Implementation and evaluation of auditory models for sound localization

Student Project

Katharina Egger

Graz, January 9, 2013

Supervisor: Dr. Piotr Majdak

Abstract

For sound localization the human auditory system uses binaural information like interaural time difference (ITD) and level difference (ILD) as well as spectral cues. The information can be described by individual head-related transfer-functions, HRTFs. They represent the spectral characteristics of the filtering process of an incoming soundwave by head, pinna and torso.
The processing of this information in the auditory system still cannot be fully explained. Hence the influence of several parameters on human perception is studied with the help of psychoacoustic experiments. However, big variation in system parameters results in great expenses of time and money. Therefore auditory models which aim in modelling response patterns of human listeners are used.
The aim of the project is to combine two existing auditory models (Langendijk and Bronkhorst, 2002 and Lopez-Poveda and Meddis, 2001) and integrate the resulting one in the Auditory Modelling Toolbox (AMT) for MATLAB. The resulting model has been evaluated with already existent data from various sound localization experiments.

Zusammenfassung

Für die Schallquellenlokalisation nutzt das auditorische System binaurale Information wie interaurale Laufzeitdifferenz (ITD) und Pegeldifferenz (ILD) sowie auch monaurale spektrale Information. Sie können durch die individuellen Außenohrübertragungsfunktionen (engl: head-related transfer-functions, HRTFs) beschrieben werden. Sie zeigen die spektralen Eigenschaften der Filterung des eintreffenden Schalls aufgrund von Kopf, Ohrmuschel und Rumpf.
Da die Verarbeitung dieser Informationen im auditorischen System noch immer nicht zur Gänze erforscht ist, wird der Einfluss verschiedenster Parameter auf die menschliche Wahrnehmung anhand von psychoakustischen Tests untersucht. Eine große Variation an Systemparametern ist jedoch zeit- und kostenintensiv. Dies führt zum Einsatz von Modellen des menschlichen auditorischen Systems, deren Ziel ist, die Ergebnisse der Probanden vorhersagen zu können.
Ziel des Projekts ist es zwei bestehende Modelle (Langendijk and Bronkhorst, 2002 and Lopez-Poveda and Meddis, 2001) zu kombinieren und in die Auditory Modelling Toolbox (AMT) für MATLAB zu integrieren. Das resultierende Modell soll mit Hilfe von vorhandenen Ergebnissen diverser Lokalisationsversuche evaluiert und konfiguriert werden.

Contents

1 Motivation

A sound is processed by several stages in the human auditory system in order to define its localization. For sound localization the human auditory system uses binaural information like interaural time difference (ITD) and level difference (ILD) as well as spectral cues, which represent the spectral characteristics of the filtering process of an incoming sound wave by head, pinna and torso.

The processing of this information in the auditory system still cannot be fully explained. Hence the influence of several parameters on human perception is studied with the help of psychoacoustic experiments. However, big variation in system parameters results in great expenses of time and money. Therefore auditory models which reproduce response patterns of human listeners are used.

Thus the long-term goal is to have an auditory model for three dimensional sound localization. There is a model which was introduced by Langendijk and Bronkhorst (2002) whose simulation is based on spectral cues in the median plane. However, it does not map the physiology of the human ear. So the aim of the project is to take this model as a basis for modeling localization and adapt it to a more human-like processing physiology. Furthermore the resulting scripts and implementations will be contributed to the Auditory Modeling Toolbox (AMT) for MATLAB. The implemented parts will be validated and the resulting new model will be evaluated and validated with already existent psychoacoustical data.

2 The Human Auditory System

This chapter gives a short overview of the structure and function of the auditory processing system. The description is based on Laback (2010) and Larsen (2010). It presents a brief insight in the physiology, as far as it is necessary to follow the subsequent chapters. For a more detailed explanation please refer to related literature.

Figure 1 shows the anatomy of the peripherical auditory system. It consists of outer, middle and inner ear which will be explained further below.

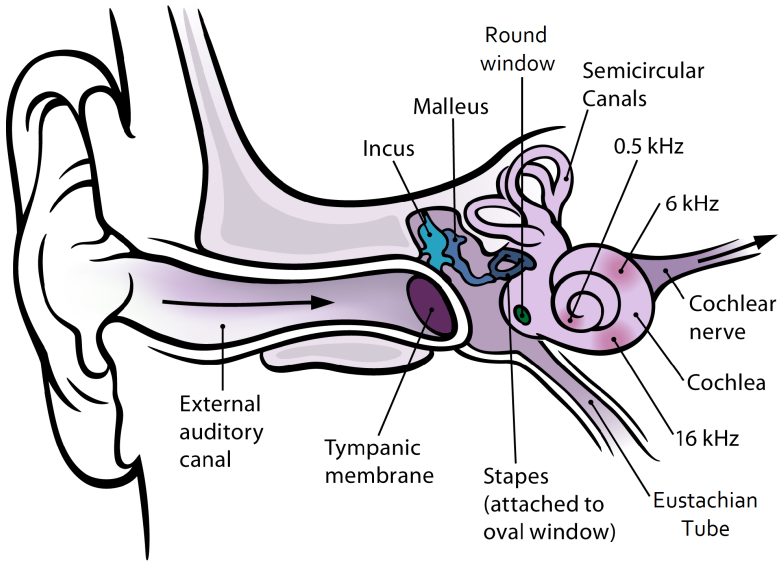

Figure 1: Anatomy of the human ear, modified from Chittka and Brockmann (2005)

2.1 Outer- and Middle Ear

The outer ear consists of the pinna and the ear canal. Sound waves enter the ear directly and via reflections of pinna, head, and torso. The pinna behaves as an acoustic cone which focuses the sound waves at higher frequencies. Due to its inhomogeneous and very inter-individually shaped structure the sound spectrum is modified which enables sound localization in sagittal planes. The pinna and the slightly crooked ear canal together form a resonance between 1.5 and 5 kHz with its maximum around 2.5 kHz.

The middle ear lies in the air-filled tympanic cavity which is separated from the outer ear by the tympanic membrane, or also called eardrum. In order to equalize the pressure between the middle ear and the atmosphere the cavity is connected to the nasal cavity via the Eustachian tube. The middle ear contains three very small ossicles, malleus (hammer), incus (anvil) and stapes (stirrup), which are connected to the eardrum. These delicate bones amplify and pass on the vibration caused by the incoming sound wave to the inner ear via the oval window. The main function of the middle ear is the impedance-matching mechanism between air and the fluid-filled cochlea in the inner ear, otherwise the incoming sound would be reflected. The middle ear behaves like a bandpass filter,

the most efficient transmission of sound happens in the frequency range from about 0.5 to 4 kHz.

2.2 Inner Ear

A spiral structure which is called cochlea, together with the vestibular balance system are known as the inner ear.
As shown in figure 2 the cochlear consists of three chambers (or scalae): Scala Vestibuli, which abuts the oval window, Scala Tympani, which is connected to the Scala Vestibuli with the so called helicotrema (apex) and terminates at the round window and third, the Scala Media. The latter lies in between the two other chambers, separated by the Reissner's membrane and the basilar membrane (BM). The BM contains the organ of Corti (OC) with the sensory hair cells. Scala Vestibuli and Scala Tympani are filled with perilymph whereas Scala Media is filled with endolymph. The potential difference between these two types of liquids is applied at the BM to further convert mechanical oscillation into electrical impulses.

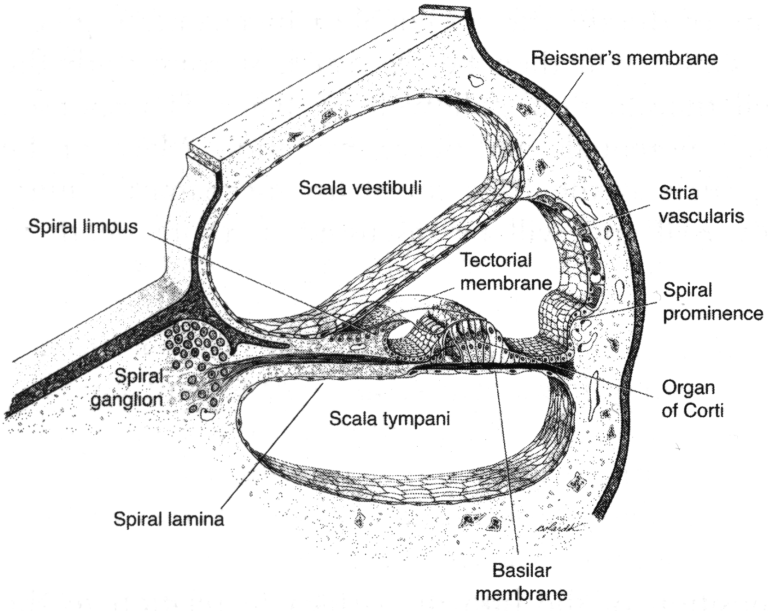

Figure 2: Cross section of the cochlear, from Pickles (2008)

Mechanical vibration of the oval window induces movement of the fluids in the cochlea which in turn makes the BM vibrate. The motion propagates itself as a traveling wave along the membrane where the frequency of a sound gives rise to a place-specific response in the BM (von Békésy, 1949). The BM appears to change in width and stiffness which is why the movement reaches its maximum at a certain resonance point on the BM after which the movement rapidly declines. High frequency sounds (up to 20 kHz) result in peak responses at the base of the BM whereas the thinner and less stiffer apical parts of the BM respond maximally to low frequency sounds (around 20 Hz). So for a certain

position on the BM, the frequency which elicits peak response is referred to as the center frequency (CF). The spacing of CFs along the BM is roughly proportional to the logarithm of the frequency. The BM does not have an endless frequency resolution. The bandwidth of frequencies around a given CF that cannot be resolved is known as critical band. The critical bands also increase with frequency resulting in a higher resolution for lower frequencies.

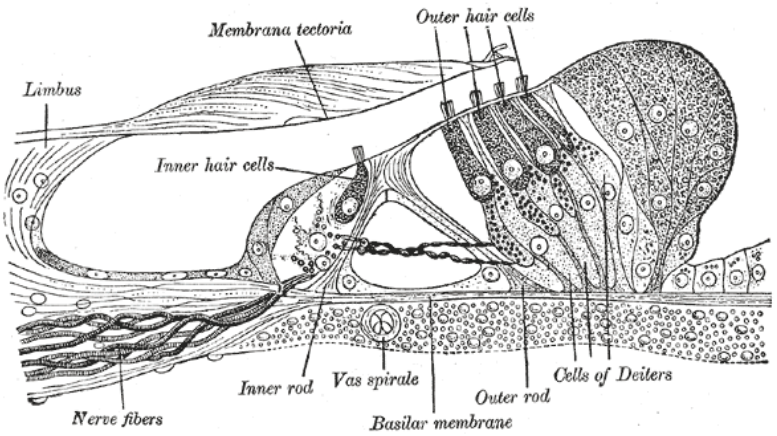

Figure 3: Section through the spiral organ of Corti, from Gray (1918)

In the OC the vibration of the BM is passed on to the hair cells (figure 3). Deflection of the hair cells induce neuronal impulses. These spikes are elicited of hair cell movement only in one direction which results in an halfwave rectified oscillation. There are two anatomically and functionally distinct types of hair cells: inner and outer hair cells. Roughly speaking the difference is, contrary to the inner hair cells (IHCs) the outer hair cells (OHCs) can be controlled actively by the brain stem.

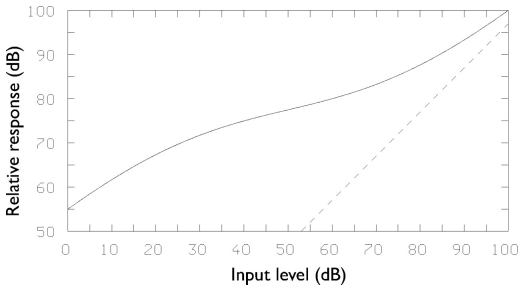

Figure 4: Schematic illustration of an input-output function of the BM for a tone with frequency close to CF (solid line); linear input-output function (dashed line, typical of what might be observed for a tone with frequency well below CF), from Moore (2002)

Physiological studies indicate a highly compressive input-output function of the BM (Moore, 2002). Figure 4 shows an schematic plot of the magnitude of BM response as a function of the magnitude of input level. The BM response is highly compressive as a function of stimulus level at frequencies close to the CF (solid line) whereas for frequencies below or above the CF the response becomes linear. This nonlinear response

maximizes the sensitivity at the CF. The increase of sensitivity for CF-near frequencies results also in broadened critical bands. The OHCs and their active contribution are thought to be the reason for this nonlinear compressive behavior. Their functionality is still not fully understood, however it depends on a complex interplay within the auditory system.

In auditory models the BM is typically represented by a bank of bandpass filters with bandwidths equal to the critical bands. Most frequently linear filters, such as a gammatone filter bank are used. However, it is obvious that linear filters cannot sufficiently represent the nonlinear properties of the BM. To describe effects like for example the above mentioned mechanism of level dependent broadening of the critical bands, it is essential to use a nonlinear filter bank.

3 The Auditory Modelling Toolbox (AMT)

The Auditory Modelling Toolbox (AMT) is a Matlab toolbox for developing and applying auditory perceptual models (Søndergaard *et al.*, 2011). It is released under a free software license, the GNU Public License version 3. Therefore it is free to download and use (`http://amtoolbox.sourceforge.net`). AMT builds on the Linear Time Frequency Analysis toolbox (LTFAT), which can be downloaded under `http://ltfat.sourceforge.net` (Søndergaard *et al.*, accepted for publication, 2011).

There is a big variety in auditory models described in the literature. AMT facilitates developing new models as it provides several implemented stages of auditory signal processing. The consistency of all functions within the toolbox helps the development process.

In the AMT the Dau et al. (Dau *et al.*, 1996a), the Zakarauskas (Zakarauskas and Cynader, 1993) and Langendijk (Langendijk and Bronkhorst, 2002) models had already been implemented. Their separately accessible model stages built the basis for our adapted localization model.
In the course of this project an experiment [1] to validate the figures of Lopez-Poveda and Meddis (2001) was contributed to the toolbox, as well as some existing functions [2] were improved.

1. use `exp_lopezpoveda2001.m` in AMT
2. see `drnl.m` and `middleearfilter.m` in AMT

4 Localization Model

4.1 Langendijk's Localization Model

In Langendijk and Bronkhorst (2002) they describe an auditory model to show the effect of spectral cues on the human median-plane sound localization. Their model [3] predicts the individual response patterns on the basis of the frequency spectrum of the signal presented to the listener and the individually measured directional transfer functions. Basically they predict the patterns by comparing the spectrum of the signal arriving at the eardrums with a set of stored spectral templates associated with particular sound source directions. Finding the template which is most similar to the input spectrum determines the most probable direction being perceived. Figure 5 shows the block diagram of their implemented localization model.

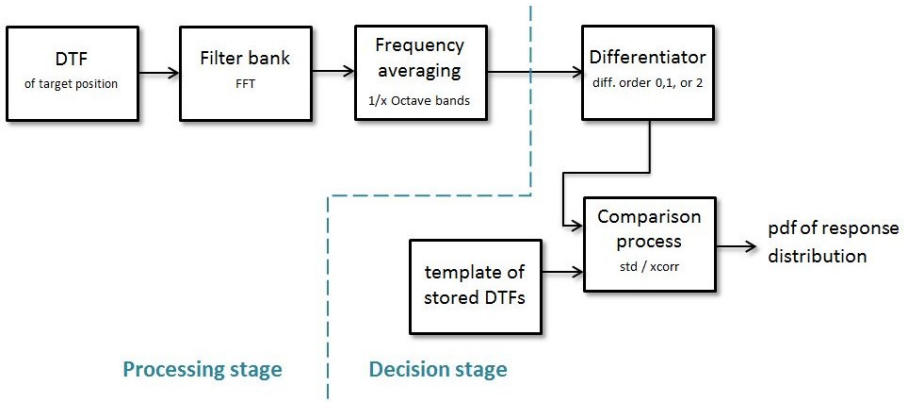

Figure 5: Block diagram of Langendijk's localization model

The model is based on the availability of a set of head-related transfer-functions (HRTFs) for several human median-plane positions. The first step is to take the directional dependent part of the HRTFs, the directional transfer function (DTF), of the target position. This DTF is processed through a linear filter bank, which is basically represented by a fast Fourier transform (FFT). Afterwards frequency averaging is applied. This process combines single spectral lines of a certain frequency bandwidth and averages them. The bandwidth can be chosen as a partial of octave, by default the DTFs get averaged in $\frac{1}{6}$-octave bands. The next step in the signal path is a differentiator which was inspired by Zakarauskas and Cynader (1993). Langendijk and Bronkhorst (2002) extended this stage with the ability to choose if the zeroth-, first-, or second-order derivative of the DTF is taken. By default no derivative is calculated.

Finally in the comparison process the similarity between the two DTFs is measured. This is done either by calculating the cross-correlation coefficient or building the standard deviation of the difference. The results of both measures are transformed to probabilities between 0 and 1.

3. use `langendijk.m` in AMT

This model is a temporal averaging model. The duration of the presented stimuli has no influence on the model results, neither the input sound level does. As the model only considers DTFs, no specific input signal can be set. Furthermore it is a linear model, thus it does not consider the nonlinear properties of the peripheral auditory processing.

4.2 Adapted Localization Model

To adapt the in chapter 4.1 described model to a more human-like processing physiology, peripheral stages from the auditory model of Lopez-Poveda and Meddis (2001) were integrated. Figure 6 shows the block diagram of the extended localization model.

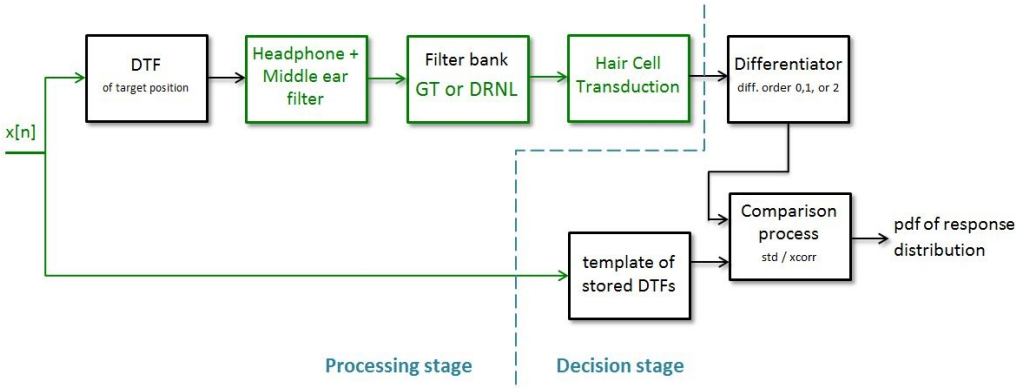

Figure 6: Block diagram of adapted localization model

The main difference is a dual resonance nonlinear filter bank (DRNL), which replaces the former linear filter bank to represent the human basilar membrane nonlinearity. Another new feature is that now any arbitrary input signal can be considered as it is first convoluted with the DTF and then further processed. Moreover new processing stages were added, they will be described in the following subsections. The decision stage of the model was not changed.

4.2.1 Headphone and middle ear filter

Like in Lopez-Poveda and Meddis (2001) a headphone and middle ear filter stage was implemented. More precisely two processes of the peripheral hearing are modeled by linear-phase, 512-point, finite impulse response (FIR) filter. First the headphone-delivered sound pressure waveform is transformed into vibration of the tympanic membrane, which, then in turn, induces vibration of the stapes, represented by the stapes velocity waveform.
The outer-ear frequency response originally was taken from Pralong and Carlile (1996) and corresponds to a typical human outer-ear pressure-gain function measured with a pair of Sennheiser HD-250 headphones. The headphone filter had already been implemented in AMT and could be applied directly. [4]

4. use `headphonefilter.m` in AMT

The middle-ear response (stapes velocity as a function of stimulus frequency) is derived from stapes displacement measurement data in cadavers by Goode *et al.* (1994). As the existing version in AMT slightly differed, amongst other things, due to a different input signal level convention as used in the models of Dau *et al.* (1996a) and Jepsen *et al.* (2008), the middle-ear filter was reimplemented. [5] Therefore data points of stapes displacement were read from Figure 1 of Goode *et al.* (1994) (shown in figure 7) and transformed to stapes velocity using the equation given in (1).

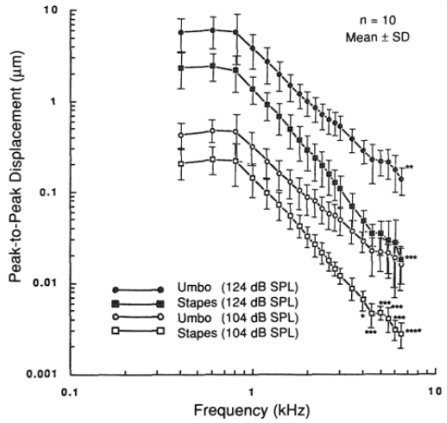

Figure 1. Umbo and stapes footplate displacement in ten human temporal bones at 104 and 124 dB SPL inputs at the TM. (Mean age 72.0 years, range 65–80 years). (* = 9 ears; ** = 8 ears; *** = 7 ears; **** = 6 ears)

Figure 7: Figure 1 from Goode *et al.* (1994)

$$Displacement_{p-p}(\mu M) = \frac{V_{p-p}}{2 \cdot \pi \cdot F} \cdot C \tag{1}$$

where the frequency F is given in kHz.

The calibration factor here is set to $C = 10^{-3}$ to use frequency points given in Hz. V_{p-p} represents the output peak-to-peak voltage of the laser Doppler measuring system and is proportional to stapes velocity. As the displacement data is measured at 104 dB SPL input level at the tympanic membrane, calculating V_{p-p} out of equation (1) refers to stapes velocity at the related 104 dB SPL input level. Observations of Goode *et al.* (1994) show a directly proportional relation between peak stapes velocity and stimulus pressure. Therefore stapes velocity results are transformed linearly to a stimulus level at the eardrum of 0 dB SPL. As the displacement data is given in μm the results get multiplied by 10^{-6} to obtain stapes velocity data in m/s.

Figure 8 shows filter functions of outer and middle-ear filter. The plots in the top row present the headphone filter, the bottom row shows the middle-ear response. The two

5. use `middleearfilter.m` in AMT

columns represent the implementation in the AMT [6] on the left and the original figure according to Lopez-Poveda and Meddis (2001) on the right.

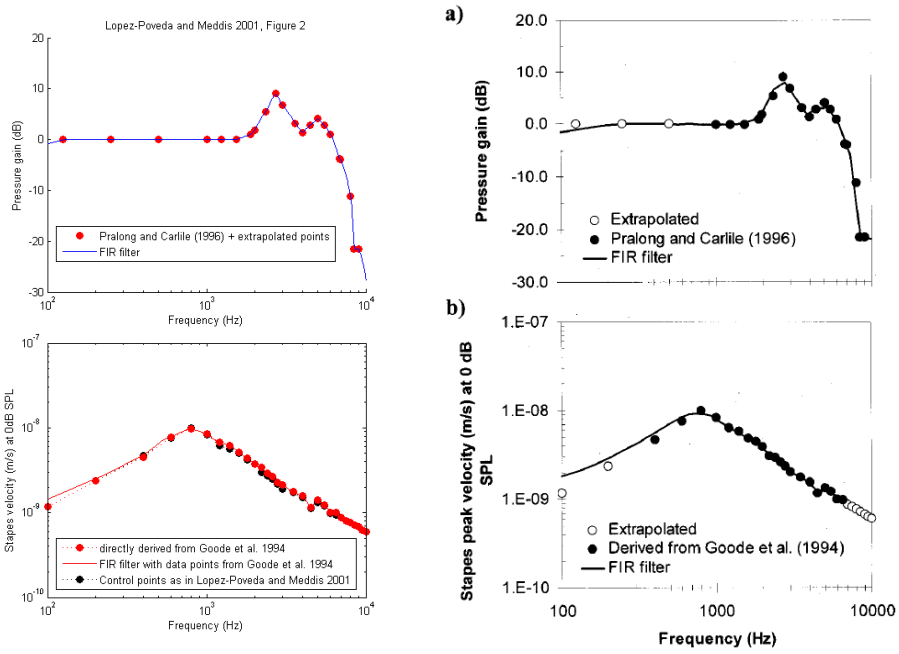

Figure 8: Headphone and middle ear filter
Implementation in AMT (left) and original according to Lopez-Poveda and Meddis (2001) (right)

4.2.2 Human Nonlinear Cochlear Filter Bank

The main improvement of the new adapted localization model is a dual resonance nonlinear filter bank (DRNL). [7] This nonlinear filter bank, contrary to the conventional FFT filter bank, is able to reflect the nonlinearity of the basilar membrane. The DRNL filter models the process of stapes motion inducing vibration of the basilar membrane (Meddis *et al.*, 2001; Lopez-Poveda and Meddis, 2001). Figure 9 shows the block diagram of the implemented filter bank.

6. use `exp_lopezpoveda2001('fig2').m` in AMT
7. use `drnl.m` in AMT

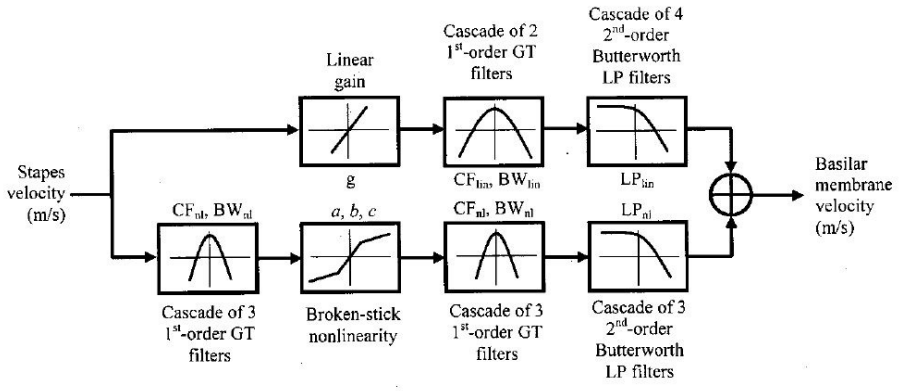

Figure 9: DRNL: Block diagram (Fig. 3a, Lopez-Poveda and Meddis, 2001)

The input signal of the filter bank is processed in two paths independently and then summed up again. In each filter of the filter bank the input signal on the one hand follows a linear signal path, where first a linear gain is applied and then the signal is filtered through a cascade of two first-order gammatone filters and subsequently through a cascade of four second-order low pass filters. On the other hand the nonlinear path is represented by three first-order gammatone filters, the nonlinear "broken-stick" gain, another cascade of three gammatone filters equal to the one before and a cascade of three second-order low pass filters. The output signal of one particular filter is the sum of both paths and represents basilar membrane velocity (Lopez-Poveda and Meddis, 2001) at a certain best frequency (BF).

All the filter parameters like cut off frequency, bandwidth and gain can be specified by two coefficients p_0 and m, describing the function of the logarithm of the BF of the respective filter like in equation (2).

$$log_{10}(parameter) = p_0 + m \cdot log_{10}(BF) \tag{2}$$

These two regression coefficients create the parameters for the whole filter bank by linear regression of each parameter at intermediate BFs. The default parameter set [8] which is used for all further simulations, unless otherwise noted, can be found in Table III (Average response) of Lopez-Poveda and Meddis (2001).

For low and high input signal levels the DRNL operates nearly linear, whereas for moderate values it shows a nonlinear behavior. To visualize the dependence of the filter form on the input level figure 10 shows the filter output as a function of signal frequency for two different input levels, 30 (upper row of plots) and 85 dB SPL. Additionally the outputs of the linear and the nonlinear path are presented.

8. in `drnl.m` in AMT

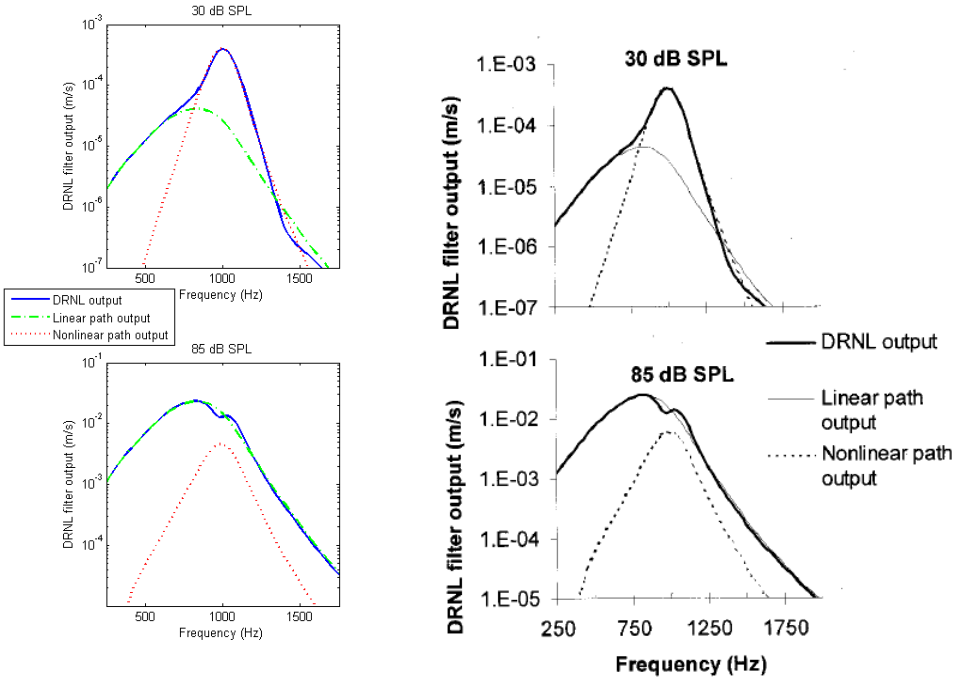

Figure 10: Filter output as a function of signal frequency at CF of 1kHz
Implementation in AMT (left) and original according to Lopez-Poveda and Meddis (2001) (right)

More precisely the plots are generated using the parameter set of one single subject at the 1-kHz site (subject YO, Table I, Lopez-Poveda and Meddis, 2001). The input signal is a pure tone with varying frequency. For the lower input level of 30 dB SPL it can be seen that both signal paths contribute to the DRNL output. However, the higher the input level gets, the more the filter shape widens, as we would expect from auditory filters. As the bottom row plots for 85 dB SPL show, the filter output equals more and more the linear path output, the nonlinear path has only marginal influence. Furthermore the shape towards low frequencies gets flattened.
Figure 10 presents the results of the implementation [9] as well as the original figure (Fig. 3b and c) from Lopez-Poveda and Meddis (2001) to compare. The original figures could be fully reproduced.

The filter shape of the DRNL is compared to the one of a gammatone filter, which is widely used to represent auditory filters. Figure 11 shows the normalized magnitude transfer functions of the DRNL filter tuned to 1kHz for different input levels of 20 (left), 50 (middle) and 80 (right) dB SPL (solid curves) and at the same time the transfer function of the corresponding fourth-order gammatone filter (dashed curves). The higher the input level, the wider the (DRNL) filter, as the auditory filters are assumed to behave. For high input levels a slight frequency shift towards low frequencies can be observed, noticeable in the rightmost plot. The gammatone filter shape is independent of input level.

9. use `exp_lopezpoveda2001('fig3bc').m` in AMT

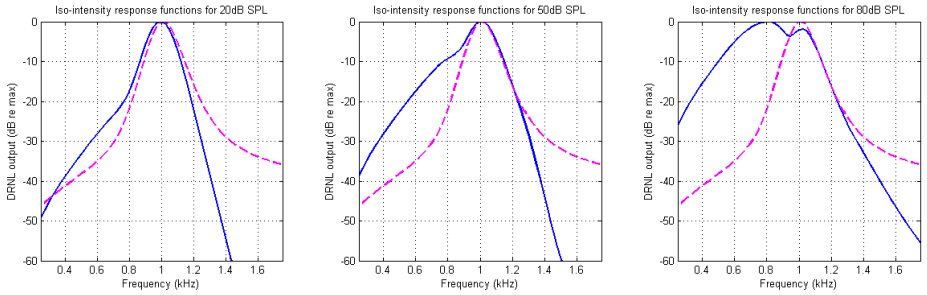

Figure 11: DRNL vs. Gammatone filter bank
Normalized magnitude transfer functions of the DRNL filter tuned to 1kHz for input levels of 20 (left), 50 (middle) and 80 (right) dB SPL (Solid curves), Transfer function of the corresponding fourth-order gammatone filter (Dashed curves)

To further validate the implemented model psychoacoustic data, more precisely experimental pulsation threshold data from Plack and Oxenham (2000), was modeled. For further details please see chapter III., A. in Lopez-Poveda and Meddis (2001). The simulation results[10] can be seen in figure 12. Figure 13 is the original figure (Fig. 4) of Lopez-Poveda and Meddis (2001), where the experimental pulsation threshold data (filled symbols for subject YO, open symbols for average data of 6 subjects) and their model results (continuous lines) are shown. The comparison of those figures prove the validation of our results.

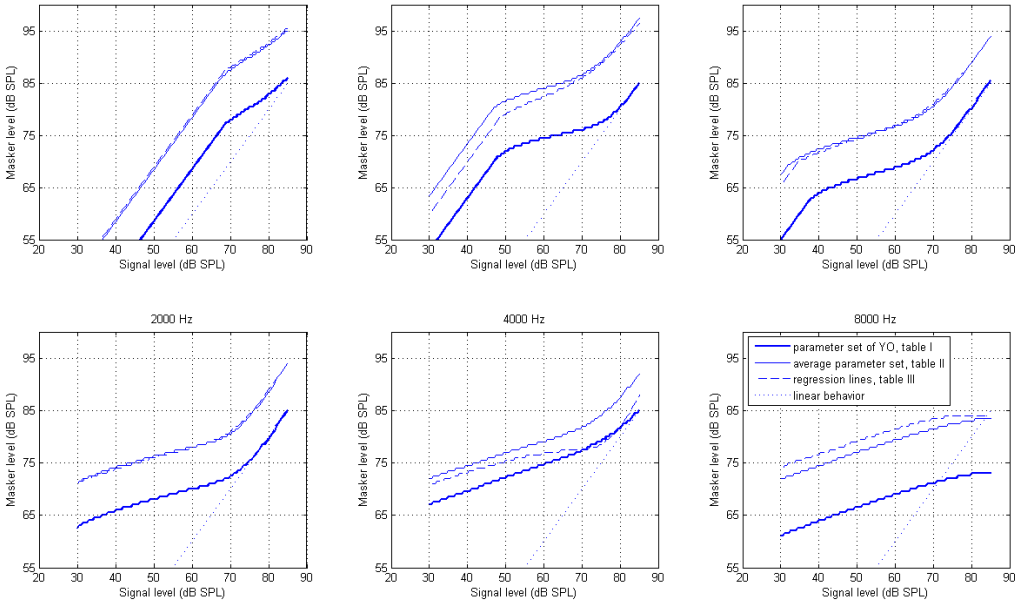

Figure 12: Modelling psychoacoustic data
(pulsation threshold Plack and Oxenham, 2000), Implementation in AMT

10. use `exp_lopezpoveda2001('fig4').m` in AMT

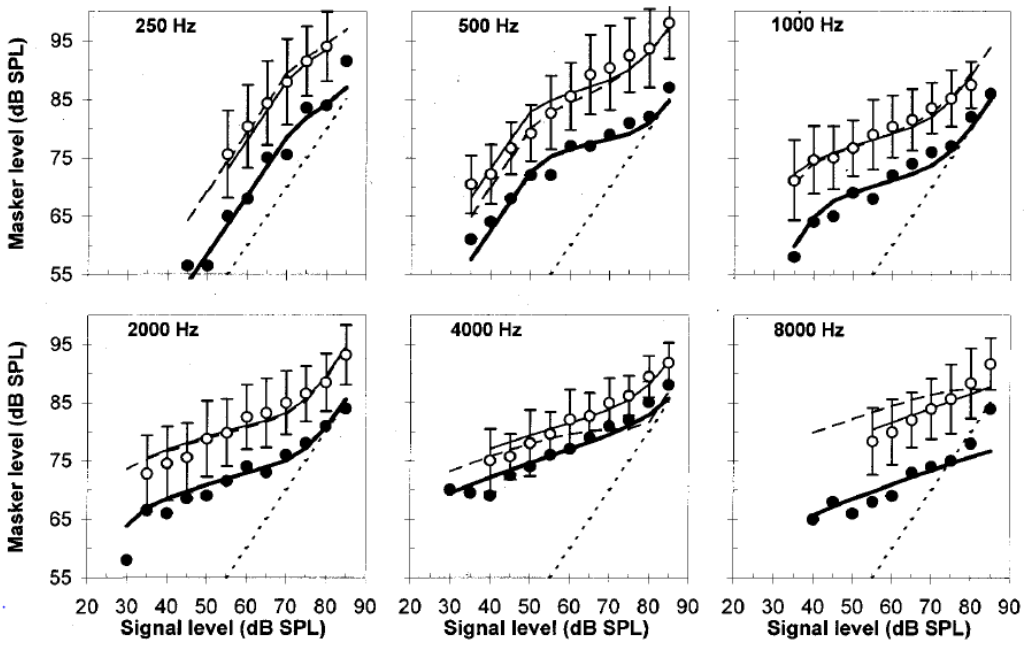

Figure 13: Modelling psychoacoustic data
(pulsation threshold Plack and Oxenham, 2000), Fig. 4 of Lopez-Poveda and Meddis (2001)

4.2.3 Hair Cell Transduction

This stage acts as the transformation of the mechanical oscillations of the basilar membrane into receptor potentials in the inner hair cells (IHC). This process is basically simulated by envelope extraction through half wave rectification followed by a low pass filter with a cutoff frequency f_c of 1000 Hz (Dau *et al.*, 1996a). This stage was already implemented in AMT [11] and directly included in the model.

The current localization model integrates every filter output over time and takes the mean (RMS). This results in one representative value for each filter output. The output of the whole filter bank builds the basis for the subsequent comparison process.
With no further stages included the IHC stage has just a small effect on the model results. More effect is expected when the temporal aspects of the stimuli are considered. Therefore in this adapted localization model the hair cell transduction stage is already included and the model can be expanded easily.

11. use `ihcenvelope.m` in AMT

5 Results - Modeling the Langendijk Experiment

The first step was to validate the new developed model simulating the localization experiment of Langendijk and Bronkhorst (2002).

In their study localization was investigated in the median plane. The influence of spectral cues on sound localization was investigated by removing spectral cues in frequency bands varying bandwidth and center frequency. They used 200 ms Gaussian noise bursts as target stimuli, whose HRTF filtered version was presented via headphones. The bursts were bandpass filtered between 200 Hz and 16 kHz with 10 ms cosine square on- and off-set ramps. The stimuli were presented at an A-weighted level of approximately 65 dB (Langendijk and Bronkhorst, 2002).

Figure 14 shows the original Fig. 7 from Langendijk and Bronkhorst (2002) [12]. In the first 5 panels you can see the probability density function (pdf) for one listener (P6) as a function of target position, plotted in a gray colored map. So the model predicts the localization probability, i.e. the ability of a proband to localize the given stimuli presented in the median plane. White areas represent a high and black ones a low probability, that the predicted response corresponds to the sound target angle. Each column in each subplot corresponds to a single pdf for all possible response locations for one target angle. The circles represent actual responses given by the listener in the conducted experiment. The 5 panels address 5 different conditions, in which Langendijk and Bronkhorst (2002) removed spectral cues from certain frequency bands and investigated its influence on the localization performance.

The panel in the right lower corner shows the likelihood statistic of the model. The bars stand for the so called actual likelihood in each condition. It evaluates the model-predicted pdf matrix on positions of the actual responses. The dots show the expected likelihood, which bases on randomly generated response patterns according to the modeled pdfs. This likelihood estimation will be averaged over 100 trials and the bars indicate the corresponding 99% confidence interval.

The model is able to predict the data if the actual likelihood lies within the error bars of the expected likelihood.

12. use also `exp_langendijk2002.m` in AMT

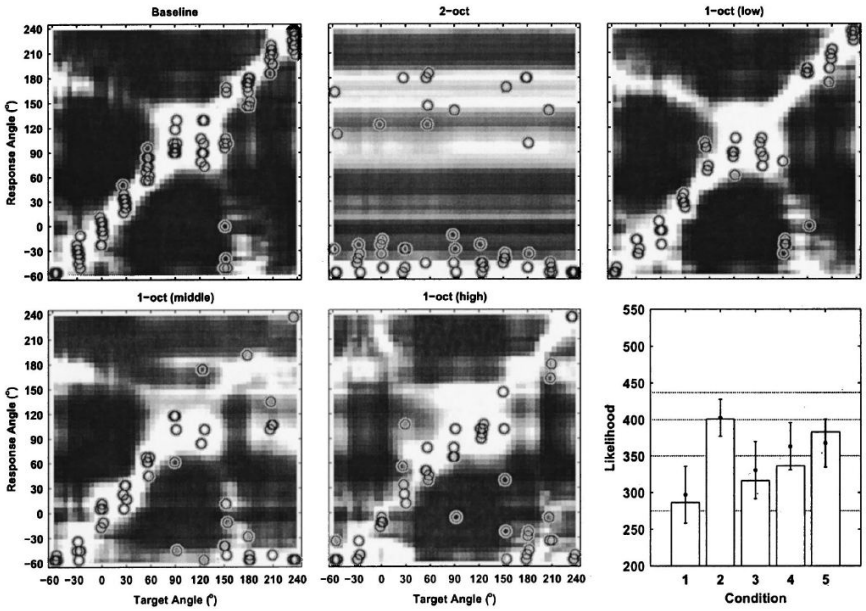

Figure 14: Fig. 7, Langendijk and Bronkhorst (2002)

Figure 15 shows analog to the original figure the pdfs predicted by the new advanced auditory model [13].

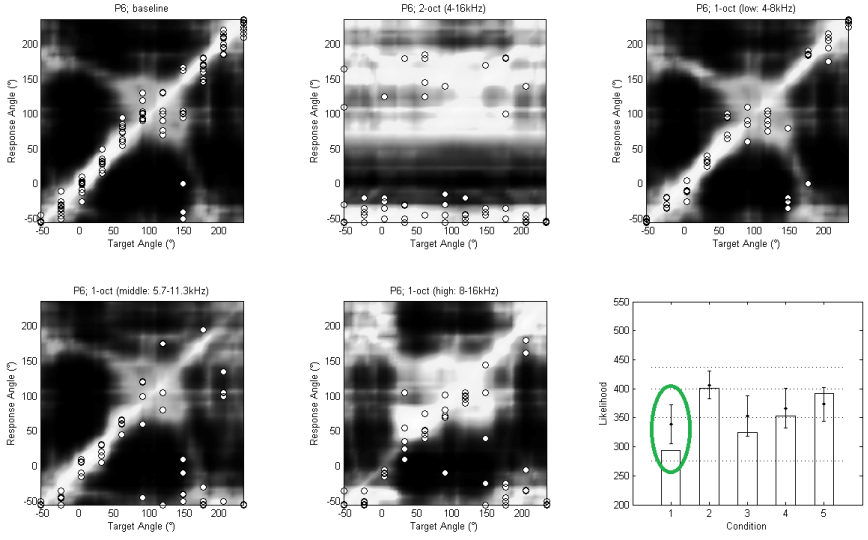

Figure 15: Gammatone filter bank; input SPL of 65 dBA

First, a gammatone filter bank was used to model the data. Contrary to the model of Langendijk and Bronkhorst (2002) where no input signal was required, Gaussian noise bursts at an input sound level of 65 dBA like in the original experiment were used. The

13. use `exp_locamo_langendijk2002.m`

DTFs provided in Langendijk and Bronkhorst (2002) [14] were applied.
The circles show, as in all other following plots in this chapter, the actual responses given by the listener in the conducted experiment as in figure 14.
In the likelihood statistic it can be seen that in the first condition (baseline) the model does not fit the data anymore. The actual likelihood bar does not lie within the confidence interval of the expected likelihood. However, as the first 5 panels show the pattern of the localization probability can be predicted by the advanced model. The results are still comparable to the original model.
As the gammatone filter bank is a linear filter bank, changing the input level of the stimuli does not affect the outcome and the model predicts the same results.

In the next step the gammatone filter bank is replaced by the DRNL. Figure 16 shows the pdfs predicted by the complete advanced auditory model.

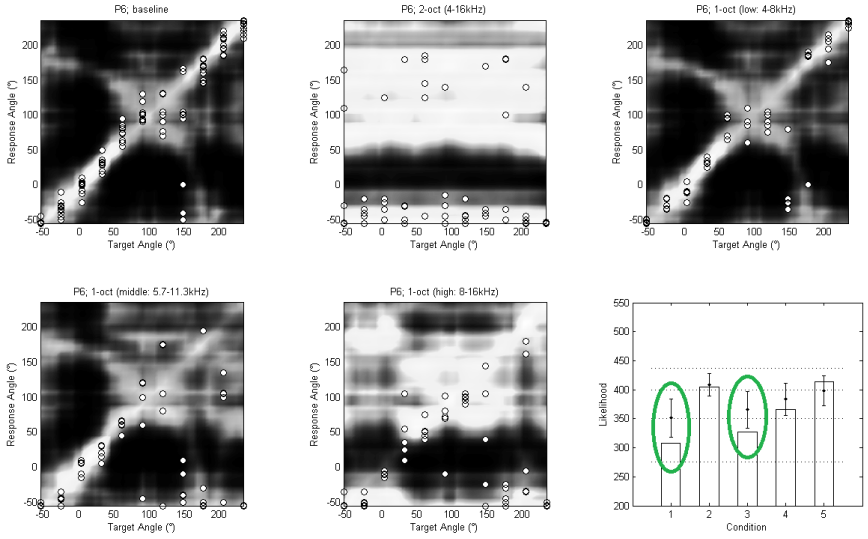

Figure 16: DRNL filter bank; input SPL of 65 dBA

As the likelihood statistic indicates, the model fit in the first condition is still worse than in the original model. Furthermore also the third condition is worse.
It seems that Langendijk and Bronkhorst (2002) mapped their model more to the data than to the physiology of the human ear.

To show the influence of the nonlinear filter bank the input level of the sound stimuli was changed.

Figure 17 shows the result plots predicted by the advanced model for an input level lowered to 40 dBA. Now the model predicts the given data set successfully. This might point out a mismatch of the input sound level and the operating point of the filters in the used model.

14. use `data_langendijk2002.m` in AMT

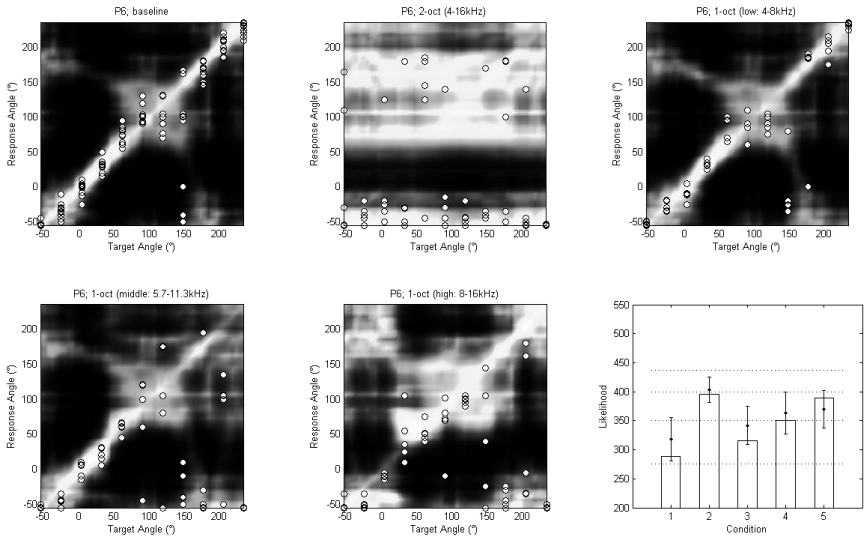

Figure 17: DRNL filter bank; input SPL of 40 dBA

Figure 18 shows the result plots predicted by the advanced model for a higher input level of 90 dBA. The details in the pdf patterns get blurred. This effect is caused by wider filters in the DRNL due to the higher input level. As expected the likelihood statistic illustrates that the model cannot fit the data and the results are bad.

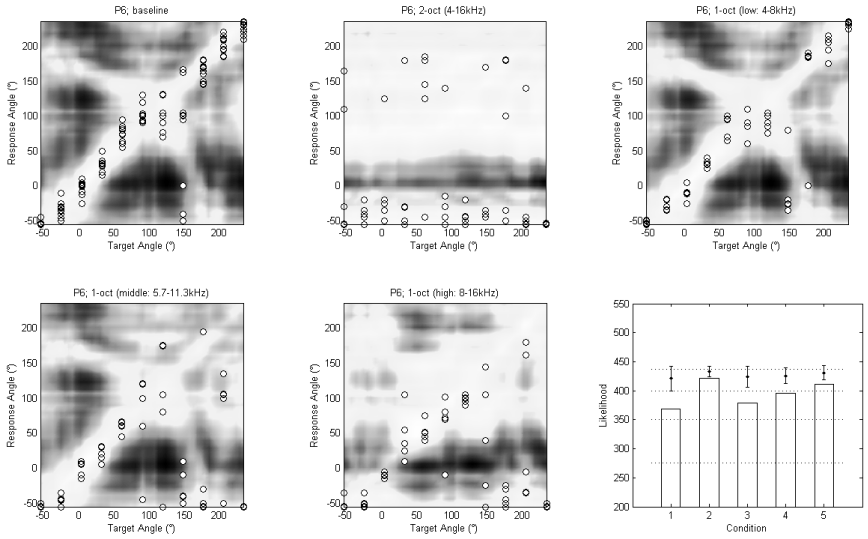

Figure 18: DRNL filter bank; input SPL of 90 dBA

In conclusion it could be demonstrated that due to the nonlinear filter bank the adapted model shows a level-depending effect, which will be discussed further in the subsequent chapter.

6 Elevation Gain

The slope of the relation between target and response for the elevation for sound localization in the vertical plane is referred to as elevation gain (Vliegen and Van Opstal, 2004). They state that the elevation gain varies with signal level and duration.

They found varying elevation gains in a non monotonic way with sound intensity. At low and high sound levels gains are low whereas for intermediate sound levels they increase, even reaching a positive level effect. Figure 19 shows the left panel of the original Figure 8 from Vliegen and Van Opstal (2004), where elevation gains are plotted as a function of sensation level. It is a schematic representation of the elevation gain effect, made by regression based on data points for one test subject (JV).

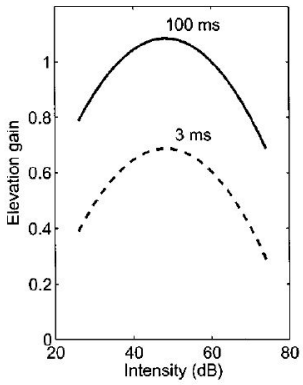

Figure 19: Elevation gain effect for test subject JV, left panel from Fig. 8, Vliegen and Van Opstal (2004)

As the adapted localization model has no stage included taking the input signal duration into account so far, the focus in this project lies on modeling the elevation gain effect regarding the signal level. The goal is to reproduce a similar level effect with the adapted model.

The parameters for the simulation were defined as follows: As target stimuli 100 ms Gaussian white noise bursts with 10 ms cosine square on- and off-set ramps were used. In Vliegen and Van Opstal (2004) sound sources were placed all over the frontal hemisphere. However, due to the model characteristics the target was placed in 21 different positions only in the median plane (ranging from $-30°$ to $+70°$ elevation) with 100 repetitions per target. The simulation was run for 9 different input sound levels between 20 and 100 dB SPL in 10 dB steps. Simulation results get averaged for HRTFs of 60 subjects. HRTFs were provided by the ARI HRTF Database which is free to download under (`http://www.kfs.oeaw.ac.at`).

The model calculates the localization probability for the given input stimulus at a certain input level for one listener [15]. Analog to the panels in the plots of chapter 5 the predicted

15. use `exp_locamo_vliegenvanopstal2004.m`

pdf can be plotted in a gray colored map as a function of target position (example plot in figure 20).

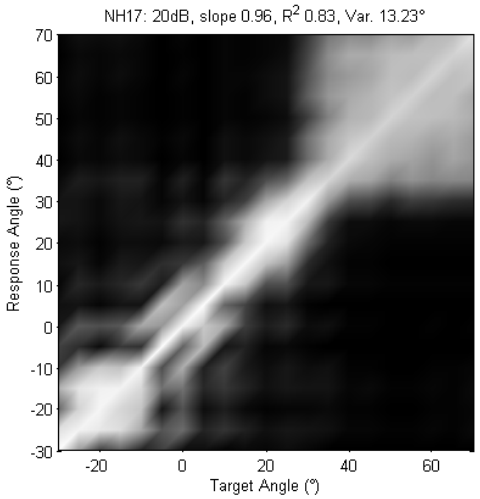

Figure 20: Example of simulated localization probability at a certain input level (20 dB SPL) for one subject (NH17)

Now for each target position a subsequent stage predicts a corresponding response angle according to the localization probability calculated before by the model. This procedure is repeated a 100 times to finally get 100 responses (repetitions) for each single target angle. Afterwards linear regression is calculated according to the resulting target/response combinations T_i/R_i (see equation 3).

$$R_i = B_1 \cdot T_i + B_0 \tag{3}$$

This regression line corresponds to the elevation gain.

As shown in figure 21 without any further processing the regression lines contain an offset at target positions located at outer ranges of the calculated values. It occurs that the lines get twisted due to an incorrect slope.

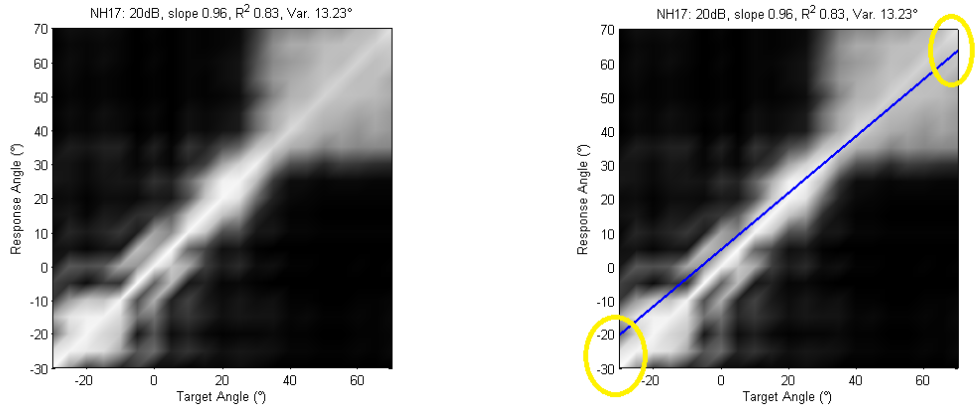

Figure 21: Simulated localization probability at an input level of 20 dB SPL for one subject (NH17), regression line containing offset

We know that every column of the localization probability plot corresponds to the response probability for one target. If we enlarge on the pdf for an intermediate target angle as in figure 22, we can assume a Gaussian normal distribution function.

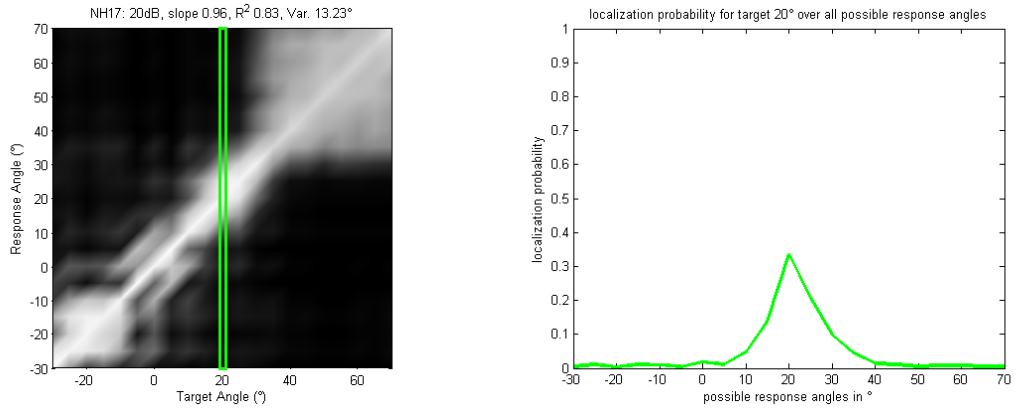

Figure 22: Simulated localization probability at an input level of 20 dB SPL for one subject (NH17), single pdf for all possible response locations for a target angle of $+20°$

The randomly generated corresponding response pattern provides a sample distribution according to the before predicted localization probability. A histogram of the number of elements for each possible response position (normalized to the maximum appearance) can be plotted as in figure 23. According to the pdf also the predicted response angles are Gaussian distributed around the ideal response position (ideal localization: response = target angle).

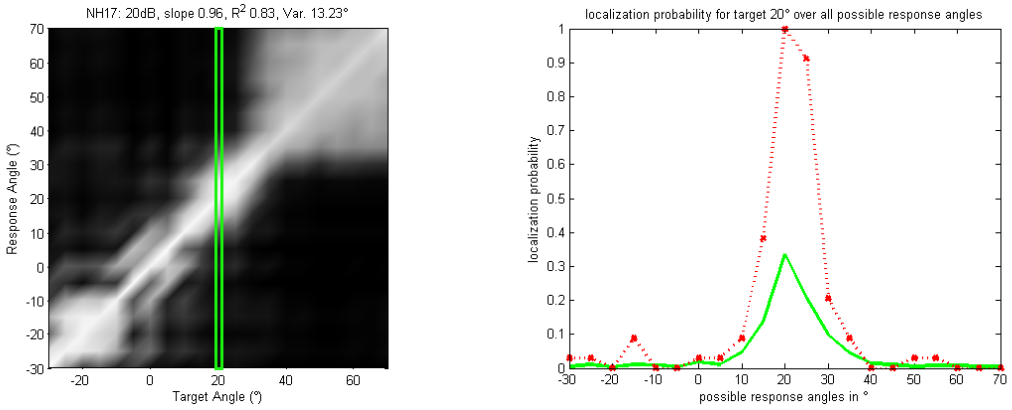

Figure 23: Simulated localization probability at an input level of 20 dB SPL for one subject (NH17), single pdf for all possible response locations for a target angle of $+20°$ (green) and the normalized histogram of number of elements for all possible response angles normalized to the maximum appearance (red)

However, the distribution of the resulting response points cannot always be assumed as Gaussian. This appears to be more likely for target angles at the outer ranges of all possible calculated values. Figure 24 shows the single pdf for all possible response locations for the target angle of $-30°$.

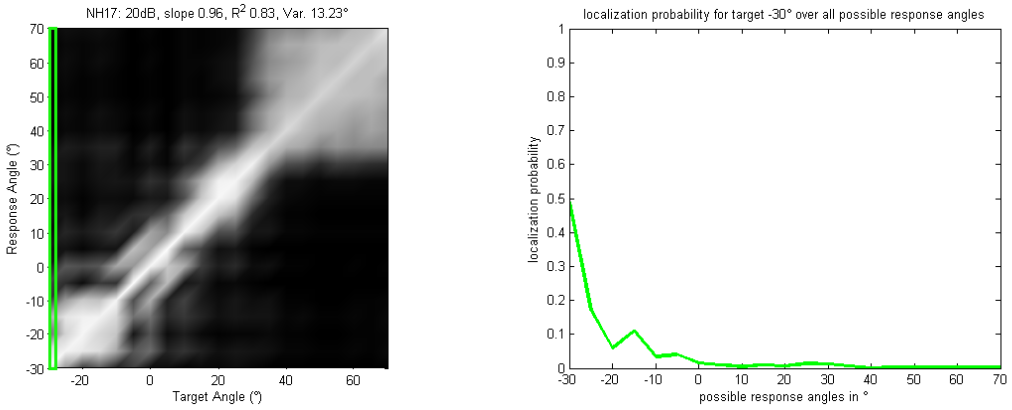

Figure 24: Simulated localization probability at an input level of 20 dB SPL for one subject (NH17), single pdf for all possible response locations for a target angle of $-30°$

If we focus on the distribution of the chosen response values, we see that the generated response pattern is not as steep as expected. It follows the pdf only roughly. This results in the offset of the linear regression lines.

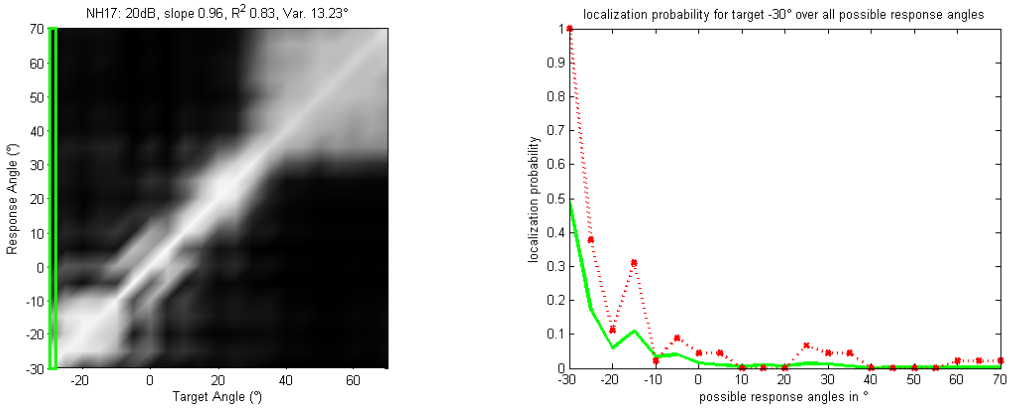

Figure 25: Simulated localization probability at an input level of 20 dB SPL for one subject (NH17), single pdf for all possible response locations for a target angle of $-30°$ (green) and the normalized histogram of number of elements for all possible response angles normalized to the maximum appearance (red)

Squaring the probability function enhances the most probable values and weakens the values with low probability (see figure 26).

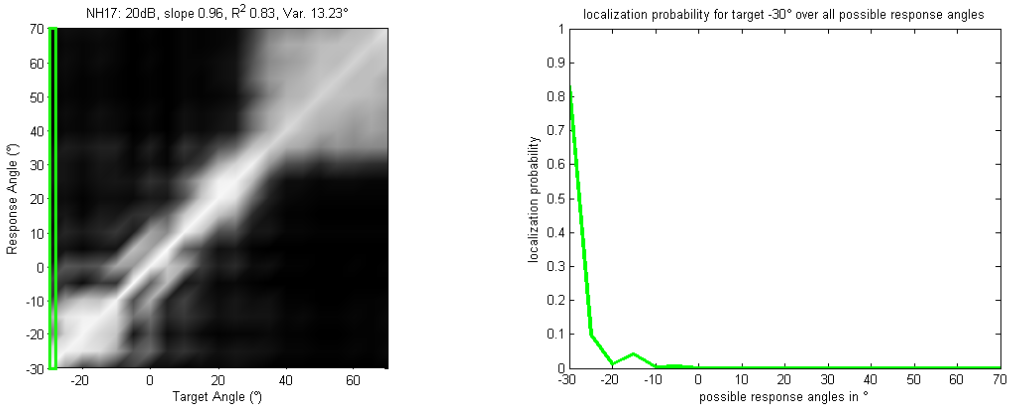

Figure 26: Simulated localization probability at an input level of 20 dB SPL for one subject (NH17), squared single pdf (right panel) for all possible response locations for a target angle of $-30°$

When we now look at the generated responses according to the squared pdfs, the sample distribution is able to approximate the localization probability (see figure 27). The offset of the regression lines could be removed.

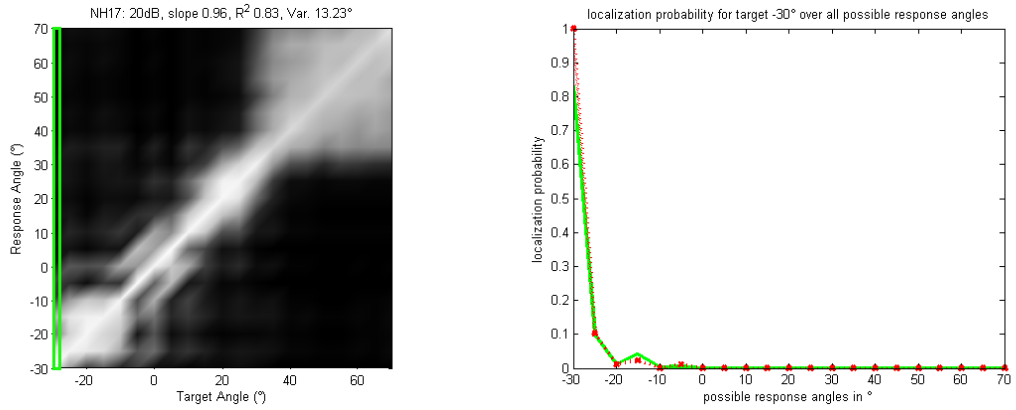

Figure 27: Simulated localization probability at an input level of 20 dB SPL for one subject (NH17), squared single pdf (right panel) for all possible response locations for a target angle of $-30°$ (green) and the normalized histogram of number of elements for all possible response angles normalized to the maximum appearance (red)

A further requirement to the simulation is, that in real experiments, subjects are allowed to respond outside the target ranges. To consider this demand the response probability of each target gets convoluted with a Gaussian normal distribution function. This extends the possible response value range. Furthermore the response probability function of positions located at outer ranges become even more Gaussian. Hence, the generated response values occur more likely at the real maximum of the probability function.

Figure 28 shows the squared single pdf for all possible response locations for a target angle of $-30°$ on the left and on the right it presents the same probability function but convoluted with a normal probability density function (a Gaussian pulse with zero mean and standard deviation of $10°$).

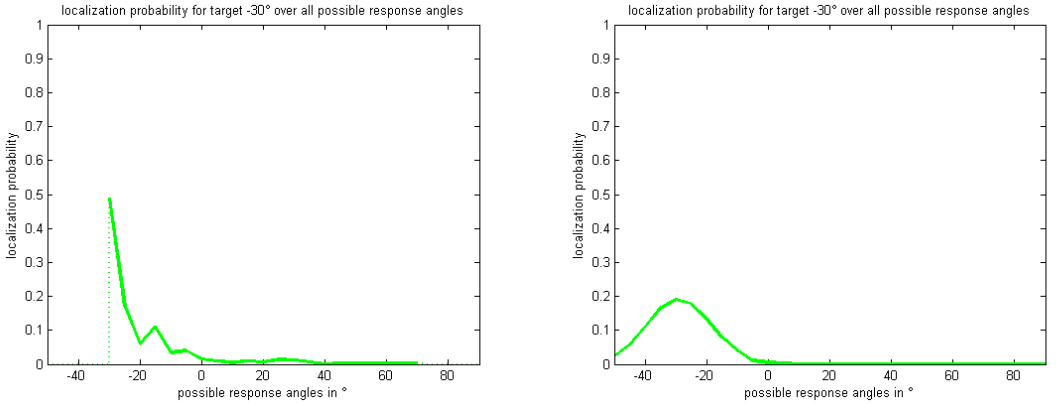

Figure 28: Squared single pdf (right panel) for all possible response locations for a target angle of $-30°$ (left), convoluted with a normal probability density function (zero mean and standard deviation of $10°$)

The convolution expands the possible response value range and additionally flattens the probability functions. Figure 29 presents in the left panel the simulated localization

probability at a certain input level (20 dB SPL) for one subject (NH17) with adjusted axes to the expanded response value range. In the right panel the corresponding squared and convoluted pdfs are pictured.

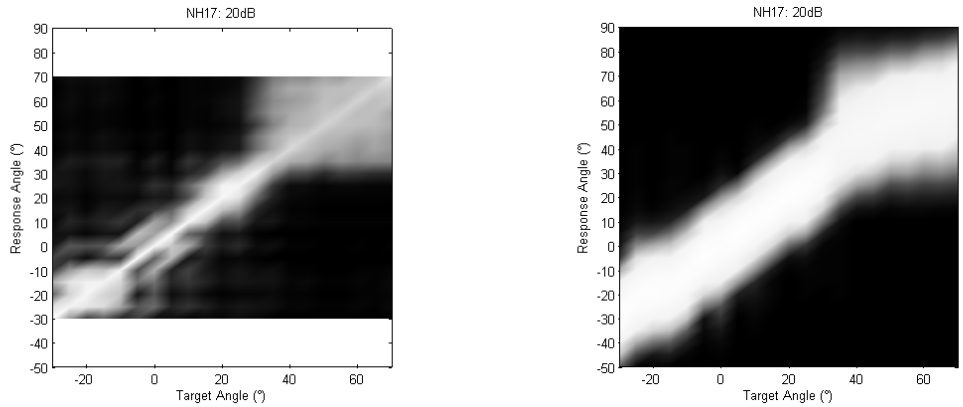

Figure 29: Simulated localization probability at a certain input level (20 dB SPL) for one subject (NH17) with adjusted the axis to the expanded response value range (left) and corresponding squared and convoluted pdfs (right)

If now linear regression is calculated out of the resulting target/response according to the squared and convoluted localization probability, the slope of the regression lines can be calculated correctly. This slope corresponds to the elevation gain. Figure 30 shows the corrected regression line fitted into the plots of figure 29.

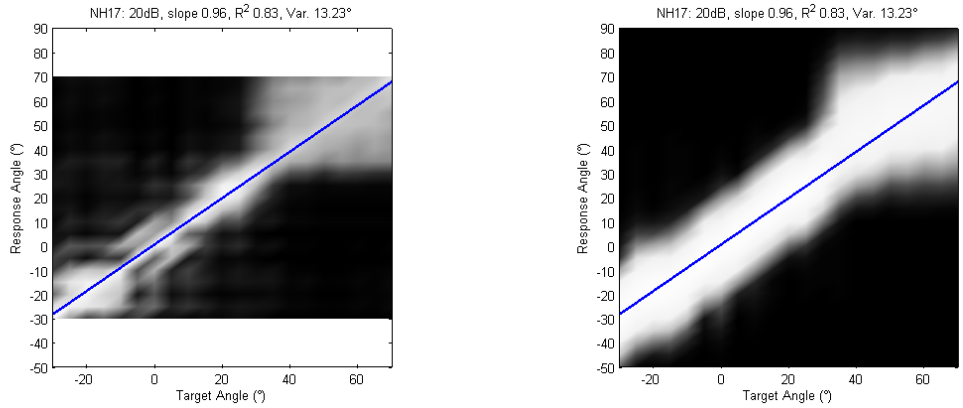

Figure 30: Simulated localization probability at a certain input level (20 dB SPL) for one subject (NH17) with adjusted the axis to the expanded response value range (left) and corresponding squared and convoluted pdfs (right), regression lines in blue

6.1 Results - Modeling Elevation Gain

Figures 31 and 32 show the predicted localization probability for all 9 calculated input levels for two different test subjects (NH17 and NH94). Furthermore the calculated regression lines are marked in blue. Their calculation is based on squared and with

a Gaussian pulse convoluted pdfs. On the contrary the pdfs are still plotted in the unprocessed version to highlight the details.

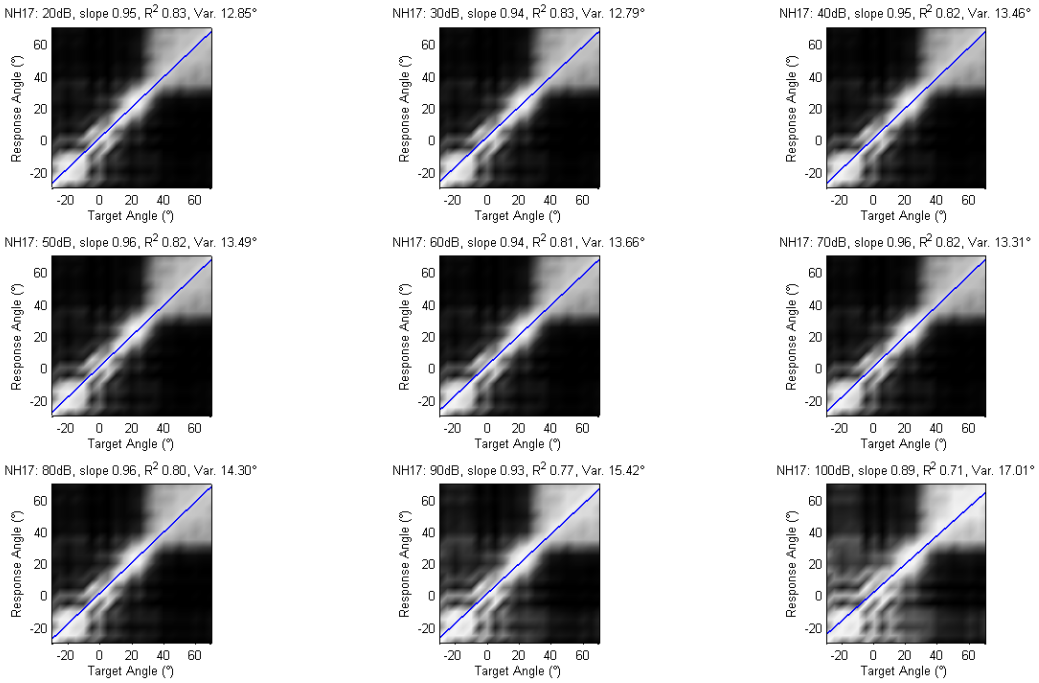

Figure 31: Modeling elevation gain, NH17

In the header of every subplot the used input level is indicated, as well as the slope of the plotted regression line which equals the elevation gain, the R-square value of the fitting and the response variability, which refers to the averaged standard deviation angle in the polar dimension of all response positions.

It can be seen that for higher input level the slope and thus the elevation gain decreases. This can also be noticed in the more and more blurred localization probabilities which therefore become whiter.

The two figures show, that the models predicts different localization probabilities for different subjects depending on what their HRTFs look like. In general HRTFs are very individual. For example in figure 32 (NH94) the plots are more blurred because of HRTFs with spectral cues obviously not as dominant as in the ones of (NH17).
For this test subject (NH94) the effect of the decreasing elevation gain for higher sound levels is even more noticeable.

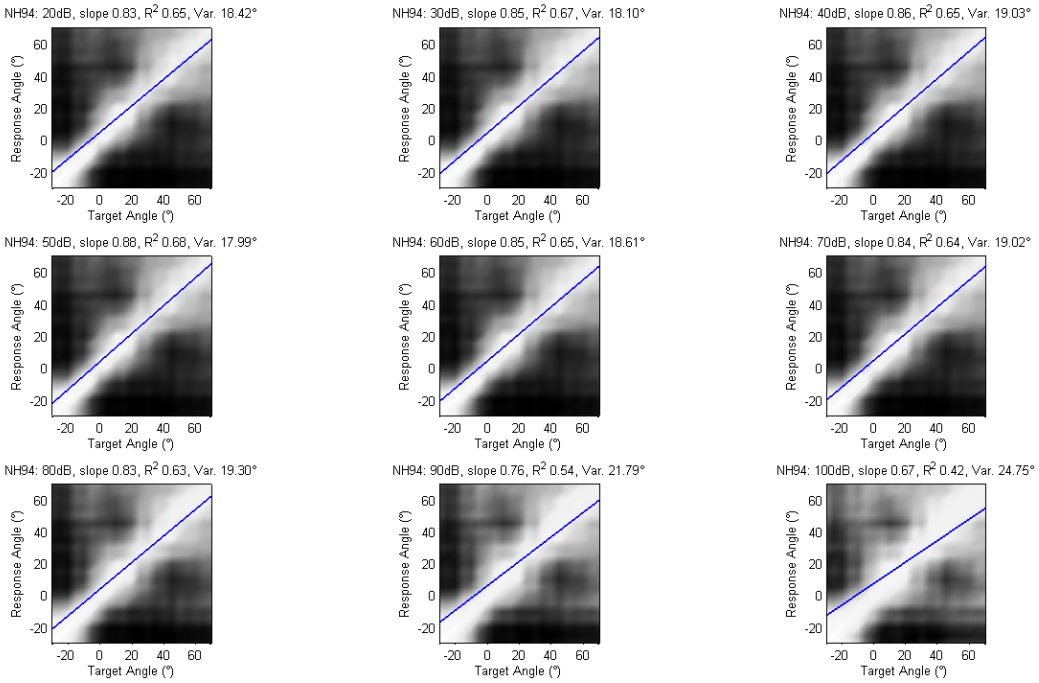

Figure 32: Modeling elevation gain, NH94

The simulation was run for 60 different HRTFs. For every listener the slope of the regression line, i.e. the elevation gain, and the response variability in the polar dimension was calculated for every different input level. Afterwards the mean average over all subjects with 95% confidence interval for both measures was calculated. The mean elevation gain and the mean response variability in $^\circ$ as functions of input signal level can be seen in figure 33. The red crosses in the left panel of the elevation gain mark the curve progression of the expected effect according to the schematic representation of Vliegen and Van Opstal (2004) in figure 19.

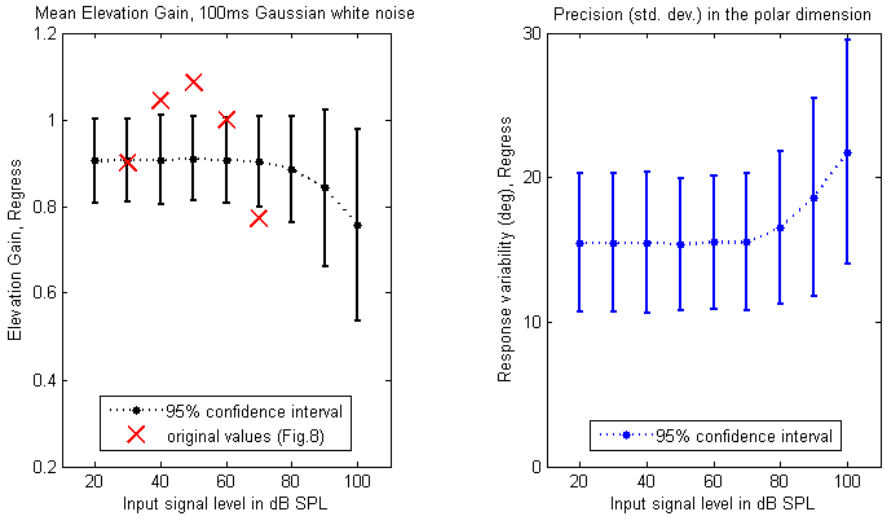

Figure 33: Averaged simulation results over all 60 subjects
Elevation gain (left) and Response variability in ° (right)

The average results show a decreasing elevation gain with increasing input signal level. So there is an elevation gain effect visible, however it occurs for levels much higher than expected. There is no effect found at low input signal levels. Furthermore no positive elevation gain could be found neither. Contrary to the elevation gain the response variability increases with increasing signal levels. This is reasonable, as a worse (lower) elevation gain implicates a worse localization performance which can be demonstrated by the higher response variability.

The results show a trend, however to define the effect with certain significance different tasks should be simulated and more statistics should be applied.

7 Conclusion and Discussion

A localization model was developed to simulate sound localization experiments in the human median plane. It was done by adapting an already existing model (Langendijk and Bronkhorst, 2002) to a more human-like processing. The main change was to include the nonlinear DRNL filter bank to simulate the behavior of filter width and compression of the auditory filters.

The resulting model was evaluated and validated with already existent data. First localization experiments from Langendijk and Bronkhorst (2002) were simulated. Their figures could be reproduced and validated, although a mismatch of the input sound level and the operating point of the filters in the used model is suspected. Furthermore, it could be demonstrated that due to the nonlinear filter bank the adapted model output depends on the input signal level. Inspired by Vliegen and Van Opstal (2004) this level-depending effect was investigated in a simulation of localization ability as a function of input signal level. In general a so called elevation gain effect could be modeled, however the results accompany other literature only to a certain degree. The results show a decreasing elevation gain for higher input signal levels. This indicates worse localization performance, as supported by an increasing response variability with increasing signal levels.

In the current model the duration of the presented stimuli has no impact on the results. However, in reality level and duration of the sound stimuli affects the localization performance. This encourages to include a time depending stage into the adapted model.

References

Chittka, L. and Brockmann, A. (**04 2005**), "Perception Space - The Final Frontier," PLoS Biol **3**(4), e137, URL http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0030137.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996a**), "A quantitative model of the effective signal processing in the auditory system. I. Model structure," The Journal of the Acoustical Society of America **99**(6), 3615–3622.

Goode, R., Killion, M., Nakamura, K., and Nishihara, S. (**1994**), "New knowledge about the function of the human middle ear: development of an improved analog model," The American journal of otology **15**(2), 145–154.

Gray, H. (**1918**), *Anatomy of the Human Body*, 20th / bartleby.com ed., URL http://www.bartleby.com/107/.

Jepsen, M., Ewert, S., and Dau, T. (**2008**), "A computational model of human auditory signal processing and perception," The Journal of the Acoustical Society of America **124**(1), 422–438.

Laback, B. (**2010**), "Psychoakustik II, Schwerpunkt: Experimentelle Audiologie, Skriptum zur Vorlesung," .

Langendijk, E. and Bronkhorst, A. (**2002**), "Contribution of spectral cues to human sound localization," The Journal of the Acoustical Society of America **112**, 1583–1596.

Larsen, A. P. (**2010**), "Modeling Binaural Signal Processing in Humans," Master's thesis, Technical University of Denmark.

Lopez-Poveda, E. A. and Meddis, R. (**2001**), "A human nonlinear cochlear filterbank," The Journal of the Acoustical Society of America **110**, 3107–3118.

Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A. (**2001**), "A computational algorithm for computing nonlinear auditory frequency selectivity," The Journal of the Acoustical Society of America **109**, 2852–2861.

Moore, B. (**2002**), "Psychoacoustics of normal and impaired hearing," British Medical Bulletin **63**, 121–134.

Pickles, J. O. (**2008**), *An introduction to the physiology of hearing* (Emerald Group Publishing, United Kingdom).

Plack, C. J. and Oxenham, A. J. (**2000**), "Basilar-membrane nonlinearity estimated by pulsation threshold," The Journal of the Acoustical Society of America **107**, 501–507.

Pralong, D. and Carlile, S. (**1996**), "The role of individualized headphone calibration for the generation of high fidelity virtual auditory space," The Journal of the Acoustical Society of America **100**, 3785.

Søndergaard, P. L., Culling, J. F., Dau, T., Goff, N. L., Jepsen, M. L., Majdak, P., and Wierstorf, H. (**2011**), "Towards a binaural modelling toolbox," .

Søndergaard, P. L., Torrésani, B., and Balazs, P. (**accepted for publication, 2011**), "The Linear Time Frequency Analysis Toolbox," International Journal of Wavelets, Multiresolution Analysis and Information Processing .

Vliegen, J. and Van Opstal, A. J. (**2004**), "The influence of duration and level on human sound localization," The Journal of the Acoustical Society of America **115**, 1705–1713.

von Békésy, G. (**1949**), "On the resonance curve and the decay period at varios points on the cochlear partition," The Journal of the Acoustical Society of America **21**, 245–254.

Zakarauskas, P. and Cynader, M. (**1993**), "A computational theory of spectral cue localization," The Journal of the Acoustical Society of America **94**, 1323–1331.