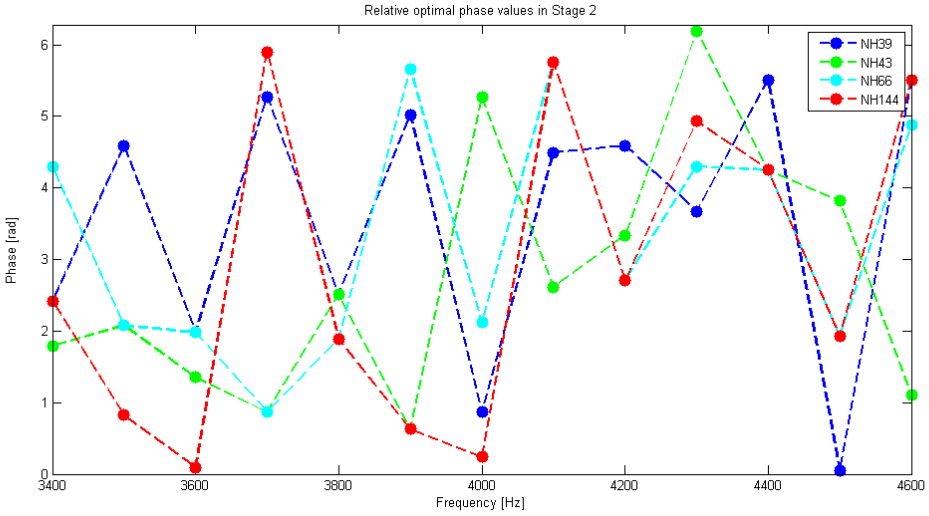
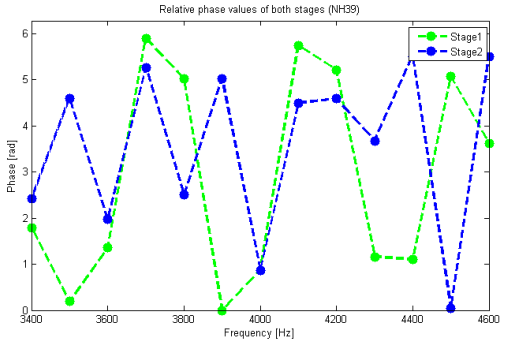


Figure 6.9: Relative phase values of stage 2 of the four subjects with highest significance.

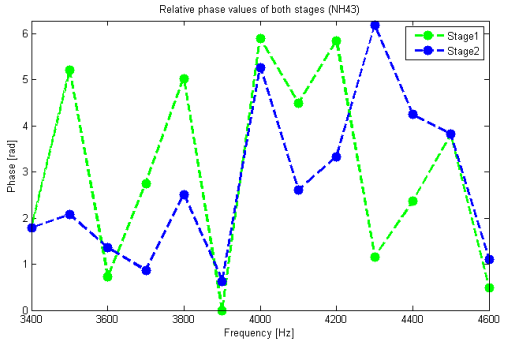
[30]The other four subjects will be used to compare results, since we assume they nevertheless contain relevant information.

Figure 6.9 shows the phase estimations from the second stage, plotted as relative phases, for the four subjects showing significant response effects as defined above[31]. The phase values show some similarities across listeners. For some components two or three subjects chose the same phase values.

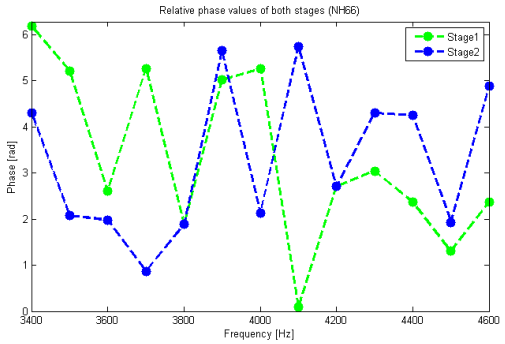
Figure 6.10 compares the relative phase results between the first and second stage for these four subjects. The results of the first stage can be seen as a rough approximation whilst the results of the second stage include mutual influences of the components and thereby appear to be more relevant. For some of the components the phase values seem to be the same or very similar [32]. In most cases the phases are less than $\frac{\pi}{2}$ apart. In some cases, however, the phases seem not to be connected.



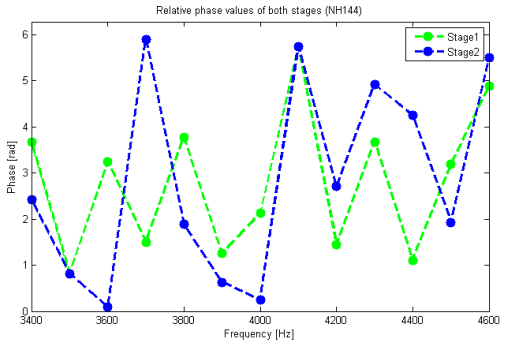
(a) NH39



(b) NH43



(c) NH66



(d) NH144

Figure 6.10: Relative phase values of first (green) and second stage (blue) for the four subjects with high significance in the second stage.

[31] A plot with the phase values of all eight subjects is attached in the appendix. The phase values across all eight subjects vary strongly.

[32] Since the phase is circular and 0 corresponds to 2π, differences in phase cannot be displayed directly in all cases.

6.4 Adjustment with Schroeder phase harmonic complexes

Comparison of the individual phase results with SPHCs

In order to compare our results with previous studies, the constant phase curvatures best-fitting the phase estimations of each subject were computed. Therefore the relative estimated phase values were compared with the relative phase values of SPHCs. The absolute distances between all pairs of phases were computed and added across components to obtain one value for the degree of accordance between the two signals. Signals with constant phase curvatures from $C = 0$ to $C = 1$ were used, since we assume from results of earlier studies that the signal best mirroring the auditory filter's phase response has a positive phase curvature (see Chapter 4.2). The distances were computed as a function of the C-value of the SPHCs.

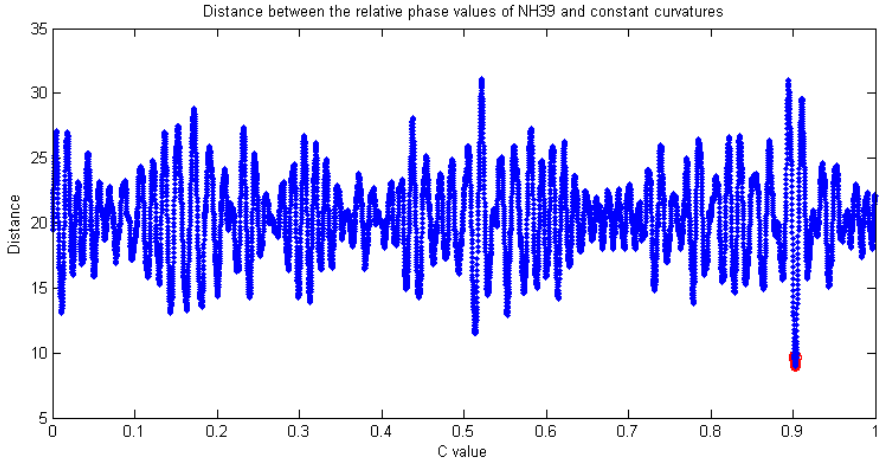


Figure 6.11: Sum of distances between the relative phases of the result stimulus in the second stage and SPHCs with $C = 0$ to $C = 1$ for one subject (NH39). The red points characterize the 10 lowest values.

Figure 6.11 shows these sum of distances in rad between stimuli and SPHCs exemplary for one subject (NH39). The best coinciding SPHC is the complex with $C = 0.90$, which differed only by 9 rad from the result values of this subject[33]. The total phase differences seem to oscillate over the range of C in fast fluctuations.

Figure 6.12 shows the relative phase results of the second stage for the same subject (NH39) together with the relative phase values of the two closest constant curvatures.

[33]The maximal difference is reached by the complex with C=0.52 with 31 rad.

The curves are very similar for the majority of components[34] but still show large differences for some components.

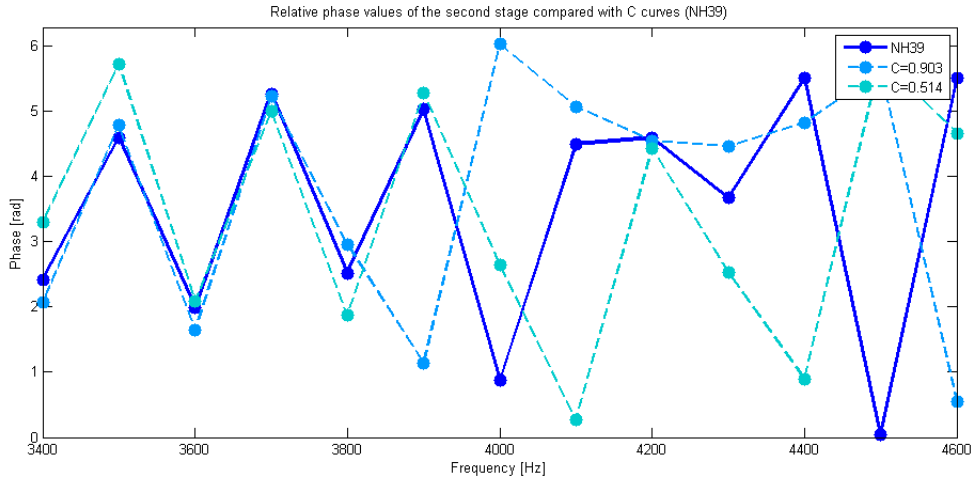


Figure 6.12: Relative phase values (Stage 2) compared with the two SPHCs that show the least differences in phase (see Figure 6.11) exemplarily for one subject (NH39).

The C-values of the best-fitting constant curvatures for each subject are displayed in Table 6.3. The analysis has been made for both stages. The curvatures seem to be spread over the whole range of C.

	NH39	NH43	NH57	NH62	NH66	NH83	NH144	NH171
C-value (Stage 1)	0.307	0.661	0.809	0.489	0.569	0.165	0.119	0.134
C-value (Stage 2)	0.903	0.219	0.660	0.537	0.481	0.657	0.683	0.080

Table 6.3: Constant phase curvatures that show the minimal distance to the phase results of each subject.

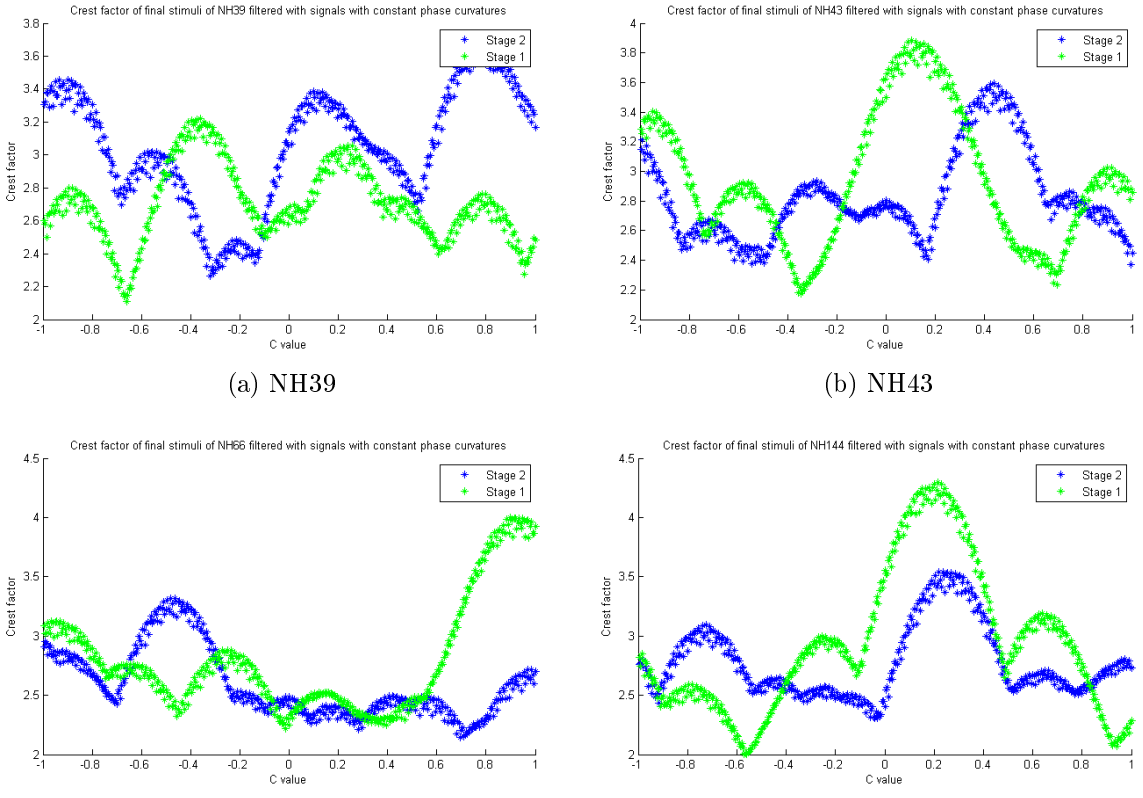
This attempt to find thebest-fitting SPHC is based just on summed distances in the phase across components. There is a possibility that this distance metric may be coincidentally very low for any random phase curvatures. Also, the method often shows not one but several very different phase curvatures that achieve almost the same summed distance. The difference in distance for very similar C-values (differences lower than 0.01), however, can actually be extremely high.

Based on these considerations, an alternative analysis with SPHCs was conducted, which is described in the following subchapter.

[34] Due to the circularity of phase the differences might again look larger for values extending 2π.

Filtering with phase responses

In order to find out which constant curvature best models the auditory system and thereby evokes the peakiest envelope, we used various phase curvatures as filters[35]. Then we calculated the crest factor of each of the filtered signals. Figure 6.13 shows the crest factors for each filter signal over the range of $C = -1$ to 1 (blue line). For comparison, the same procedure has been applied to the phase estimations from the first stage (green line). The plots are shown for the four subjects, who performed above-average in terms of significance. The curvatures with the highest crest factors are expected to evoke the peakiest signals and thus to best model the filtering of the auditory system. We expect the maximal crest factor to lie in the positive range of C.



(a) NH39

(b) NH43

(c) NH66

(d) NH144

Figure 6.13: Crest factors of the individual result signals of the first and second stage filtered with constant curvatures with C-values between $C = -1$ and $C = 1$ for four subjects in stage 1 (green) and stage 2 (blue).

[35]This filter is merely a hypothetical filter, where the phase filtering is already considered in the synthesis.

Figure 6.14 shows smoothed versions of the curves for the second stage shown in in Figure 6.13). In this plot all four subjects can be compared directly. The highest maxima lie in the positive range of 0.25 to 0.9 for three of the subjects. For one subject (NH66) the crest factor maximum lies in the negative range at about -0.45. One subject (NH39) showing the largest maximum in the positive range shows two additional (only slightly smaller) maxima one of which is in the negative range. The other subjects seem to have only one distinct maximum[36].

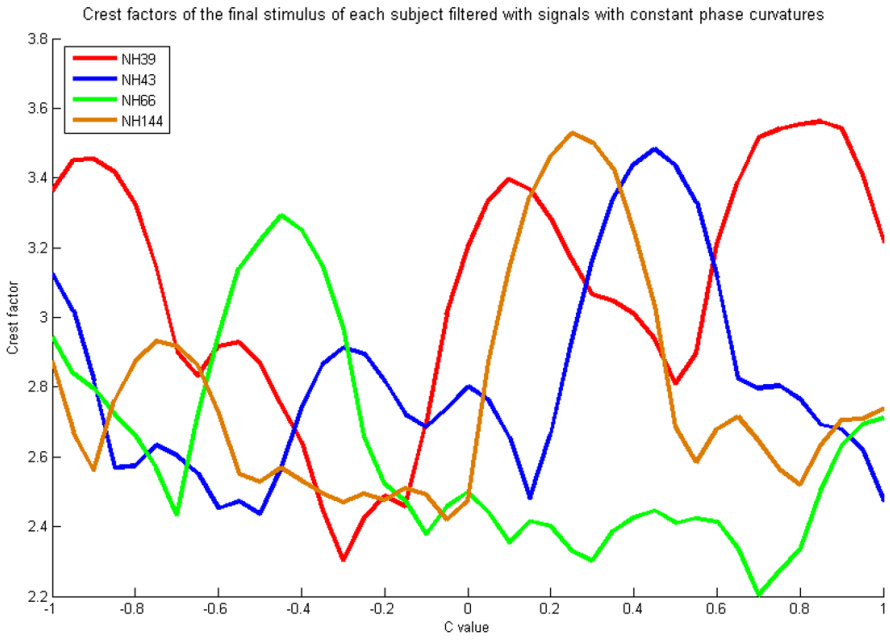


Figure 6.14: Smoothed crest factors of Figure 6.13.

Figure 6.15 shows the C-value of the filter signals causing the highest crest factors for all eight subjects in both stages. It can be seen that in the first stage just 5 out of 8 subjects had their maximal crest factor for a signal with a positive curvature. In the second stage 7 out of 8 subjects chose a signal with phase values that are mirrored best with a positive curvature.

A sign test was conducted to examine the significance of these results [Karas and Savage 1967]. The one-tail sign test revealed that, for the second stage, the number of subjects, who chose phase values similar to positive phase curvatures is significantly higher than chance ($p = 0.035$).

[36] Additional plots of the other subjects are attached in the appendix.

For most of the subjects, the crest factor maximum changed considerably between first and second stage, for example for NH66, who had the highest C-value of all subjects in the first and the lowest value in the second stage. That corresponds with the results of the estimated relative phase values, which vary strongly between first and second stage (see Figure 6.10) and also seem to be most similar amongst the four significant subjects in the second stage.

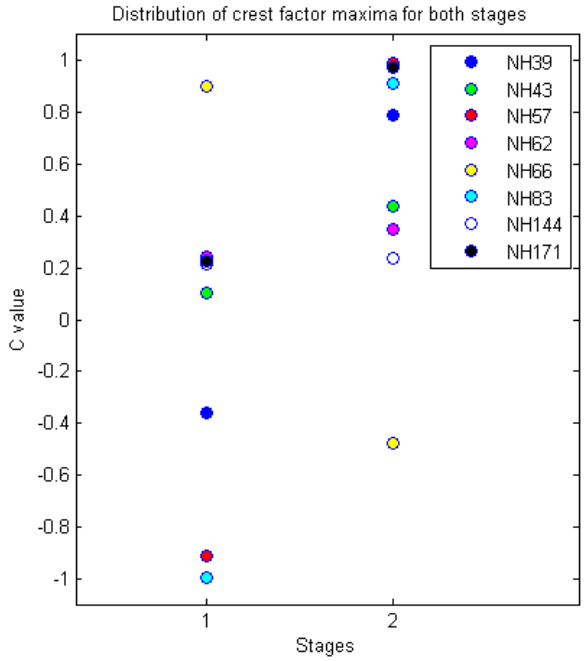


Figure 6.15: C-values of the maximal crest factors for the filtered signal for all subjects (Stage 1 and 2).

For a better understanding of the filtering process, Figure 6.16 shows the waveforms of the used signals. The first waveform is the result signal of stage 2 for subject NH39. The second waveform shows a signal with a constant phase curvature that was used to filter the result signal. In this example, the filter phase response ($C = 0.79$), which evokes the highest crest factor in the output signal, is shown. The waveform of this signal is shown in the last plot. It is assumed to be the peakiest internal waveform for the class of harmonic complexes used in the experiment.

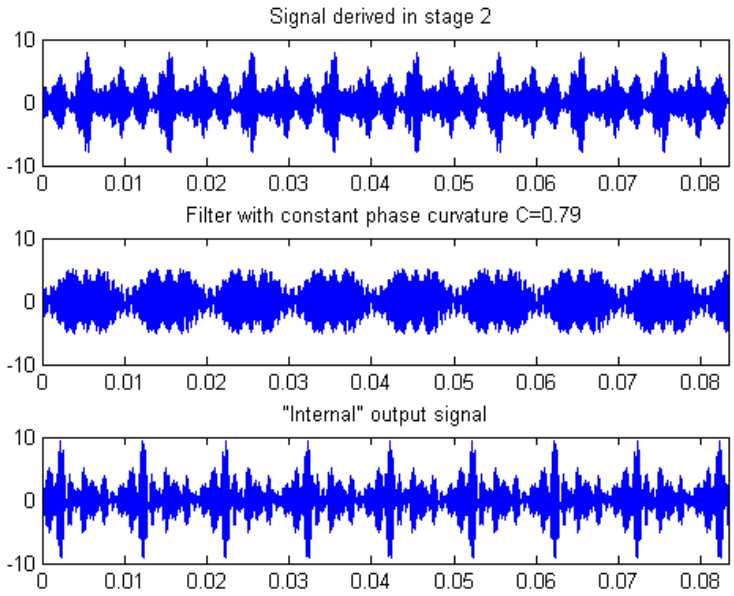


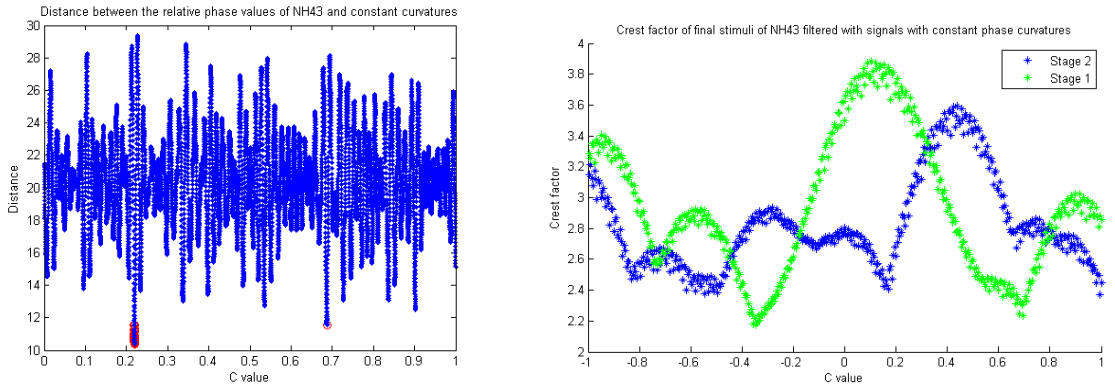
Figure 6.16: Waveforms of the result stimulus in stage 2 (upper plot) filtered with a constant phase curvature (waveform of SPHC with this phase response shown in middle plot) that evokes the output signal with the highest crest factor (lower plot).

Comparison of previous analyses with constant phase curvatures

Both methods outlined above attempted to find constant phase curvature signals that best match the experimentally adjusted stimuli in order to estimate the human phase response of each subject. In the first case the relative phase values between result stimuli and SPHCs were compared. In the second case the signal was filtered with constant phase curvatures and the crest factor of the output signal was calculated.

The first method is based solely on the relative phase values. As can be seen in Figure 6.11, the distances fluctuate strongly as a function of C. This method shows no wider range of C that matches best, as suggested by other studies. Each maximum is followed by a minimum within a short distance and only a small aberration in C might result in a large change in distance.

In the second method the crest factor was computed for each filtered phase curvature. For this method, the fluctuations across the range of C are smaller. In most cases the maximum is in the positive range of C. The data of most subjects show just one distinct maximum. Figure 6.17 compares the results of both methods for one subject.



(a) Relative distances of phases

(b) Filtering with constant phase curvatures

Figure 6.17: Comparison of analysis methods exemplarily for one subject (NH43).

Even when the first method doesn't show large-scale changes over C it can still contain useful information. In Table 6.4 the C-values of the best fitting signals of both methods are listed. The distance model was applied for the range of $C = [0, 1]$, the filtering model for the range $C = [-1, 1]$. The best fitting phase curvatures of the two methods seem to be not related to each other. This discrepancy is not surprising given that the distance method revealed a wide range of similarly fitting C-values with no distinct best value.

	NH39	NH43	NH57	NH62	NH66	NH83	NH144	NH171
Distances Stage 2	0.903	0.219	0.660	0.537	0.481	0.657	0.683	0.080
Filtering Stage 2	0.79	0.44	0.99	0.35	-0.48	0.915	0.235	0.975

Table 6.4: Best fitting constant phase curvature C for both analysis methods.

7 Discussion

In this study the auditory phase response of normal hearing subjects was estimated by using an ITD lateralization task. The results show that phase relations play an important role in ITD perception and thereby an ITD task is suitable to discriminate between different phase relations within one stimulus. The results show a large amount of variability across the different subjects but also show some similarities to earlier studies on the phase response.

Studies on the human phase response so far have been mostly based upon masking tasks and been conducted mostly with normal hearing subjects. They discovered that signals with constant phase curvatures lead to a strong masker-phase effect and assumed from their results that the auditory phase response of NH listeners has a positive constant phase curvature [Smith et al 1986, Kohlrausch and Sander 1995]. Later studies also suggested that this curvature decreases with increasing center frequency of the corresponding auditory filter [Lentz and Leek 2001]. Matching the subjects' phase estimations resulting from the current experiment to constant-curvature phase responses yields a tendency towards positive curvatures, however, spread over a large range of C for the different subjects.

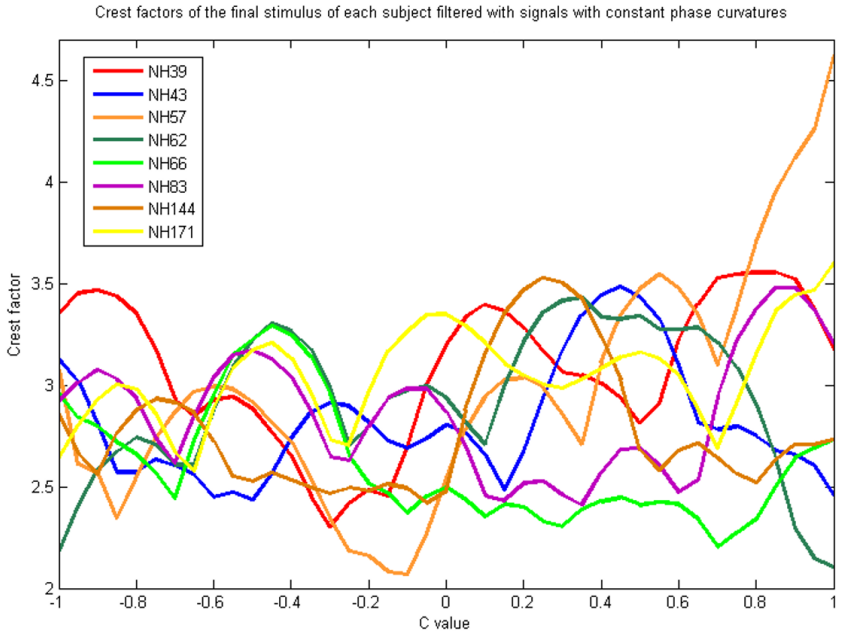


Figure 7.1: Crest factors for the filtered signal in Chapter 6.4 for all subjects.

The masking study by Lentz and Leek [2001] determined the phase response of the same auditory filter (at 4000 Hz) and found the best fitting C-values for their four subjects to be $C = 0.7$,0.9,0.5 and 0.7 (see Figure 7.2). All subjects showed a broad maximum in this area, however, the absolute masking values differed strongly between the subjects. In comparison, in our experiment the maximal crest factors for the filtered signals were $C = \{-0.48, 0.235, 0.35, 0.44, 0.79, 0.915, 0.975, 0.99\}$ (see Chapter 6.4). The curves of the crest factors for all eight listeners are shown in Figure7.1.

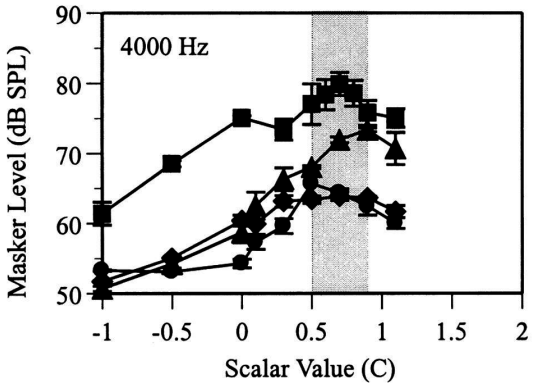


Figure 7.2: Masker level for SPHC maskers at 4000 Hz with different phase curvatures between $C = -1$ and 1 [Lentz and Leek 2001].

Another study by Shen and Lentz [2009] examining the level dependency of auditory phase curvatures, also used stimuli with a center frequency of 4000 Hz and a fundamental frequency of 100 Hz. Figure 7.3 shows the results for a fixed masker level of 70 dB (comparable to the level in our experiment). The strongest masker-phase effect is found for $C = 0.5$ (for three subjects) and $C = 0.75$ (for one subject).

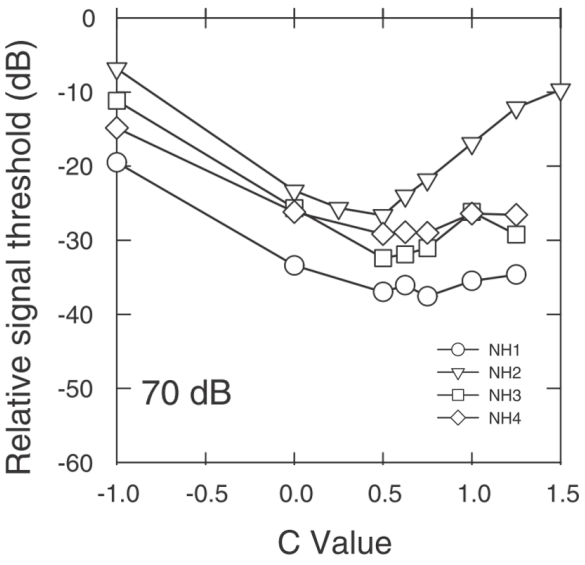


Figure 7.3: Relative signal threshold for SPHC maskers at 4000 Hz with different phase curvatures between $C = -1$ and 1.5 for four individual subjects at 70 dB SPL masker level [Shen and Lentz 2009, rearranged by the author].

Several similar studies using signals in signals at various frequency ranges and fundamental frequencies were performed, e.g. by Oxenham and Dau [2001] and Oxenham and Ewert [2005]. For all conditions masking signals with positive curvatures achived the lowest thresholds. The studies were performed with low numbers of subjects and showed some variability across subjects.

The large amount of inter-individual variability in the results of this experiment may be caused by various reasons some of which are addressed in the following.

The statistical analysis showed that the subjects differed largely in the quality of their lateralization responses. Some subjects gave significantly distinct lateralization responses for different conditions, while others obviously perceived less differences between the phase conditions. These differences could arise due to a individual differences in their sensitivity to ITD cues. ITD sensitivity is known to elicit large inter-individual differences across the normal hearing population [Saberi and Antonio 2003]. Alternatively, it is possible that the listeners, showing less differences between phase conditions, could perceive differences but they could not map them on the lateralization scale as used in our experiment.

Some of the subjects participated in earlier ITD studies and were shown to be sensitive to ITDs (NH39, NH43, NH144). These three subjects also performed well in the present experiment. Since the previous ITD study was carried out over a relatively long period of time, learning effects might have been strong. However, we do not know exactly if these listeners' ITD sensitivity is better than that of the subjects showing less significant phase effects in the current study.

Since the method of the current study is based on subjective responses of the subjects it may not be as accurate as objective measurements. Measuring ITD thresholds would be a more precise alterative, but it is a very time-consuming task and would not be accomplishable for the amount of conditions tested in this experiment. However, it might be useful to examine ITD thresholds of the stimuli with phases determined in the experiment for each subject.

Yet unpublished ITD JNDs measurements[37] using SPHCs showed strong effects of varying the C-parameter (i.e., the phase response), which were very consistent across the pool of seven NH listeners tested. This may suggest that the variability between listeners in the present study was more due to their ability to deal with the particular methodology and stimulus conditions of the present study rather than to their overall ITD sensitivity.

The results obtained in this experiment do not allow to either support or reject a uniform curvature of the phase response. It is unlikely but might be possible that the human phase responses of normal hearing listeners do not reveal a uniform phase curvature. While many established studies have suggested this (see Chapter 4), objective studies, physiologically measuring the phase response in the cochlea, are only available for animals ,e.g. Summers et al. [2003]. For the human auditory system the uniformity of the cochlear phase response has not been proven entirely.

The method was designed to be suitable for hearing impaired subjects. Most studies on human phase response used masking tasks, which work well for normal hearing subjects but can't be performed adequately by hearing impaired subjects due to their reduced cochlear compression (see Chapter 4, [Oxenham and Dau 2004]). This was the reason for approaching the problem using the lateralization task. The task seems to be more difficult for NH subjects than a masking task and leads to less clear results,

[37] Tabucci and Laback (2014), unpublished internal communication.

e.g. compared to the study of Lentz and Leek in 2001. However, it is not known how HI subjects would perform in this ITD task. On the one hand the task is harder for HI subjects since they have an overall reduction in ITD sensitivity [Moore 1996]. On the other hand, only relative differences between phase conditions are important for the lateralization task, thus absolute ITD sensitivity may not play an important role. Also, envelope the ITD perception sensitivity is reduced to a lesser amount than fine-structure ITD sensitivity other lateralization cues and might thus be a more salient cue important for HI listeners than for NH listeners.

The method used in this experiment shows that phase relations play an important role in ITD perception and thereby in turn, ITD tasks can be used to gain information about the phase response in the human auditory system. Although the accuracy of the new ITD-based lateralization paradigm used in this experiment seems to be lower than that for the masking paradigm, it appears worth to further work on this method. The big advantage of the new paradigm is that it is suitable for both normal hearing and hearing impaired subjects. While the masking paradigm relies on cochlear compression, which is reduced or even absent in HI listeners, the estimation of the phase response with the ITD-based lateralization paradigm does not rely on the presence of cochlear compression.

The results showed that the subjects chose fairly different phase relations for the first and second stage of the experiment. The second stage is assumed to provide more reliable results since it involves interactions between the subjects. Therefore it might be reasonable to expand this part of the experiment for receiving more precise results. If the experimental time limit allows it, the procedure could be conducted over more iterations or a more precise method such as the ILD-pointer method [Bernstein and Trahiotis 2011] might be used. The first stage could be reduced or even replaced with a shorter procedure. Since we expect hearing impaired subjects to have very different individual phase responses, this first stage is still important to find phase values that approximate their phase response and allow a fine adjustment in the second stage. For normal hearing subjects presumptions based on existing knowledge could be used to replace the first stage. Optimally, alternative experimental procedures would be tested

with the same subjects, preferably including those yielding significant phase effects in their lateralization responses. This would allow to compare the results with the present data.

Additionally, the analysis of results could be extended. In this thesis the output waveform of the result stimuli filtered with the phase values of SPHCs was only examined for the crest factor. The crest factor is an straight-forward method to characterize the peakedness of a signal, however, it only depends on the ratio of maximum to effective amplitude and disregards the shape of the waveform. For example, temporally inverting the output signal in Figure 6.16 would lead to different slopes of the onsets in each modulation cycle , potentially changing the salience of ITD cues, but still the crest factor remains constant. Including other parameters of the waveform like the steepness of the envelope slopes or the duration of the off-times, which are known to be important for envelope ITD perception (see Chapter 3.3, [Laback et al. 2011, Klein-Hennig et al. 2011]) might improve the quality of the analysis. Incorporating such analyses would, however, have exceeded the time constraints for this thesis.

The filtering of the result stimuli in Chapter 6.4 has been made only with the phase values of the SPHCs. Including also the amplitudes of these signals and thereby the amplitude response of the underlying auditory filter, might improve the analysis.

The results of this experiment could additionally be compared with the phase properties of established computational models of the basilar membrane. Older models not or not appropriately considering active processes in the cochlea (e.g. the gammachirp model by Patternson or earlier models like the basilar membrane filter by Strube (1985) or the transmission line model of cochlear processing by Giguere and Woodland (1994)) have been shown to not account for phase effects (see, e.g., Oxenham and Dau [2001]). However, comparisons could be made to more recent approaches such as the model by Zilany et al. [2009].

8 Summary and conclusion

The experiment conducted in this work used ITD perception to estimate the phase response of auditory filters of individual subjects. Although auditory filtering is a highly nonlinear mechanism inside the cochlea, the phase response is assumed to have a constant curvature for normal hearing listeners. Hearing impaired listeners, however, potentially have individually different phase responses due to their damages at specific locations of the basilar membrane. ITD is a localization cue that can be perceived by NH and HI listeners. The ITD sensitivity for HI listeners is reduced but is still expected to be high enough to allow measurements of the phase response, especially in the case of high frequency envelope ITD, which has been used in this experiment. The experiment used harmonic complexes as stimuli, which were presented with various phase relations. The complexes were first constructed by successively adding each component to an initial three-component complex. For each additional component the phase value was varied over one period and the phase value producing the largest magnitude of lateralization was chosen for this component. After the complex has been constructed in this way, in a second stage, which was using the entire complex, the phase values of each component were individually changed in random order. The signals that evoke the strongest magnitude of lateralization are assumed to best mirror the phase response of the subject and to produce a highly peaky internal signal at the output of auditory filtering. The experiment was conducted by eight normal hearing subjects.

All subjects showed a dependency of lateralization on the phase relations. Some subjects used a large range on the lateralization scale for their answers, others seemed to restrict their answers to a small range but in some cases also had smaller standard deviations. The experiment showed a large inter-subject variability in the estimated phase values. In the first stage no common result can be obtained from the values. For the phase results of the second stage, some similarities can be seen between the four subjects that achieved statistically significant phase effects in their lateralization responses.

Two methods were applied to compare the resulting phase values to signals with constant phase curvatures. In the first method, the relative phase values of the result stimuli of each subject were compared with the relative phases of SPHCs. This method showed large fluctuations for slight differences of the phase curvature and no pronounced

maximum in any region. It was thereby assumed that the comparison of relative phase values is too vague and can be affected by chance effects that don't reflect the real properties of the phase response.

In the second method the result signal was filtered with different constant phase curvatures. In theory the phase curvature best reflecting the human phase curvature will evoke the peakiest waveform for the output signal. For this reason the crest factors of the output signal were calculated for all phase curvatures. With the exception of only one subject, the phase data of the subjects show the largest maximum of the crest factor for a positive phase curvature, consistent with the literature.

The results show large differences across the subjects. Future studies may extend the experimental procedure to clarify to what extent this variability is due to procedural aspects. Also the analysis of the results could be extended by including some well-established auditory processing stages up to the binaural comparison stage. These future studies should then allow to determine to what extent envelope ITD might be an appropriate tool to measure the human auditory phase response.

References

- Bacon, S.P. (2006). "Auditory compression and hearing loss", Acoustics Today, Vol. 2, Issue 2, 30 -34.
- Bech, S., and Zacharov, N. (2006). "Perceptual audio evaluation - theory, method and application", (John Wiley & Sons Ltd, Chichester, England).
- Bernstein, L.R., and Trahiotis, C. (1985). "Lateralization of low-frequency, complex waveforms: the use of envelope-based temporal disparities", J. Acoustic. Soc. Am. 77, 1868 - 1880.
- Bernstein, L.R., and Trahiotis, C. (2002). "Enhancing sensitivity to interaural delays at high frequencies by using "transposed stimuli" ", J. Acoustic. Soc. Am. 112, 1026 - 1036.
- Bernstein, L.R., and Trahiotis, C. (2003). "Enhancing interaural-delay-based extents of laterality at high frequencies by using "transposed stimuli" ", J. Acoustic. Soc. Am. 113, 3335 - 3347.
- Bernstein, L.R., and Trahiotis, C. (2009). "How sensitivity to ongoing interaural temporal disparities is affected by manipulations of temporal features of the envelopes of high-frequency stimuli", J. Acoustic. Soc. Am. 125, 3234 - 3242.
- Bernstein, L.R., and Trahiotis, C. (2011). " Lateralization produced by envelope-based interaural temporal disparities of high-frequency, raised-sine stimuli: Empirical data and modeling", J. Acoustic. Soc. Am. 129, 1501 - 1508.
- Blauert, J. (1983) "Spatial hearing: the psychophysics of human sound localization", (MIT Press, Cambridge).
- Blodgett, H.C., Wilbanks, W.A., and Jeffress, L.A. (1956). "Effects of large interaural time differences upon the judgment of sideness", J. Acoustic. Soc. Am. 28, 639 - 643.
- Buell, T.N., and Hafter, E.R. (1988). "Discrimination of interaural differences of time in the envelopes of high-frequency signals: Integration times", J. Acoust. Soc. Am. 84, 2063 - 2066.

- Carlyon, R.P., and Datta, A.J. (1997). "Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system", J. Acoustic. Soc. Am. 101, 3636 - 3647.

- Carlyon, R.P., and Datta, A.J. (1997). "Masking period patterns of Schroeder-phase complexes: Effects of level, number of components, and phase of flanking components", J. Acoustic. Soc. Am. 101, 3648 - 3657.

- Columbia College Chicago, School of Media Arts, "Acousticlab" http://acousticslab.org/psychoacoustics/PMFiles/PMImages/placecodingb.jpg (page last called on 11.11.2014).

- Crankshaft Publishing, http://what-when-how.com/wp-content/uploads/2012/04/tmp15F73.jpg (page last called on 11.11.2014).

- Gelfand., S.A. (1997). "Essentials of Audiology", 2nd edition, (Thieme, New York).

- Glasberg, B. R., and Moore, B. C. (1990). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. 47, 103–138.

- Goldstein, J.L. (1967). "Auditory spectral filtering and monaural phase perception", J. Acoustic. Soc. Am. 41, 458 - 479.

- Greenwood, D.D. (1971). "Aural Combination Tones and Auditory Masking", J. Acoustic. Soc. Am. 50, 502 - 543.

- Grothe, B., Pecka, M., and McAlpine, D. (2010). "Mechanisms of Sound Localization in Mammals", Physiol. Rev. 90, 983 - 1012.

- Hartmann, W.M., and Rakerd, B. (1989). "On the minimum audible angle - A decision theory approach", J. Acoustic. Soc. Am. 85, 2031 - 2041.

- Hawkins, D.B., and Wightman, F.L. (1980). "Interaural Time Discrimination Ability of Listeners with Sensorineural Hearing Loss", Int. J. Audiol. 19, 495-507.

- Karas, J., and Savage, I.R. (1967). "Publications of Frank Wilcoxon (1892–1965)", Biometrics 23(1), 1 – 10.

- Klein-Hennig, M., Dietz, M., Hohmann, V., and Ewert, S. (2011). "The influence of different segments of the ongoing envelope on sensitivity to interaural time delays", J. Acoustic. Soc. Am. 129, 3856 - 3872.

- Kohlrausch, A. and Sander. A. (1995). "Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets", J. Acoustic. Soc. Am. 97, 1817 - 1829.

- Laback., B. (2010), "Psychoakustik II, Schwerpunkt: Experimentelle Audiologie", lecture notes for the course "Psychoakustik 2" at the University of Music and Performing Arts Graz.

- Laback, B., Zimmermann, I., Majdak, P., Baumgartner, W., and Pok, S. (2011). "Effects of envelope shape on interaural envelope delay sensitivity in acoustic and electric hearing", J. Acoustic. Soc. Am. 130, 1515 - 1529.

- Lacher-Fourgère, S., and Demany, L. (2005). "Consequences of cochlear damage for the detection of interaural phase differences", J. Acoustic. Soc. Am. 188, 2519 - 2526.

- Leek, M. R., and Summers, V. (1993). "The effect of temporal waveform shape on spectral discrimination by normal-hearing and hearing-impaired listeners", J. Acoustic. Soc. Am. 94, 2074 - 2082.

- Lentz, J.J., and Leek, M.R. (2001). "Psychophysical estimates of cochlear phase response: masking by harmonic complexes", J. Assoc. Res. Otolaryngol. 2, 408 - 422.

- Majdak, P., and Laback, B. (2009). "Effects of center frequency and rate on the sensitivity to interaural delay in high-frequency click trains", J. Acoustic. Soc. Am. 125, 3903 - 3913.

- Mathes, R. C., and Miller, R. L. (1947). ''Phase effects in monaural phase perception'', J. Acoustic. Soc. Am. 19, 780 – 797.

- Moore, B.C.J. (2012). "An introduction to the psychology of hearing", (Macmillan Press, London), 6th edition, published by Brill.

- Moore, B.C.J. (1996). "Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids", 3rd Int. Hearing Aid Conference 1995, University of Iowa.

- Mossop, J.E., and Culling, J.F. (1998). "Lateralization of large interaural delays", J. Acoustic. Soc. Am. 104, 1574 - 1579.

- Noel, V.A, Delgutte, B., and Eddington, D.K. (2013). "Lateralization and binaural fusion of long interaural time delays with bilateral cochlear implants", Conference on Implantable Auditory Prosthesis 2013, Granlibakken.

- Oxenham, A.J., and Dau, T. (2001). "Towards a measure of auditory-filter phase response", J. Acoustic. Soc. Am. 110, 3169 - 3178.

- Oxenham, A.J., and Dau, T. (2004). "Masker phase effects in normal-hearing and hearing-impaired listeners: evidence for peripheral compression at low signal frequencies", J. Acoustic. Soc. Am. 116, 2248 - 2257.

- Oxenham, A.J., and Ewert, S.D. (2005). "Estimates of auditory filter phase response at and below characteristic frequency (L)", J. Acoustic. Soc. Am. 117 , 1713 - 1716.

- Rasumow, E. (2009). "Messung und Simulation der Maskierung von harmonischen Tonkomplexen mit Schroederphasen auf Sinustönen", Master thesis at Universität Oldenburg, Inst. f. Hörtechnik und Audiologie.

- Ruggero, M.A., and Rich, N.C. (1991). "Furosemide Alters Organ of Corti Mechanics: Evidence for Feedback of Outer Hair Cells upon the Basilar Membrane", J. Neurosci. 11, 1057–1067.

- Saberi, K., and Antonio, J.V. (2003). "Precedence-effect thresholds for a population of untrained listeners as a function of stimulus intensity and interclick interval", J. Acoustic. Soc. Am. 144, 420 - 429.

- Schroeder, M.R. (1970). "Synthesis of low-peak-factor signals and binary sequences with low autocorrelation", IEEE Trans. Inf. Theory 16, 85 - 89.

- Shen, Y. and Lentz, J.J. (2009), "Level dependence in behavioral measurements of auditory-filter phase characteristics", J. Acoustic. Soc. Am. 126, 2501 - 2510.

- Shera, C.A. (2001). "Frequency glides in click responses of the basilar membrane and auditory nerve: Their scaling behavior and origin in traveling-wave dispersion", J. Acoustic. Soc. Am. 109, 2023 - 2034.

- Smith, B.K., Sieben, U.K., Kohlrausch, A., and Schroeder, M.R. (1986). "Phase effects in masking related to dispersion in the inner ear", J. Acoustic. Soc. Am. 80, 1631 - 1637.

- Summers, V., and Leek, M.R. (1998). "Masking of tones and speech by Schroeder-phase harmonic complexes in normally hearing and hearing-impaired listeners", Hear. Res. 118, 139 - 150.

- Summers, V. (2000). "Effects of hearing impairment and presentation level on masking period patterns for Schroeder-phase harmonic complexes", J. Acoustic. Soc. Am. 108, 2307 - 2317.

- Summers, V., de Boer, E., Nuttall, A.L. (2003) . "Basilar-membrane responses to multicomponent (Schroeder-phase) signals: Understanding intensity effects", J. Acoust. Soc. Am. 114, 294 - 306.

- University of Minnesota Duluth, Lectures "Sensory Physiology" http://www.d.umn.edu/~jfitzake/Lectures/UndergradPharmacy/SensoryPhysiology /Audition/Figures/CochleaStructure.jpg (page last called on 11.11.2014).

- Yost, W.A., and Hafter, E.R. (1987). Chapter "Lateralization" in "Directional Hearing", edited by W.A. Yost and G. Gourevitch (Springer, New York).

- Zilany, M.S, Bruce, I.C., Nelson, P.C., and Carney, L.H. (2009). "A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics.", J. Acoustic. Soc. Am. 126, 2390 - 2412.

- Zwislocki, J., and Feldman, R.D. (1956). "Just noticeable differences in dichotic phase", J. Acoustic. Soc. Am. 28, 860-864.

List of Figures

Appendix

Part A: Audiogram of each subject

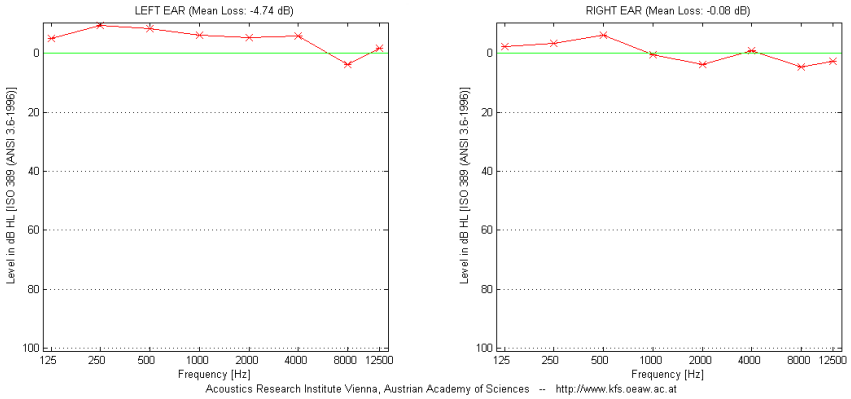


Figure 8.1: Audiogram NH39 from 21.12.2010

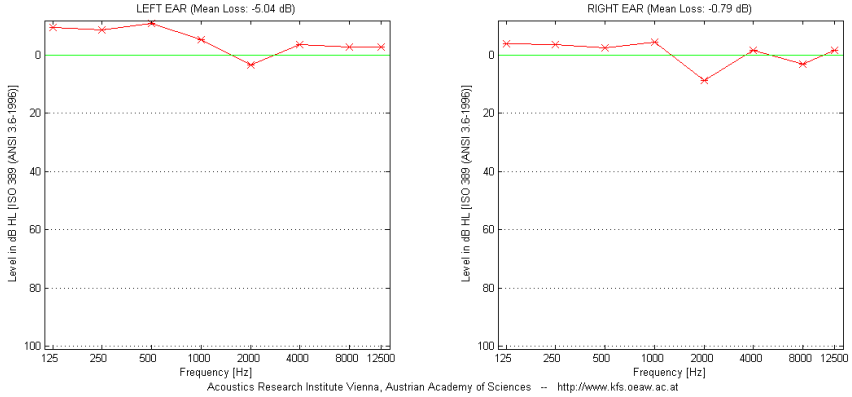


Figure 8.2: Audiogram NH43 from 17.12.2010

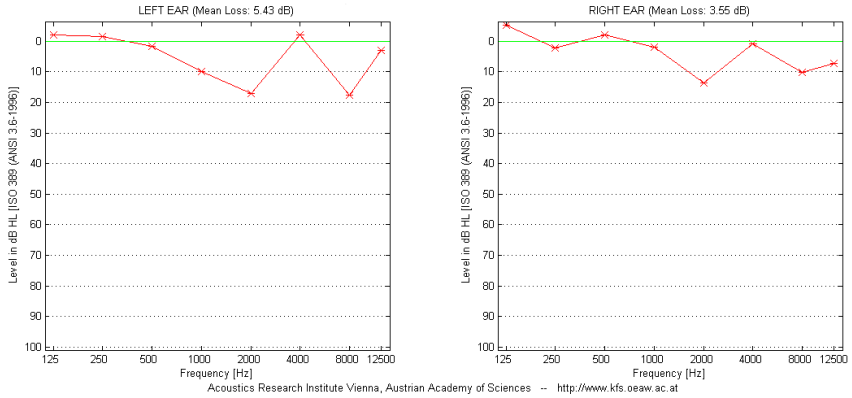


Figure 8.3: Audiogram NH57 from 26.08.2009

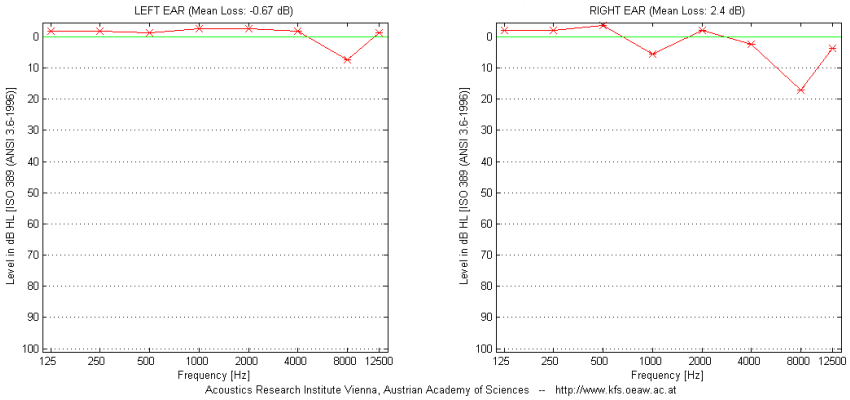


Figure 8.4: Audiogram NH62 from 10.09.2009

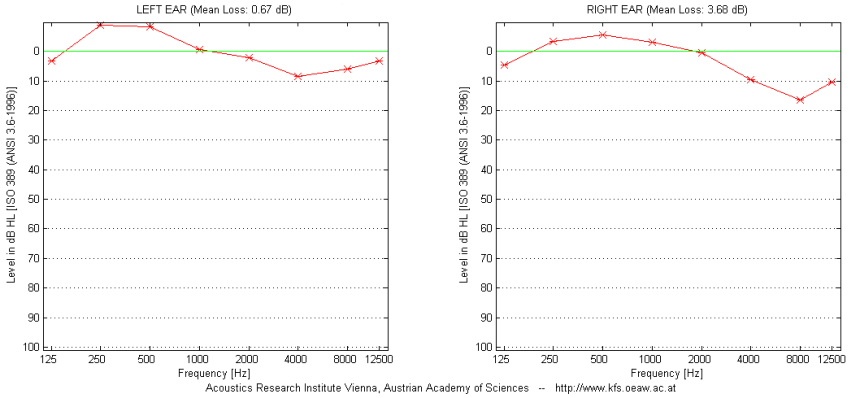


Figure 8.5: Audiogram NH66 from 16.09.2009

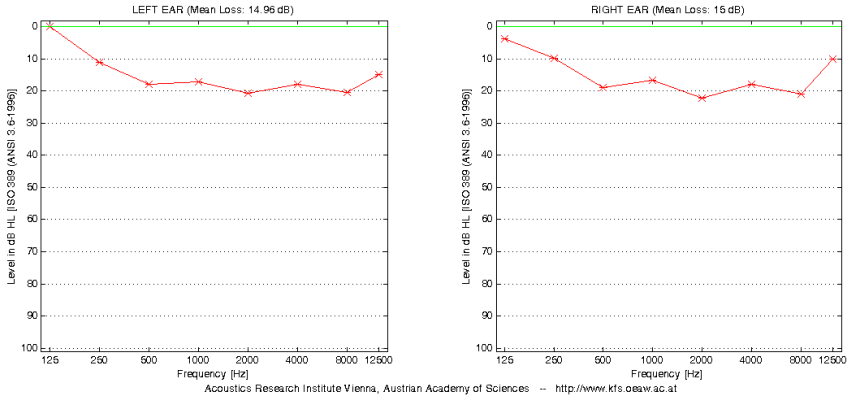


Figure 8.6: Audiogram NH83 from 19.05.2014

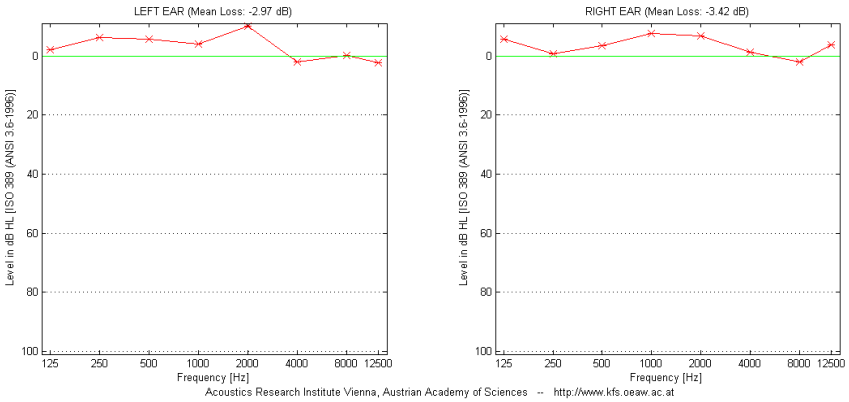


Figure 8.7: Audiogram NH144 from 20.05.2014

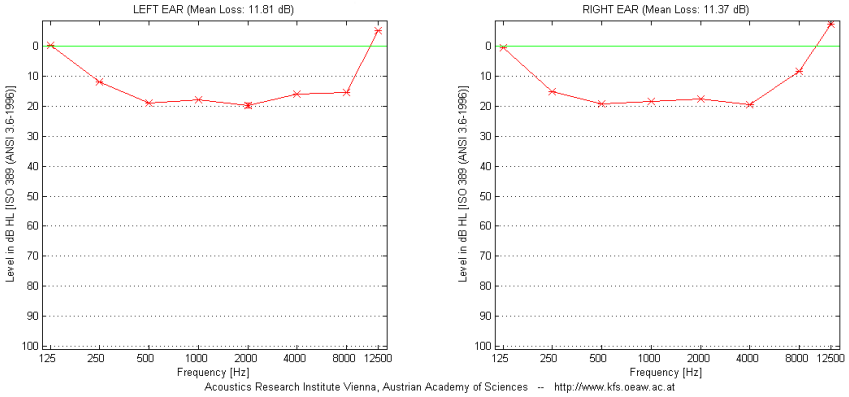


Figure 8.8: Audiogram NH171 from 14.05.2014

Part B: Instructions for the subjects

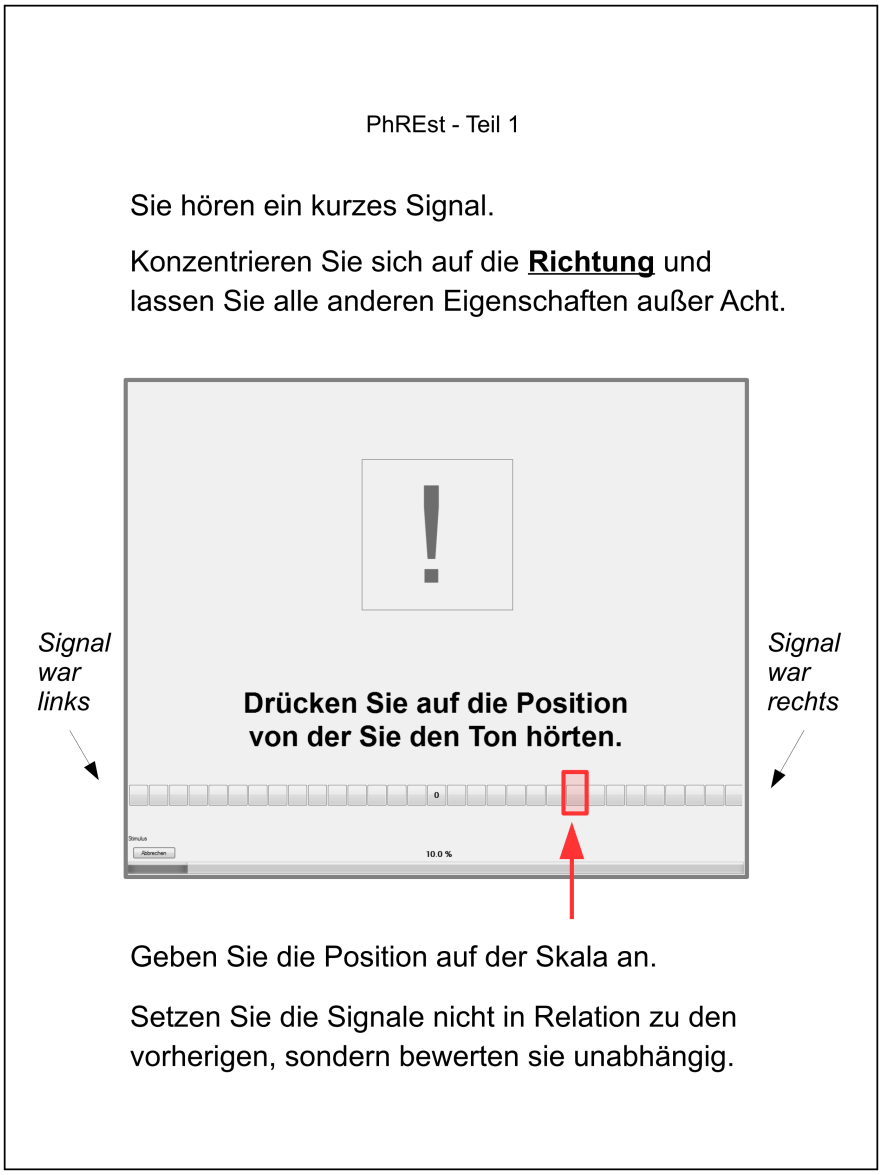


Figure 8.9: Instruction page of the magnitude estimation task (used in stage 1 and stage 2).

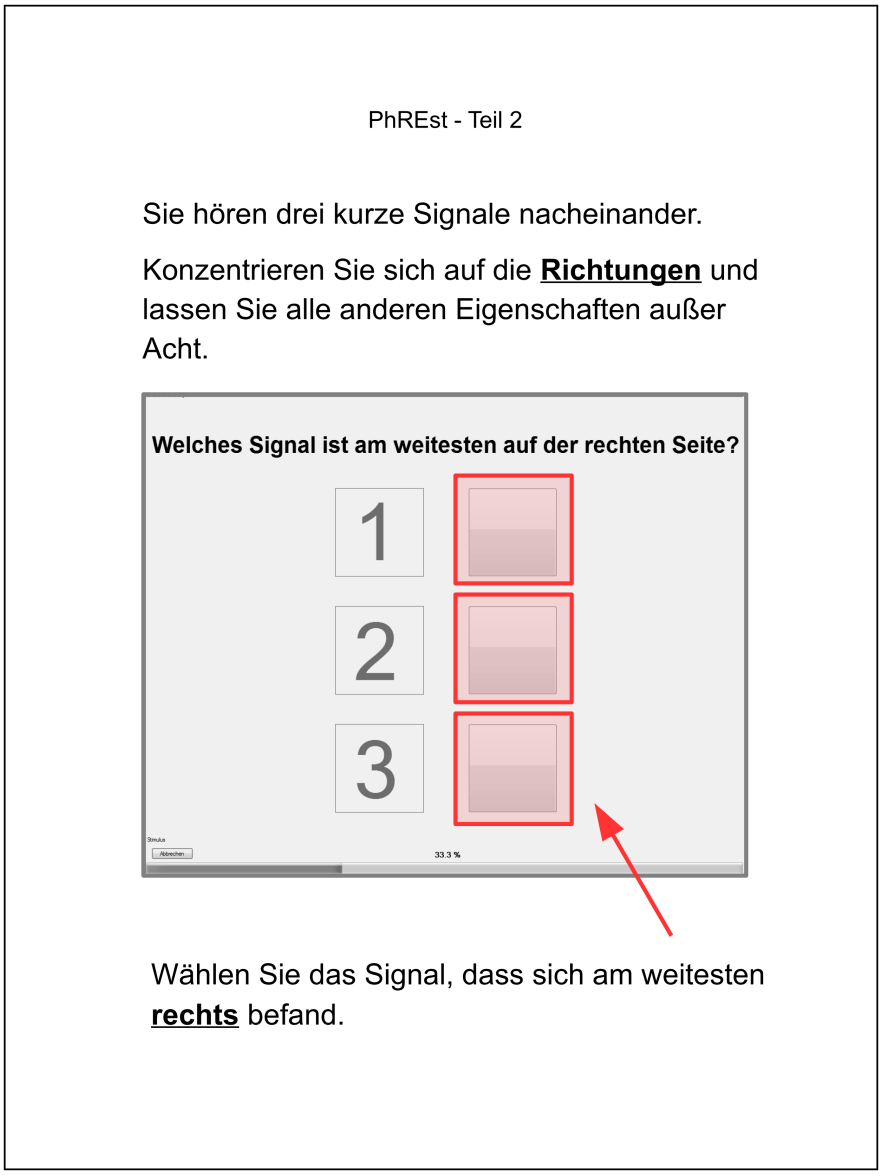


Figure 8.10: Instruction page of the ranking task (used only in stage 1).

Part C: Lateralization plots for all subjects

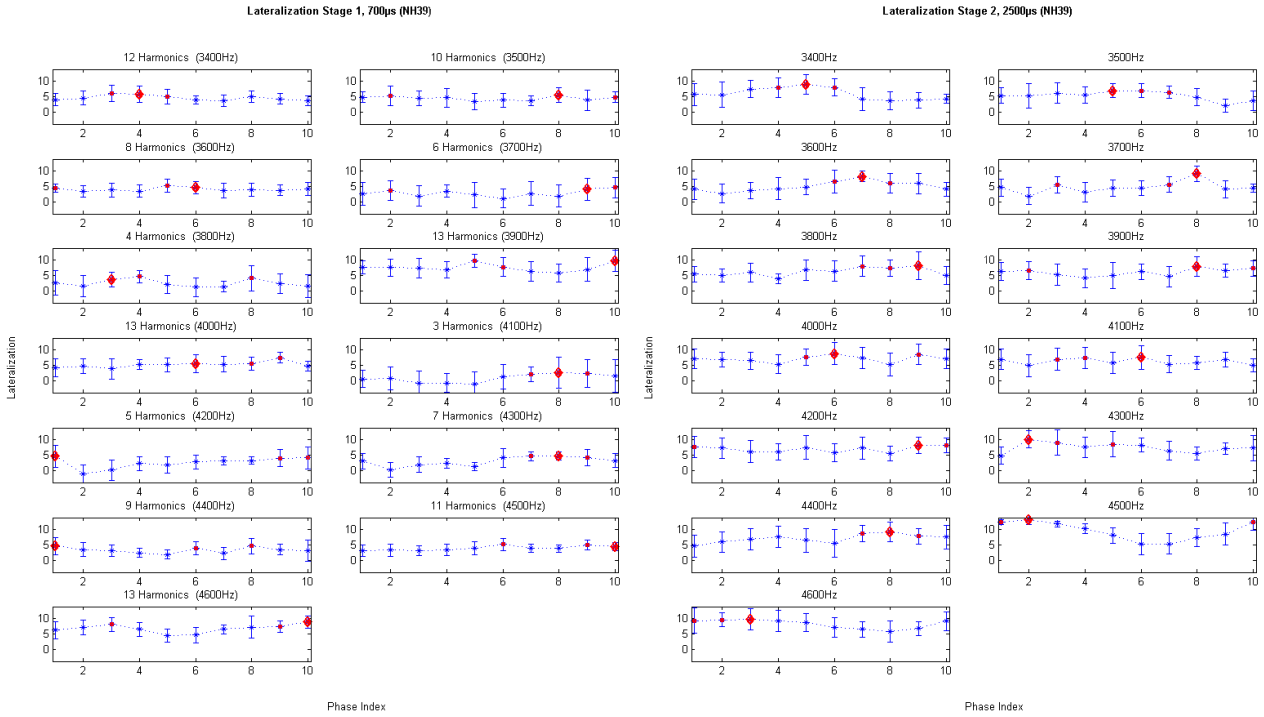


Figure 8.11: Lateralization values of both stages for NH39

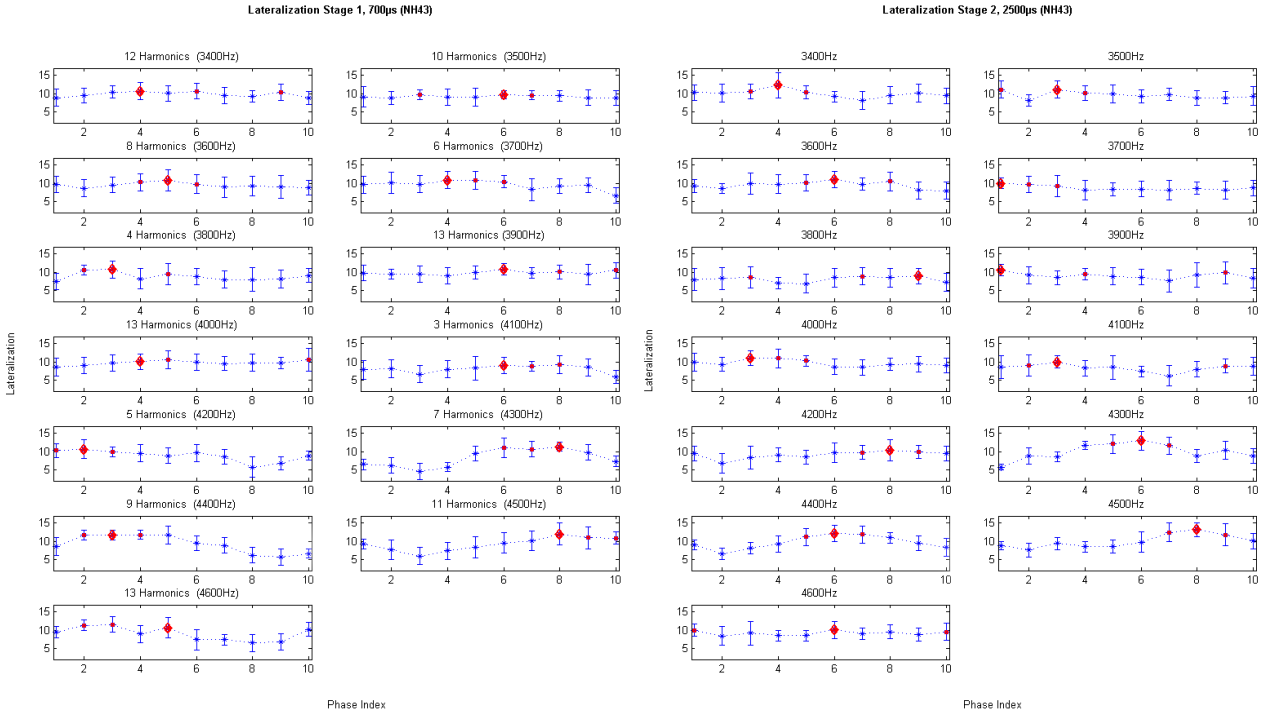


Figure 8.12: Lateralization values of both stages for NH43

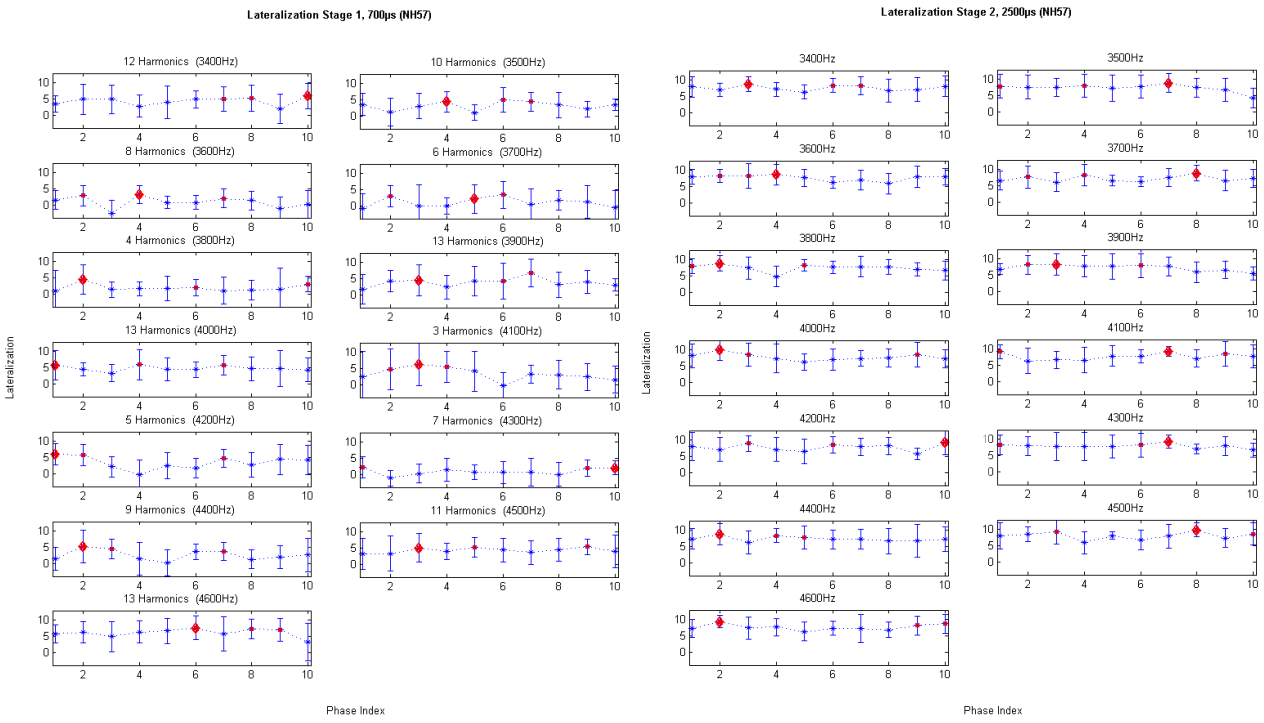


Figure 8.13: Lateralization values of both stages for NH57

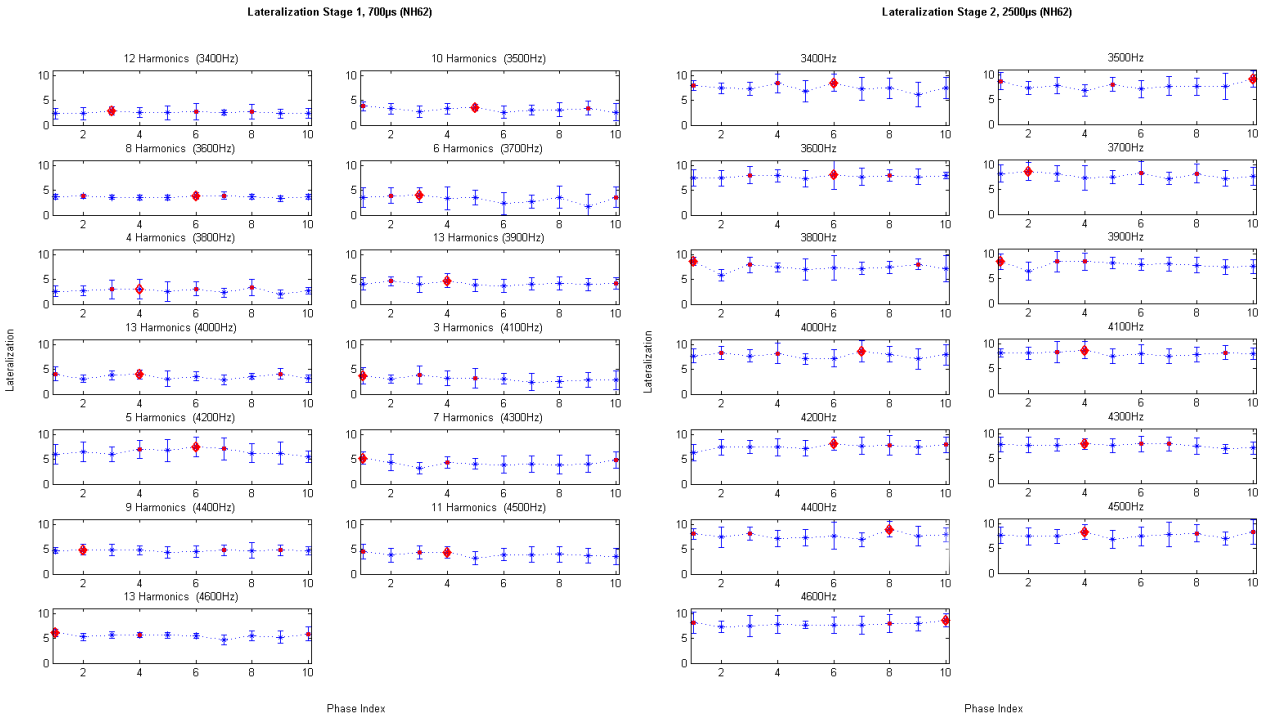


Figure 8.14: Lateralization values of both stages for NH62

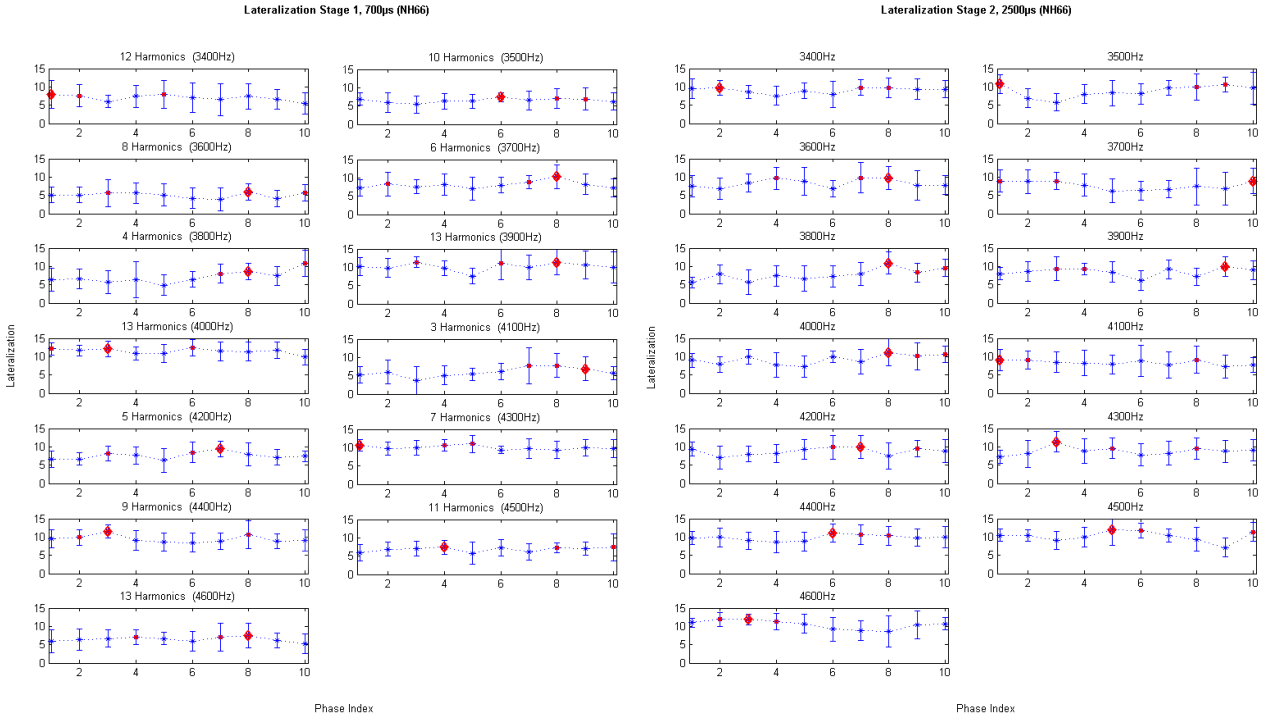


Figure 8.15: Lateralization values of both stages for NH66

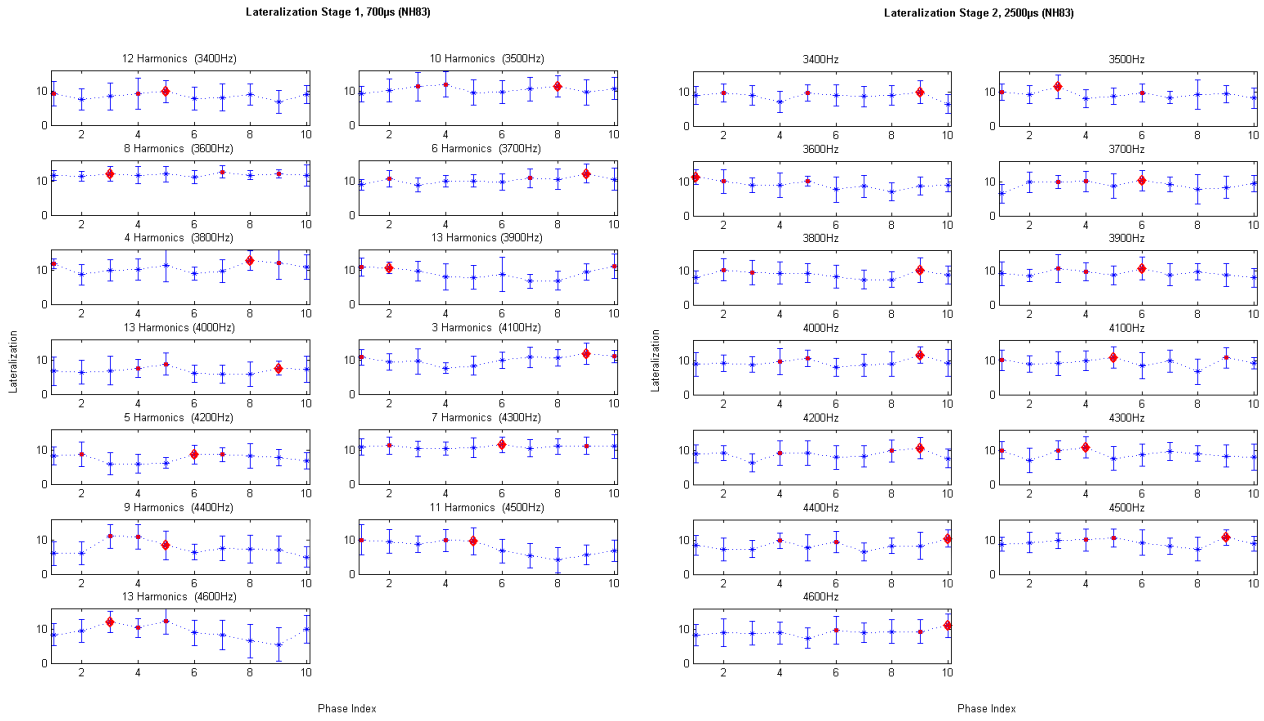


Figure 8.16: Lateralization values of both stages for NH83

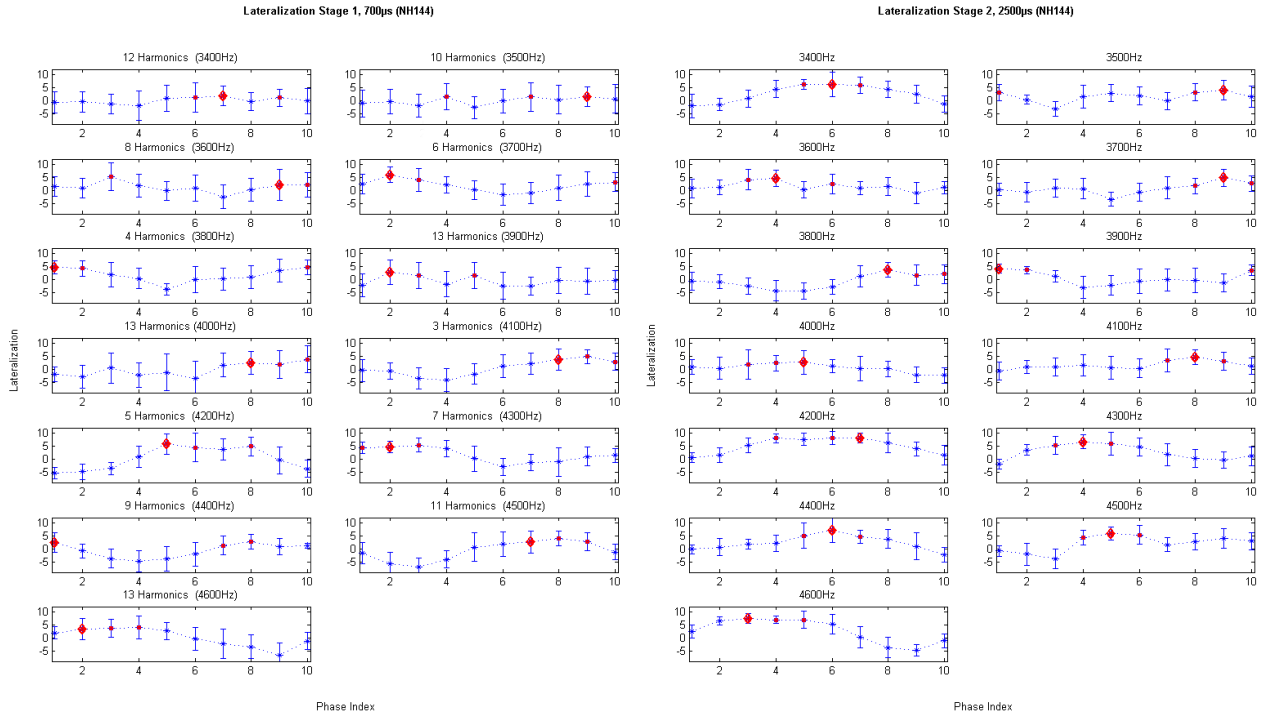


Figure 8.17: Lateralization values of both stages for NH144

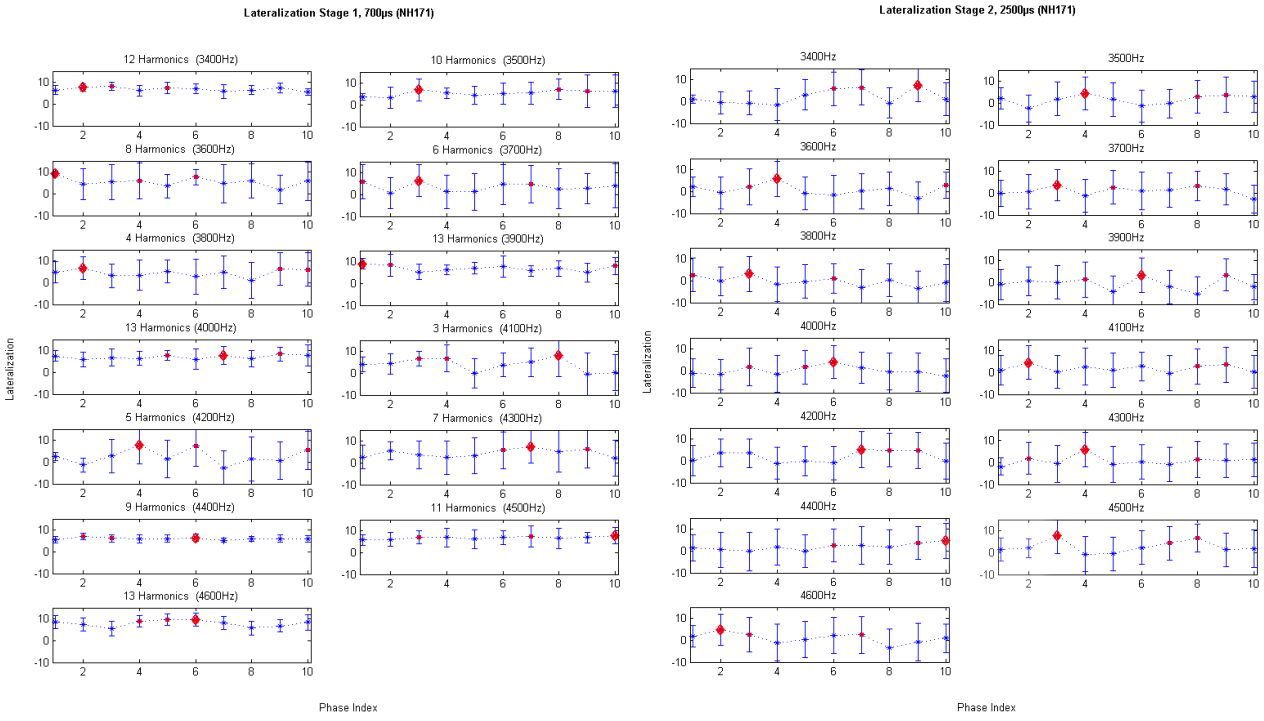


Figure 8.18: Lateralization values of both stages for NH171

Part D: Significance plots for all subjects

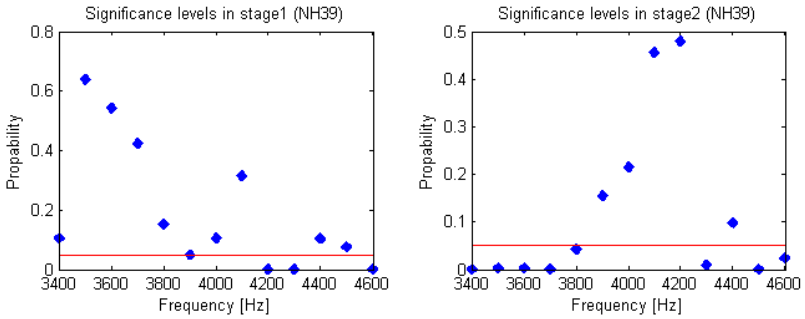


Figure 8.19: Significance levels for both stages for NH39

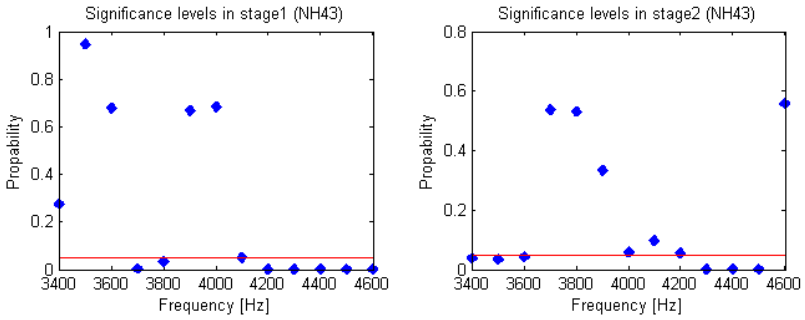


Figure 8.20: Significance levels for both stages for NH43

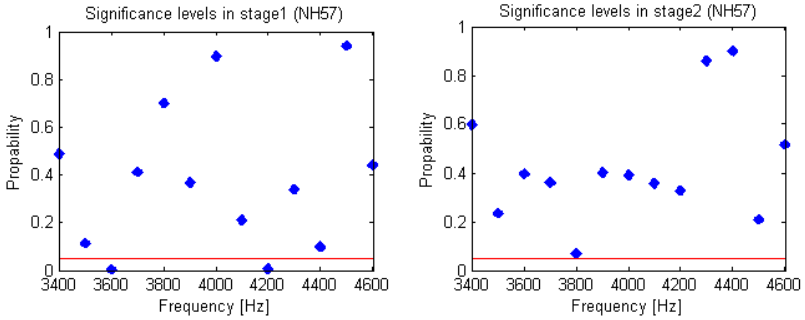


Figure 8.21: Significance levels for both stages for NH57

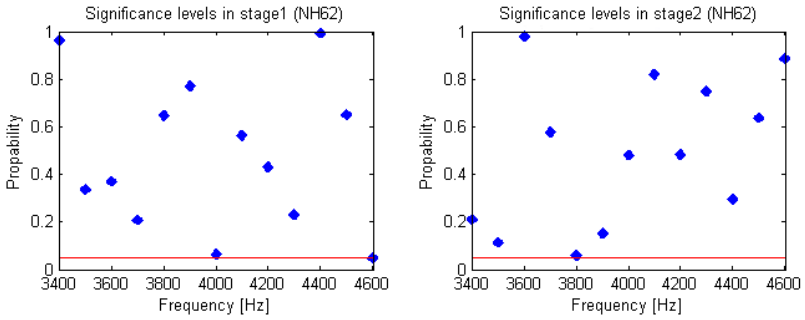


Figure 8.22: Significance levels for both stages for NH62

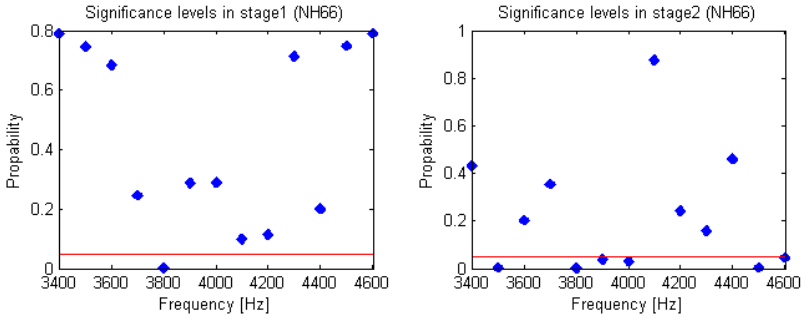


Figure 8.23: Significance levels for both stages for NH66

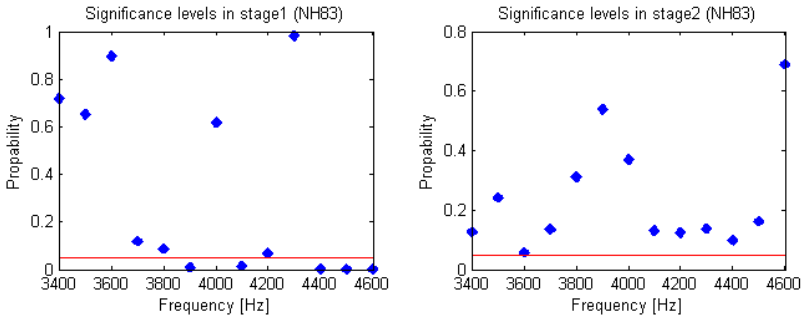


Figure 8.24: Significance levels for both stages for NH83

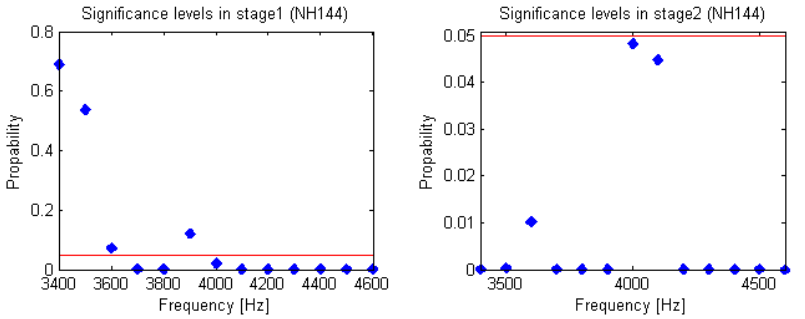


Figure 8.25: Significance levels for both stages for NH144

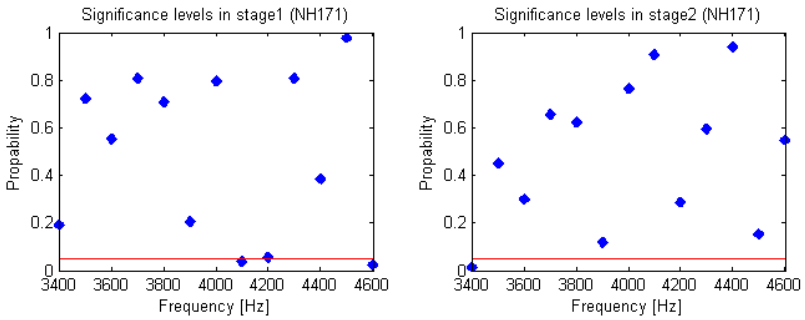


Figure 8.26: Significance levels for both stages for NH171

Part E: Relative phase values for all subjects

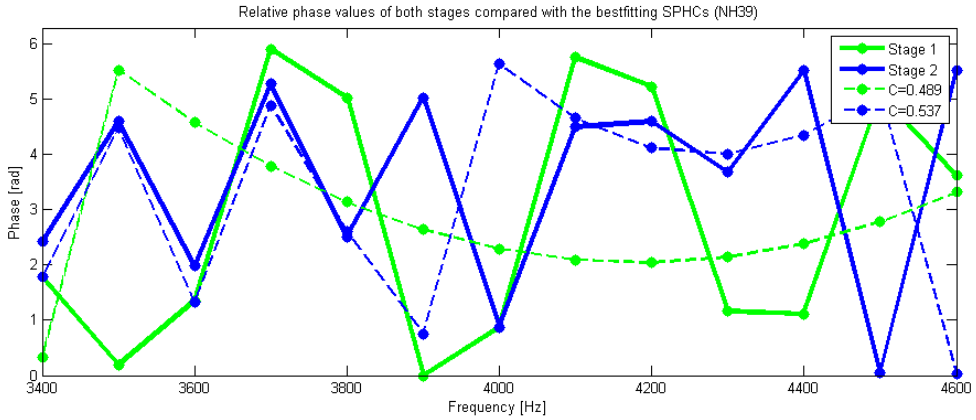


Figure 8.27: Relative phase values and the bestfitting SPHC in the distance method for both stages for NH39

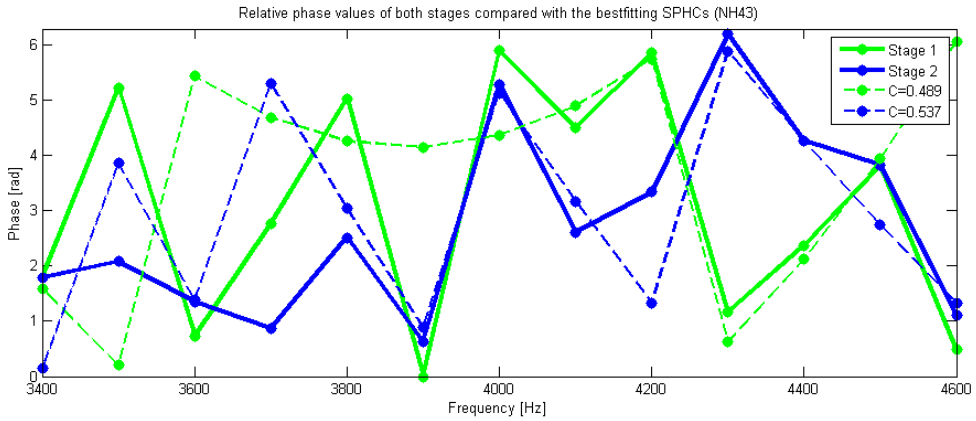


Figure 8.28: Relative phase values and the bestfitting SPHC in the distance method for both stages for NH43

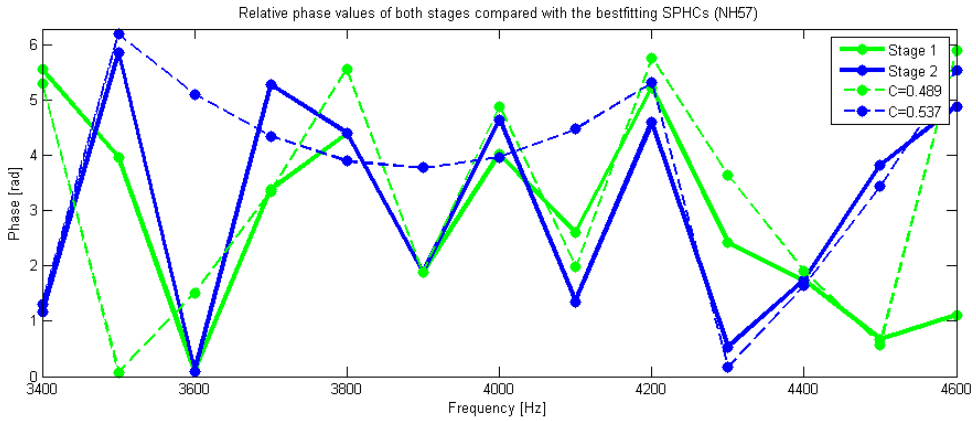


Figure 8.29: Relative phase values and the bestfitting SPHC in the distance method for both stages for NH57

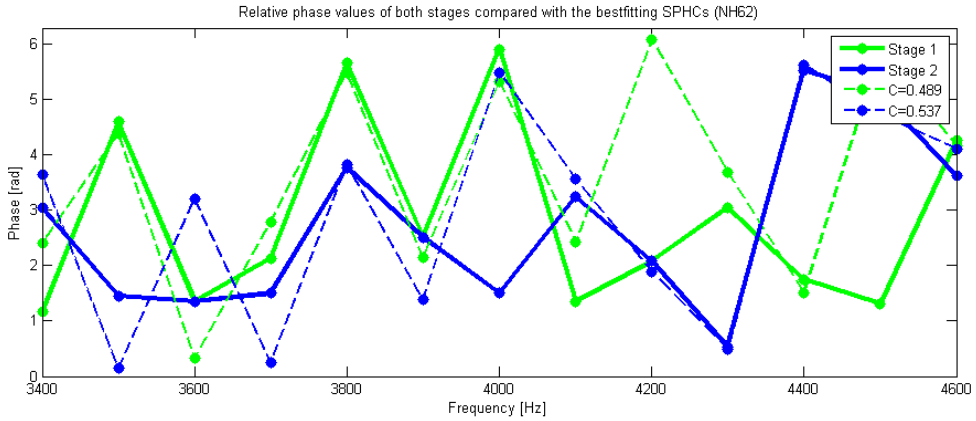


Figure 8.30: Relative phase values and the bestfitting SPHC in the distance method for both stages for NH62

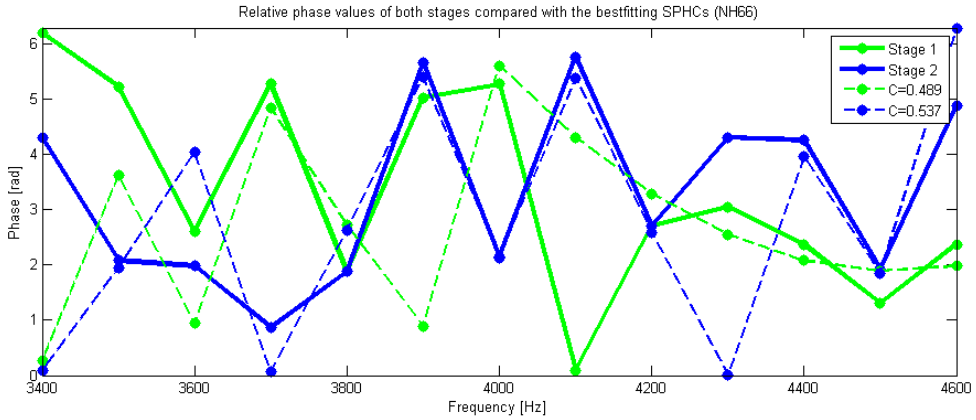


Figure 8.31: Relative phase values and the bestfitting SPHC in the distance method for both stages for NH66

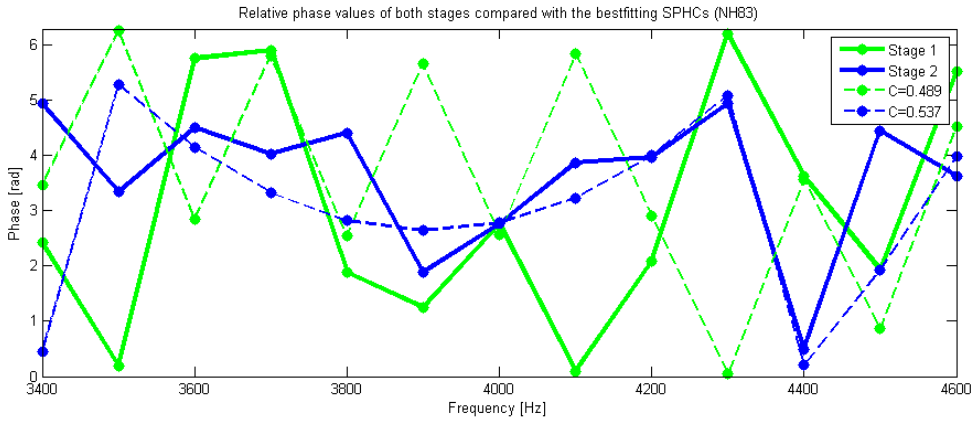


Figure 8.32: Relative phase values and the bestfitting SPHC in the distance method for both stages for NH83

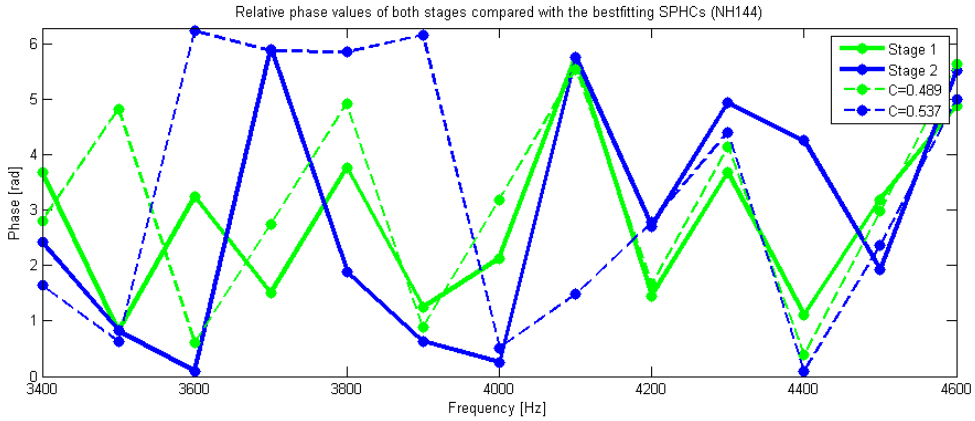


Figure 8.33: Relative phase values and the bestfitting SPHC in the distance method for both stages for NH144

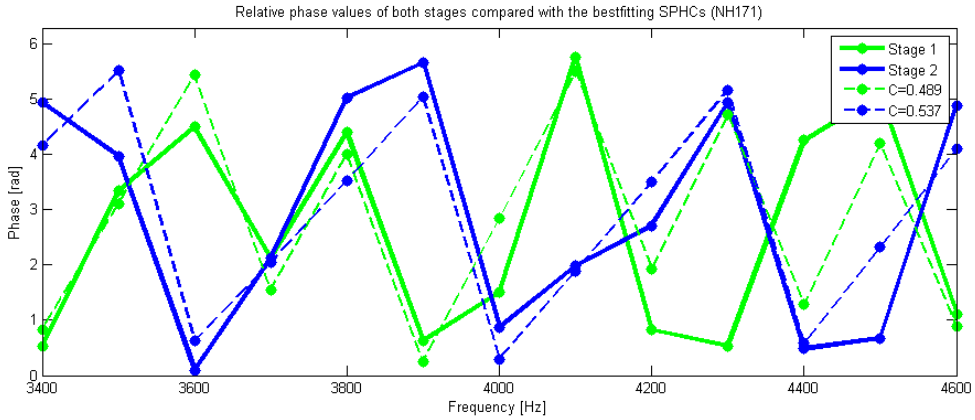


Figure 8.34: Relative phase values and the bestfitting SPHC in the distance method for both stages for NH171

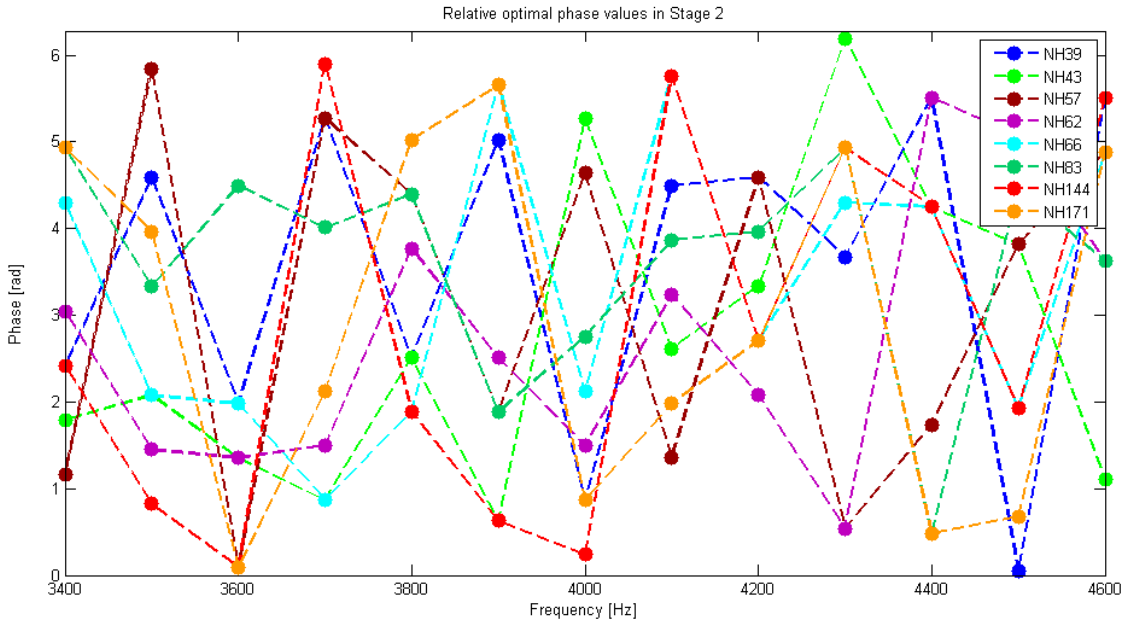


Figure 8.35: Relative phase values of stage 2 for all subjects

Part F: Crest factors of filtered signals for all subjects

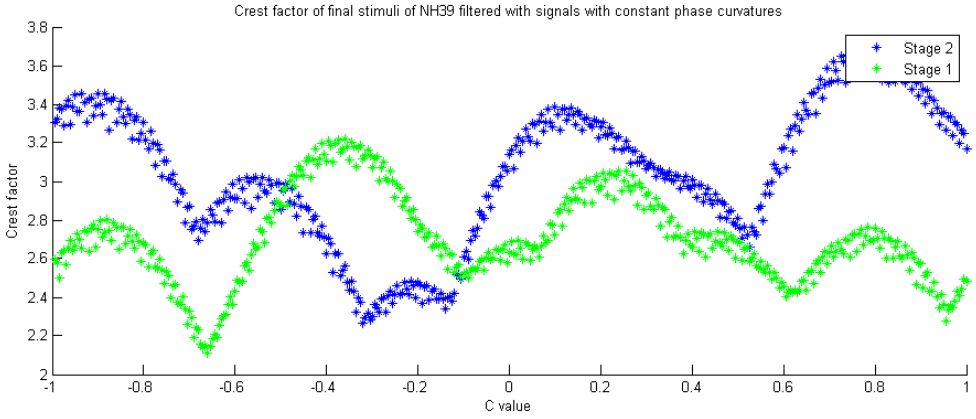


Figure 8.36: Crest factors NH39

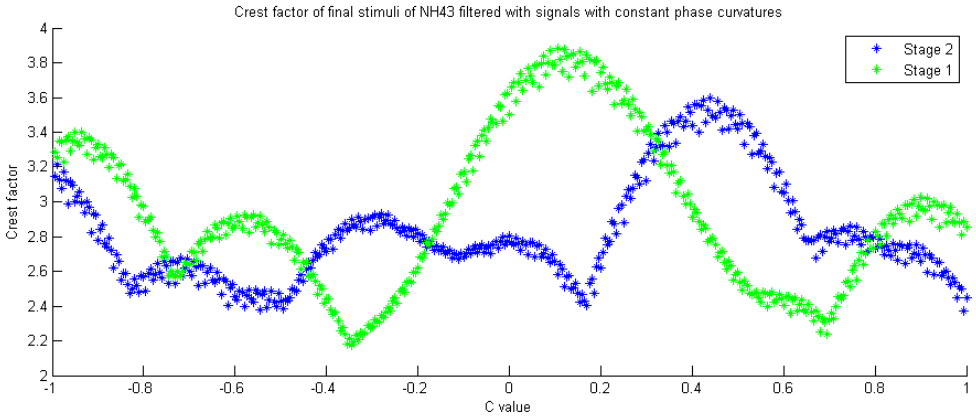


Figure 8.37: Crest factors NH43

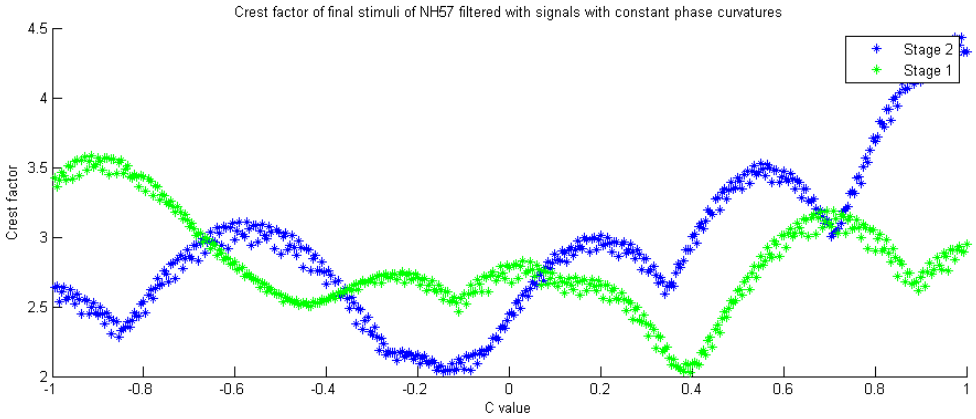


Figure 8.38: Crest factors NH57

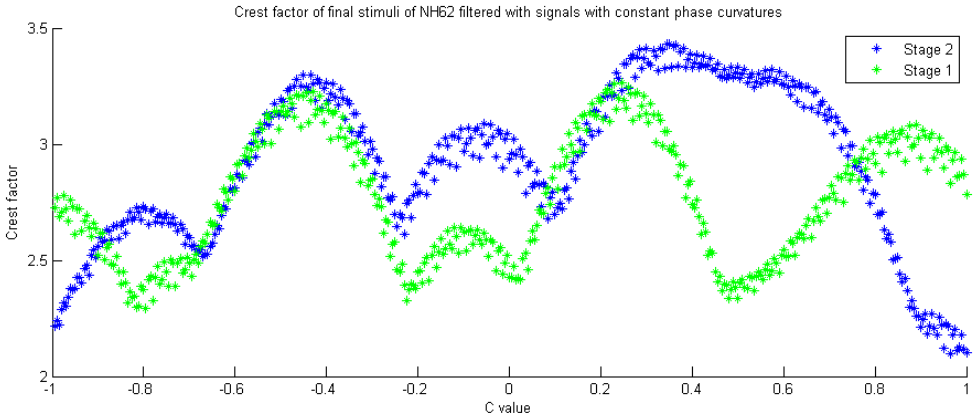


Figure 8.39: Crest factors NH62

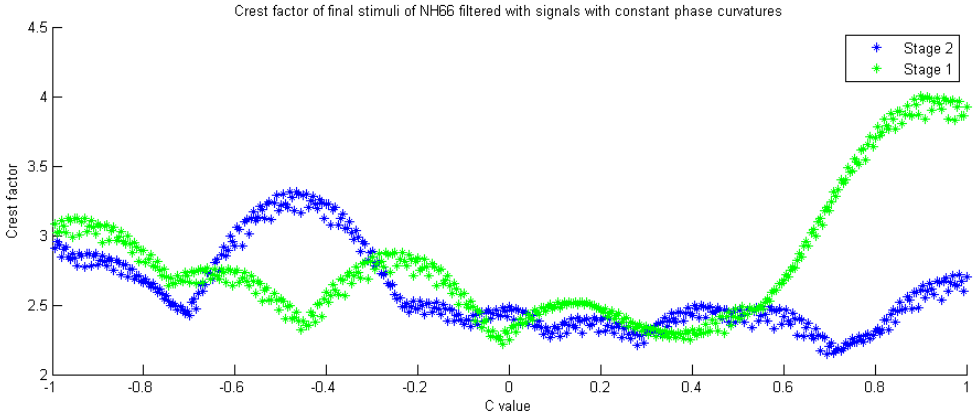


Figure 8.40: Crest factors NH66

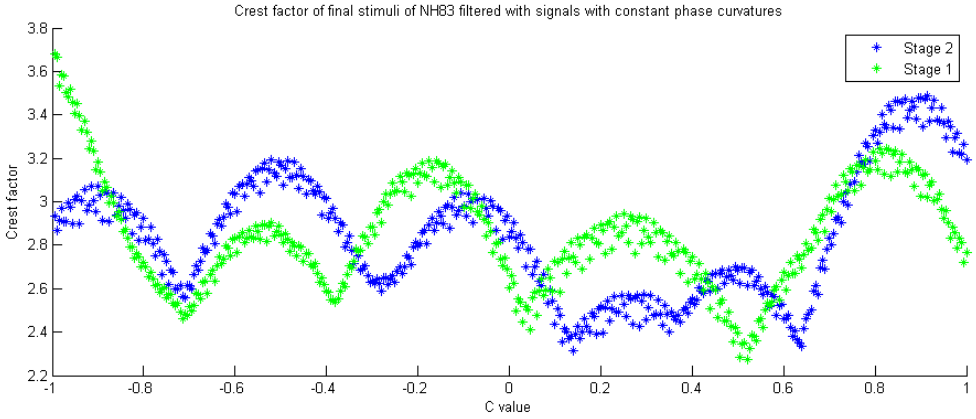


Figure 8.41: Crest factors NH83

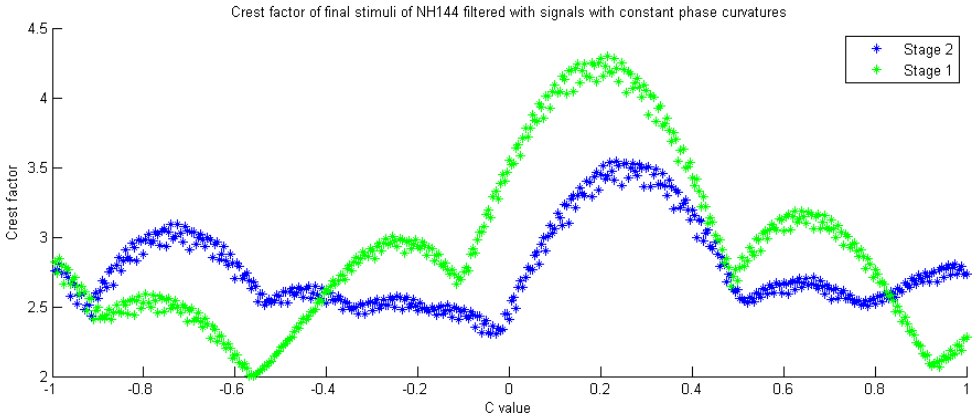


Figure 8.42: Crest factors NH144

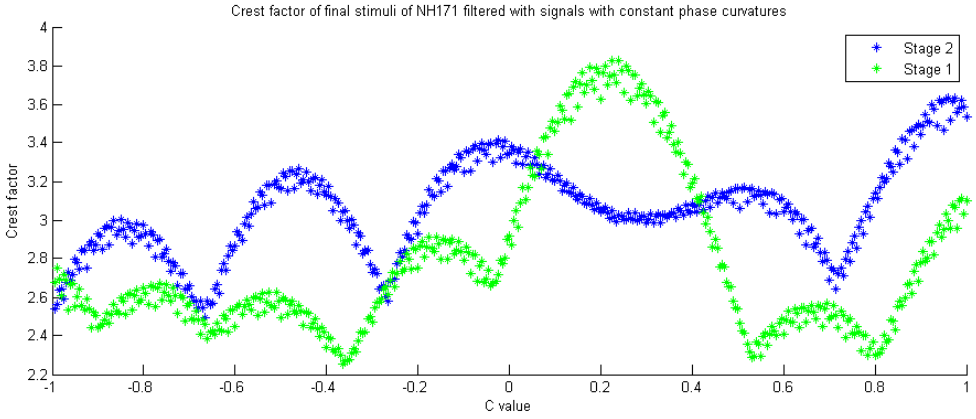


Figure 8.43: Crest factors NH171