# Framework for Modeling Forced-Choice Experiments in the Auditory Modeling Toolbox

Student Project

Martina Kreuzbichler

Supervisors:
Doz. Dr. Piotr Majdak (ARI)
DI Ph.D. Matthias Frank (IEM)

Graz, December 2016

ÖAW AUSTRIAN ACADEMY OF SCIENCES
ACOUSTICS RESEARCH INSTITUTE
ARI

kunst uni graz
Institute of Electronic Music and Acoustics
iem

## Abstract

As an interdisciplinary field, hearing research combines scientific and technical methods to gain new knowledge and understanding. Thereby modeling in an essential tool. Collected data from perceptual experiments is processed in logical structures to achieve a better understanding of the human hearing.
The focus in this project is directed on modeling forced-choice experiments (specifically alternative forced choice AFC). The aim is to include the process of the experiment in the Auditory Modeling Toolbox (AMT) for Matlab and to test this experimental procedure with a model. In this work a binaural signal detection model described by [Breebaart et al., 2001a] is implemented and tested for various listening conditions.
Finally, all AFC-functions are designed in such a way that they generally can be used for other models in the AMT.

## Zusammenfassung

Hörforschung ist ein interdisziplinäres Gebiet, in dem naturwissenschaftliche und technische Methoden verbunden werden, um zu neuen Erkenntnissen zu gelangen. Ein wichtiges Werkzeug der Hörforschung ist die Modellierung. Dabei werden die in Wahrnehmungsexperimenten gewonnenen Daten in logischen Strukturen verarbeitet, mit dem Ziel ein besseres Verständnis über den Hörapparat zu erhalten.
Der Fokus in diesem Projekt richtet sich auf die Modellierung psychoakustischer Experimente mit erzwungener Wahl (engl: alternative forced choice; AFC). Ziel ist es, diesen experimentellen Ablauf in die Auditory Modeling Toolbox (AMT) für Matlab zu integrieren und die Expermentalprozedur mit einem Modell zu testen. In dieser Arbeit werden dafür binaurale Maskierungsexperimente nach dem Modell von [Breebaart et al., 2001a] implementiert und für verschiedene Hörbedingungen getestet.
Schlussendlich werden die AFC-Funktionen so aufbereitet, dass sie als eine allgemeine Struktur innerhalb der AMT für andere Modelle verwendet werden können.

# Contents

# 1 Introduction

## 1.1 Motivation

The auditory sense helps people to orient in everyday life. Therefore research deals with the processing of the audio information in our ears and brain. With the help of psychoacoustic experiments the impact of various parameters on human perception is analyzed. A commonly used testing approach is a forced-choice paradigm in which the subject is presented with two or more alternatives from which he/she must choose a response.
Psychoacoustic experiments can result in huge expenses in time and money and therefore hand we need another approach to understand particular aspects of hearing and to analyze the results of the experiments. That is where auditory models come into play.
In this project the focus lies on modeling forced-choice experiments for the Auditory Modeling Toolbox (AMT) for Matlab. Furthermore the experimental procedure is tested with a binaural signal detection model. Already implemented components are examined and if necessary revised, while the new parts are evaluated and compared with experimental data as well as result of the original model implementation by [Breebaart et al., 2001a].

## 1.2 The Human Auditory System

The human auditory system is described in the following chapter. The overview is based on [Larsen, 2010], [Laback, 2010] and [Breebaart et al., 2001a]. The interested reader is referred to related literature.

Figure 1 shows the anatomy and the structures of the peripheral auditory system.

### 1.2.1 Outer and Middle Ear

The pinna and the ear canal form the outer ear. The structure of the pinna is different for every person and for safety reasons the ear canal is lightly bent. The sound which enters the outer ear consists of direct sound and reflected sound waves of the pinna head and upper torso. The pinna behaves as an acoustic cone which focuses the sound waves at higher frequencies. Pinna and ear canal produce a resonance between 1.5 and 5 kHz. For frequencies higher than 4 - 5 kHz ($\lambda <=$ dimensions of the pinna) a filtering of the sound wave occurs which is important to localize the elevation of sounds.

Via the tympanic membrane, or also called eardrum, the sound waves enter the middle ear. It consists of three small ossicles (malleus, incus and stapes) which serve to make an energy efficient transduction between the outer ear (air) and the inner ear (filled with fluid). In total there is an amplification with a factor of 50 of the pressure (force per area). The most efficient transmission happens at 0.5 - 4 kHz.

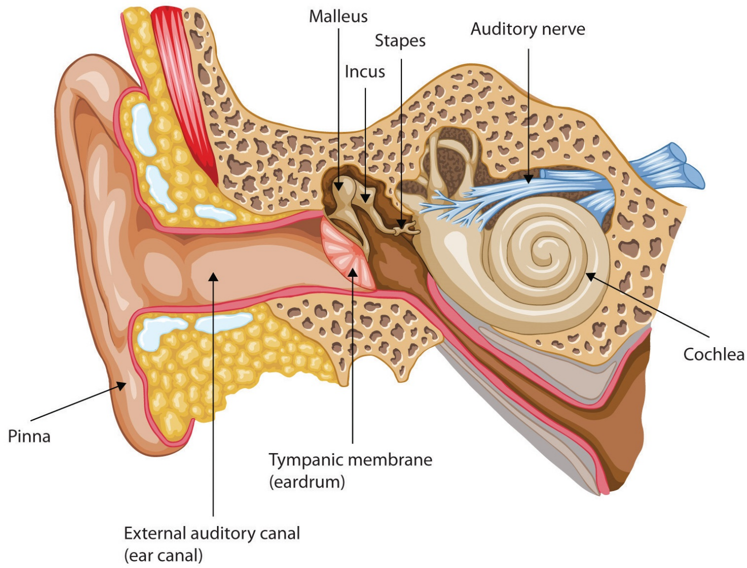


Figure 1: Anatomy of the human ear [Stangor, 2011]

### 1.2.2 Inner Ear

The inner ear consists of the vestibular balance system and the cochlear. The spiral structure has a length of 35 mm and is composed of three components (see fig. 2):

- Scala Vestibuli filled with perilymph: The stapes transfers movements via the oval window to the scala vestibuli.
- Scala Tympani filled with perilymph: The helicotrema at the end of the cochlear connects the scala vestibuli with the scala tympani.
- Scala Media filled with endolymph: This chamber lies in between the two other chambers separated by the basilar membrane.

Due to mechanical vibrations at the oval window the fluids begin to move and as a consequence the basilar membrane vibrates. The vibration appears as a traveling wave along the membrane. High frequencies have a maximum at the oval window (base of the cochlear) and low frequencies at the helicotrema. The amplitude increases from the oval window untill it reaches the maximum and increases abruptly. For a certain position of the basilar membrane one frequency causes the peak response. This frequency is referred to as the center frequency (CF).The CFs are spaced logarithmically with respect to the frequency. The basilar membrane has a frequency selectivity due to the lack of an infinite frequency resolution. The bandwidth of frequencies around a given CF that cannot be resolved in the auditory system is called critical band. The vibration of the traveling wave is passed on the hair cells (inner hair cells). Due to displacement neuronal impulses (spikes) are sent to the auditory nerve. In contrast to the inner (internal) hair cells the outer (external) hair cells can be controlled actively by the brain stem and therefore act as "amplifier for the cochlear".

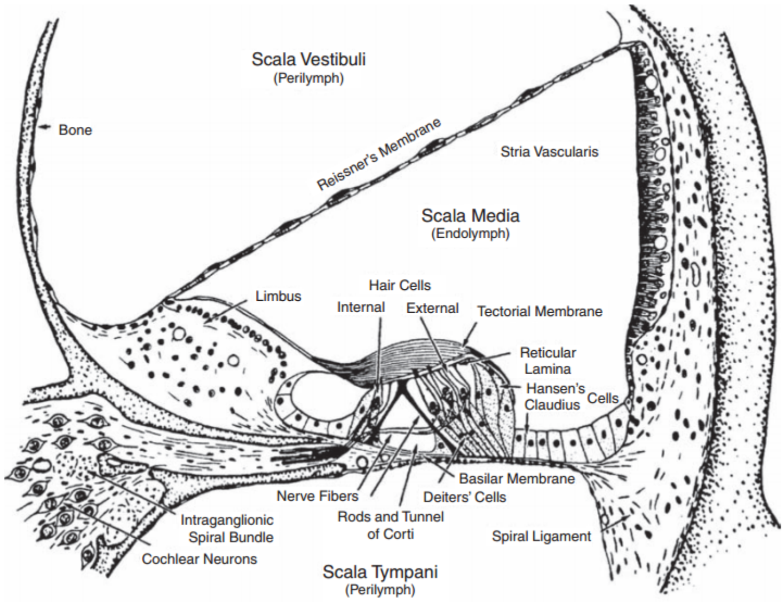


Figure 2: Anatomy of the cochlear [Gelfand, 2009]

### 1.2.3 Neuronal Structure in the Brain

For binaural hearing the central auditory system in our brain is essential. Important information for localization including the interaural time difference (ITD) and the interaural level difference (ILD) can be explained with cells of the superior olive.
An important theory of hearing is the equalization-cancellation (EC) theory [Durlach, 1972]. This theory is used for the prediction of binaural masking level differences. E.g.: Broadband noise, which is in phase to both ears, and pure tones out of phase to each ear are presented simultaneously ($N_0S_\pi$). The E-process transforms the stimuli presented to the two ears to eliminate masking components by transforming and equalizing. The C-process subtracts the stimulus in one ear from the stimulus in the other ear. Consequently part of the masker energy cannot be canceled.
A subgroup of cells in the lateral superior olive is excited by the ipsilateral ear and inhibited by the contraleral ear. These cells are therefore called excitation-inhibition (EI) cells.

## 1.3 AMToolbox

The Auditory Modeling Toolbox, AMToolbox, is a Matlab/Octave toolbox for developing and applying auditory perceptual models with a particular focus on binaural models [Søndergaard and Majdak, 2013].

The AMToolbox is published under the GNU general public license version 3 (`www.gnu.org/licenses/gpl.html`), a free and open source license that guarantees the freedom to share and modify it for all its users and all its future versions. The AMToolbox, including its source code, is available from SourgeForge (`amtoolbox.sourceforge.net/` and `sourgeforge.net/projects/amtoolbox`).

AMToolbox is build on top of the large time-frequency-analysis toolbox, LTFAT. LTFAT is a Matlab/Octave toolbox for time-frequency analysis and multichannel digital signal processing [Søndergaard et al., 2012].

Aim of this project is to implement an alternative forced choice procedure in the AMToolbox. This procedure shall be evaluated by a binaural model (in this case the Breebaart model), but shall be as general as possible that later on other models can be evaluated as well.

Parts of the Breebaart model [Breebaart et al., 2001a] have already been implemented. The model stages are separately accessible and could be used with a few improvements.

In this project the existing parts are evaluated and new functions are contributed.

# 2 Alternative Forced Choice Experiments

The information in this chapter is based on [Gelfand, 2009], [Laback, 2010] and [Levitt, 1971].

## 2.1 Experimental Procedure

A commonly used testing approach is a forced choice paradigm in which the subject is presented with two or more alternatives from which he must choose a response. With the Breebaart model we want to find out whether a subject can hear a tone in the presence of noise. In a two-alternative forced choice (2AFC) method, two stimuli are presented successively, whereby only one interval contains the target. After listening to the two intervals the subject must decide in which interval the target was present. The performance of the subject can vary between 50 % (target not detectable) and 100 % (target is very loud).

In modern psychophysics the AFC procedure is the method which is mostly used, because a potential bias has no influence on the performance.

### 2.1.1 Strategy of Detection

To detect a target the subject must decide whether the stimulation in the auditory system is e.g. due to noise alone (N = reference) or due to signal plus noise (SN = target). The representation of this process can be shown by distributions along a decision axis (see fig. 3). Here the x-axis can represent the energy contained in a noise and in the noise plus signal interval, the number of neuronal spikes in the auditory nerve,... . The y-axis denotes the probability of an event occurring. The separation between between the noise only and the signal plus noise curves becomes a measure of sensitivity measured with the parameter $dprime(d')$:

$$d' = \frac{\bar{x}_{SN} - \bar{x}_N}{\sigma} \tag{1}$$

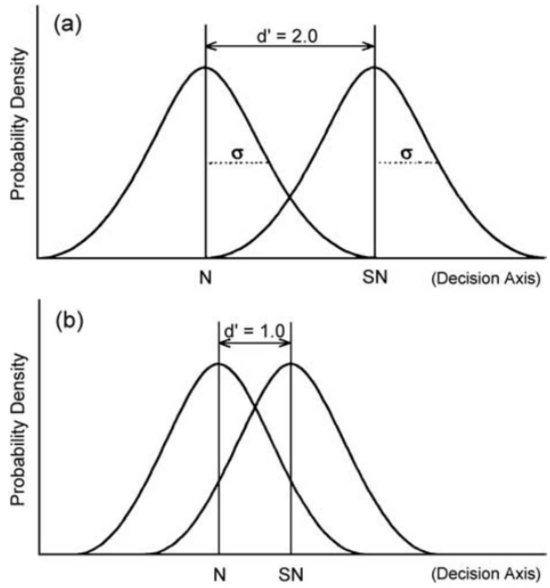


Figure 3: Distributions along a decision axis [Gelfand, 2009]

The subject selects the interval (yes = target in this interval, no = target not in this interval) depending on the subjects criterion on responding. The vertical line in figure 4 represents this criterion. Whenever the value is above the line the subject says yes. The criterion's value depends on three factors:

1. The probability ratio $\beta$ between N and SN curve is affected by the overlap.
2. The number of intervals adjusts the criterion $\beta$. If the signal is only presented one-third of a time, the subject adapts a stricter criterion compared to a fifty-fifty basis.
3. The instructions given to the subject change the criterion $\beta$. If the subject is told he/she shall be really strict when saying yes, the vertical line in figure 4 would be further right.

## 2.2 Adaptive Procedures

In this psychoacoustic method the level of a stimulus changes depending on the response of the subject. It tends to converge upon the threshold level and maximizes the efficiency because most of the trials are near the to the threshold. No prior knowledge of threshold is required. The step sizes are large at first and get smaller as the threshold is approached.

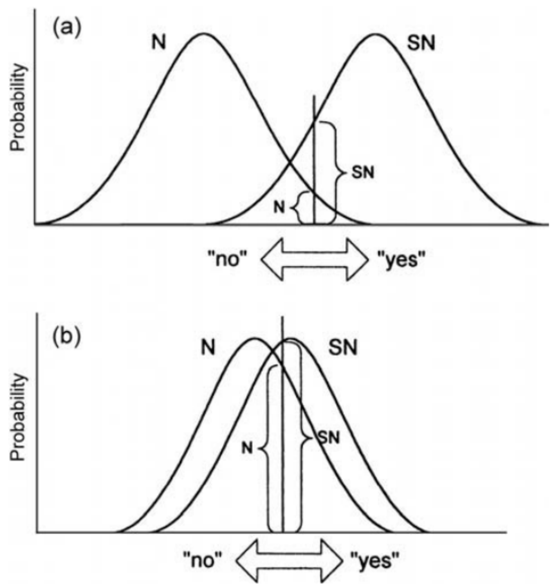


Figure 4: Distributions along a decision axis with criterion points (vertical lines) [Gelfand, 2009]

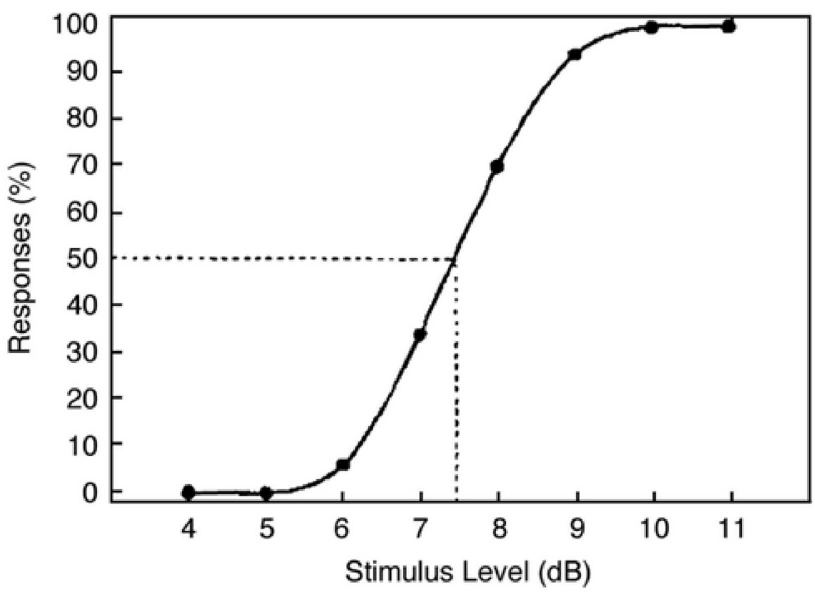


Figure 5: Psychometric function [Gelfand, 2009]

### 2.2.1 Simple Up Down Method

Here the stimulus increases when the response of the subject is wrong and decreases when the response is correct. Therefore a descending run starts with a positive response continues downward till there is a negative response and vice versa. The also called staircase method converges at the 50 % point of the psychometric function (see fig. 5). The procedure runs for some reversals and the arithmetic, quadratic or geometric mean or the median is calculated by taking a number of the last reversals. One problem is that for the simple up down method the probability of a positive response is the same as of a negative one.

### 2.2.2 Transformed Up and Down Procedure

To get a convergence at a higher point of the psychometric function (see fig. 5) the up down rule can be modified. If we are interested in a point around 70 % the following rules are established:

- Up rule: (o) or (+,o)
- Down rule: (+,+)

The probability of two successive positive responses is $p^2$. The up down procedure converges at a point where the rules for up and down have the same probabilities (0.5). Therefore the estimated point $p$ on the psychometric function is 0.707. A schematic representation of the experimental progress can be seen in fig. 6.

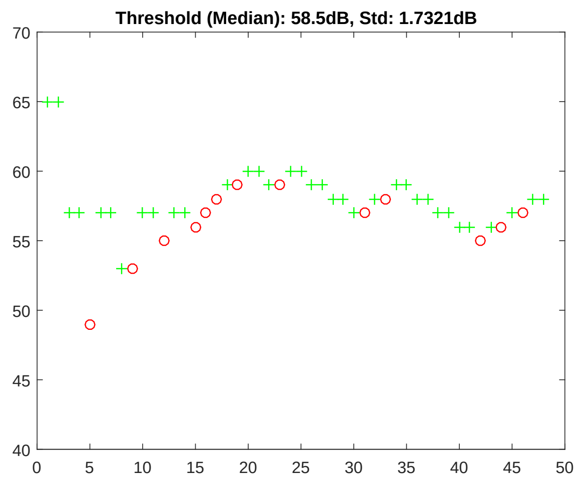


Figure 6: Experimental progress

## 2.3 Emulation of an AFC Experiment in AMT

### 2.3.1 General

To run an alternative forced choice experiment in AMToolbox the following parameters have to be initialized first:

- experimental parameters
- model parameters
- signal parameters
- decision parameters

All set parameters are stored in a structure. Afterwards the experiment can be be started. It has the following work flow (see fig. 7)

- start experiment with run
- call signal generation function in each interval
- call model function in each interval
- call decision function
- change experimental parameter
- start again by calling the signal generation function

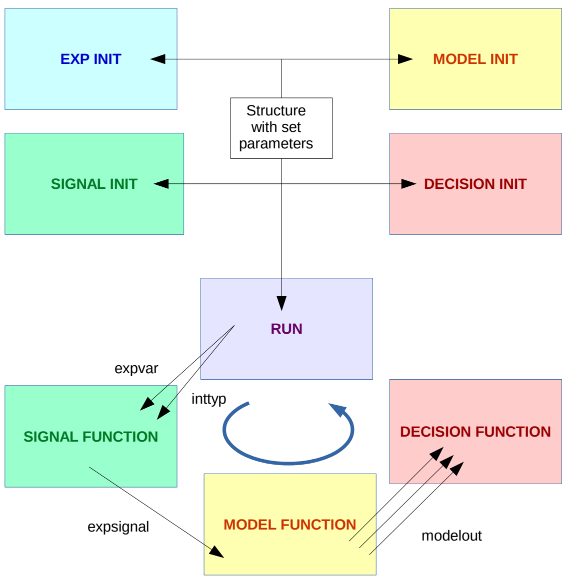


Figure 7: Emulation of an AFC experiment in AMT

### 2.3.2 Initialization of Parameters

**Initialization for Experiment**

***par=emuafcexp('expinit',[ ],exp)*** initilizes the experiment with key-value pairs provided in a cell array ***exp***. The following pairs are required:

- 'intnum',intnum
  number of intervals in a trial, e.g. 3 sets up a 3-afc experiment.
- 'rule',down_up
  vector with down-up-rule, e.g. [2 1] sets up a 2-down, 1-up
- 'expvarstepstart',expvarstepstart
  step size of the experimental variable at the beginning of the experiment
- 'expvarsteprule',factor_turns
  vector with a factor and number of turn arounds, The factor affects the step size of the experimental variable after the number of turn arounds, e.g. [0.5 2] multiplies the stepsize by 0.5 after two turn arounds.
- 'stepmin',min_threshturn
  vector with minimal step size and number of turn arounds after reaching that minimal step size for the threshold calculation, e.g. [1 8] means that after reaching the step size 1, the experiment will continue for another 8 reversals before terminating.

All key-value pairs can be defined in a cell beforehand: e.g.
exp = {'intnum', 3, 'rule', [2 1], 'expvarstepstart', 8, 'expvarsteprule', [0.5 2],... 'stepmin', [1 8],'expvarstart', 65};

To check the set parameters use ***par.exp***.

**Initialization for Signal Generator**

***par=emuafcexp('signalinit',par,sig)*** initializes the signal generator. It is creating signals for the model with key-value pairs provided in the cell array ***sig***. The signal generator is called with those parameters in each trial of the experiment. The already set parameters from other initializations are stored in ***par***. Up to 15 input parameters are supported. One of the inputs must be ***inttyp***: In each experimental interval, this input will be replaced by ***target*** or ***reference*** depending on the interval type. One of the inputs must be ***expvar***: In each trial, this input will be replaced by the value of the experimental variable. The following pairs are required:

- 'name',name
  string which defines the name of the signal generation
- 'inputX',inputX
  input parameter X needed for the signal generator, up to 10 parameters possible

All key-value pairs can be defined in a cell beforehand: e.g.
sig = {'name', 'breebaart2001siggen', 'input1', 'inttyp', 'input2', 500,...
'input3', 'expvar', 'input4', 0.3, 'input5', pi, 'input6', 5,'input7', 65, 'input8', 0.4,...
'input9', 0, 'input10', 0.05, 'input11', 32000};

To check the set parameters use ***par.signal***.

### Initialization of Model called in each Interval

***par=emuafcexp('modelinit',par,mod)*** initializes the model called in each interval with the key-value pairs provided in ***mod***. Up to 10 input parameters are supported. One of the inputs must contain the keyword ***expsignal***. This keyword is replaced in the ***run*** routine with the output of the signal generation function:

- 'name',name
  string which defines the name of the model function
- 'inputX',intputX
  input parameter X needed by the model
- 'outputs',outputs
  indicies of used model outputs for the decision, e.g. [1 2 6]: output 1,2 and 6 used

All key-value pairs can be defined in a cell beforehand: e.g.
mod = {'name', 'breebaart2001preproc', 'input1', 'expsignal', 'input2', 32000,... 'input3', 0, 'input4', 0, 'outputs', [1 3 4]};

To see how the model is called during a run, the initialization returns a structure entry called ***par.callstring.model*** where the string, with which the model function is called, is stored. To check the set parameters use ***par.model***.

### Initialization of Decision Stage called in each Trial

***par=emuafcexp('decisioninit',par,dec)*** initializes the decision stage of the experiment with key-value pairs provided in ***dec***. Up to 10 input parameters are supported. All inputs containing the keyword ***modelout*** are replaced with the outputs of the model function during an experimental run. Therefore the number of inputs with the keyword ***modelout*** must be equal to number of ***outputs*** defined in ***modelinit***. An output of the model function contains a cell with an entry for each interval. E.g. param1{1} contains the first output of the model function of the first interval and param3{2} contains the third output of the model function of the second interval. Therefore the decision function must be implemented so that the inputs of the decision function are cells with entries for each interval. Following parameters are required:

- 'name',name
  string which defines the name of the decision function

- 'inputX',intputX
  input parameter X needed by the decision function

All key-value pairs can be defined in a cell beforehand: e.g.
dec = {'name', 'breebaartcentralproc', 'input1', 'modelout', 'input2', 'modelout', 'input3', 'modelout', 'input4', 'lbr'};

To check the set parameters use ***par.decision***.

### 2.3.3 Running the Experiment

After the initialization, the experiment can be started by ***out = emuafcexp('run',par)***. The ***out*** vector contains the experiment output. ***out(:,1)*** is the median threshold of the values of number of turns defined in the initialization of the experiment. ***out(:,2)*** is the standard deviation of the values of number of turns defined in the initialization of the experiment. ***out(:,3:end)*** provides the values of the individual experimental variable used in the trial (see fig. 8).

***out = emuafcexp('run',par,'plot')*** runs the experiment and plots the experimental progress. A plot of a progress can be seen in figure 9.

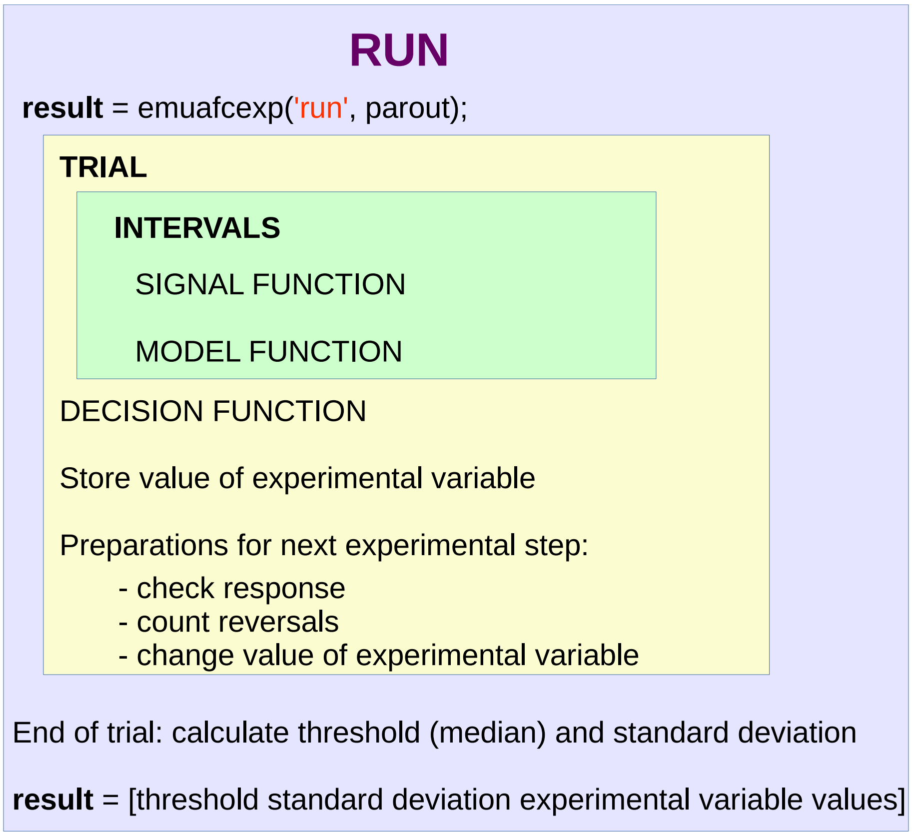


Figure 8: Schematic representation of a trial

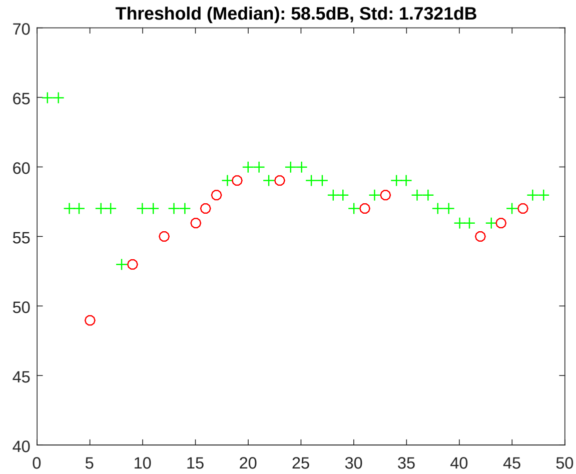


Figure 9: Experimental progress

In summary (see also fig. 7), an experimental signal is generated by the signal generation function for each interval. Beforehand, the keywords ***expvar*** and ***inttyp*** are replaced by the current value of the experimental variable and the word target or reference depending on the interval type. The output of the signal generation function replaces the keyword ***expsignal***. Afterwards the model function simulates the human hearing and the keywords ***modelout*** are exchanged with the output of the model functions. Following the computing of all internal representations for every interval a decision function is called. It decides in which interval the target was heard.

# 3 Example

To evaluate the experimental procedure of alternative-forced-choice an example is needed. Therefore a binaural signal detection model described by [Breebaart et al., 2001a] is implemented and tested for various listening conditions.

## 3.1 Breebaart Model

The information in this chapter is based on [Breebaart et al., 2001a], [Breebaart, 2001] and [Larsen, 2010]. In [Breebaart et al., 2001a] an auditory model is described which is able to detect binaural signals. The model is evaluated in an alternative forced choice procedure. In the model there are different stages for monaural and binaural processing and a final decision state. In short the model consists of three different stages (see fig. 10):

1. a peripheral processor which models the outer middle and inner ear
2. a binaural processor which computes the differences between the two ears
3. a central processor which analyzes the output of the binaural processor and gives a decision

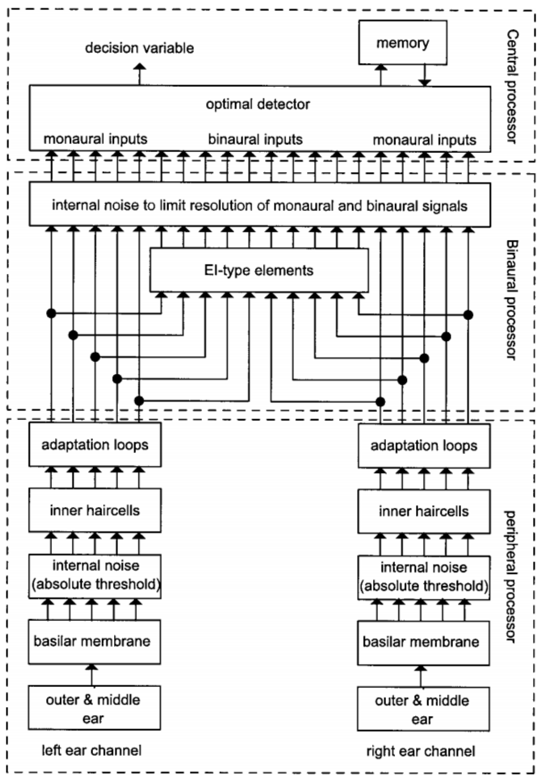


Figure 10: Stages of the model [Breebaart et al., 2001a]

### 3.1.1 Peripheral Processor

The peripheral processor in the model described by [Breebaart et al., 2001a] consists of 5 parts:

1. a outer and middle ear transfer function
2. a model of the basilar membrane
3. the incorporation of an absolute threshold
4. the signal processing of the inner hair cells
5. the adaptation of the auditory nerve

**Outer and Middle Ear**
The transfer function of the outer and middle ear is modeled as a bandpass filter with cutoff frequencies of 1 kHz and 4 kHz. The input-output relation of the time-domain filter is as follows:

$$y[n] = (1-q)rx[n] - (1-q)rx[n-1] + (q+r)(y[n-1] - qry[n-2] \tag{2}$$

$$q = 2 - cos\left(\frac{2\pi 4000}{f_s}\right) - \sqrt{cos\left(\frac{2\pi 4000}{f_s} - 2\right)^2 - 1} \tag{3}$$

$$r = 2 - cos\left(\frac{2\pi 1000}{f_s}\right) - \sqrt{cos\left(\frac{2\pi 1000}{f_s} - 2\right)^2 - 1} \tag{4}$$

[Breebaart et al., 2001a] mentioned that only headphone experiments are modeled and evaluated and therefore HRTF (head-related transfer function) filtering is not included.

**Cochlear with Basilar Membrane**
The basilar membrane has a frequency selectivity. It is modeled by a gammatone filterbank [Patterson et al., 1987]. The bandwidth of the filters is corresponding to the equivalent rectangular bandwidth (ERB) [Glasberg and Moore, 1990]. The filters are spaced with two filters per ERB.

**Absolute Threshold**
The absolute threshold of human hearing is incorporated as an independent Gaussian noise which is added to each signal (see [Breebaart et al., 2001a] section IV (3) for details).

**Inner Hair Cells**
The firing of the inner hair cells is modeled by a half-wave rectifier and a fifth-order low-pass filter. The first step simulates that only a deflection of the hair cells cilia leads to action potential firing in the Auditory nerve. The second step, a low-pass filter with a -3 dB cutoff frequency of 770 Hz simulates a decrease of phase locking at higher frequencies [Larsen, 2010].

**Auditory Nerve**
The adaptation of the auditory nerve is included with a chain of five adaptation loops [Dau et al., 1996]. In the steady state part the input-output characteristic is almost logarithmic. The output is expressed in model units (MU). Sound pressure levels of 0 to 100 dB are scaled to 0 to 100 MU. The time constants of the adaptation loops are linear spaced and have the following values: 5, 129, 253, 376 and 500 ms.

### 3.1.2 Binaural Processor

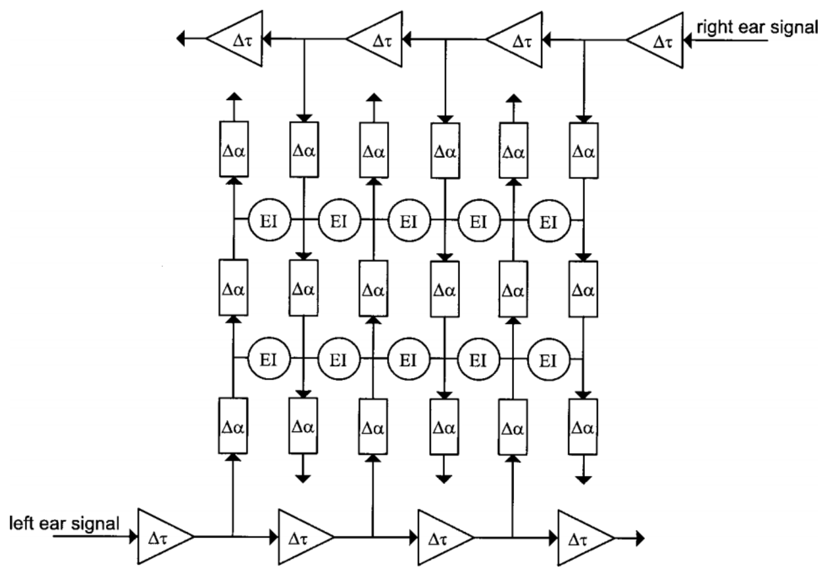


Figure 11: Structure of the binaural processor [Breebaart et al., 2001a]

The binaural processor compares signals from corresponding auditory channels with excitation-inhibition (EI) elements. Each element has a characteristic ITD and a characteristic ILD. The first and last line in figure 11 carry in the internal representations from corresponding auditory channels from the right and the left ears. The delays result in a time lag of the right-ear signal at the left side and vice versa. In addition a chain of attenuators is added. This leads to an equalization. Afterwards the signals are subtracted and the result is squared:

$$E(i,t,\tau,\alpha) = (10^{\alpha/40}L_i(t+\tau/2) - 10^{-\alpha/40}R_o(t-\tau/2))^2 \quad (5)$$

with $i$ = frequency channel, $t$ = time in seconds, $\tau$ = time adjustment in seconds and $\alpha$ = level adjustment in dB

To include a finite binaural resolution the output is processed by a sliding temporal integrator:

$$E'(i,t,\tau,\alpha) = \int_{-\infty}^{\infty} E(i,t+t_{int},\tau,\alpha)w(t_int)dt_{int}, \quad (6)$$

with

$$w(t) = \frac{exp(-|t|/c)}{2c} \quad (7)$$

For a relation of the output to the interaural correlation of the ear signals a logarithmic compression function is applied:

$$E''(i,t,\tau,\alpha) = ap(\tau)ln(bE'(i,t,\tau,\alpha)+1) \quad (8)$$

The constants a and b describe the sensitivity for binaural differences. The weighting function $p(\tau)$ includes the fact that cells with larger characteristic interaural delays are less frequent than cells with smaller characteristic delays.

$$p(\tau) = 10^{-|\tau|/5} \quad (9)$$

with $\tau$ in ms.

The interested reader is referred to Appendix A in [Breebaart et al., 2001a] for more information.

### 3.1.3 Central Processor

Monaural and binaural representations $E''(i)[n]$ enter the central processor (decision stage) in parallel, whereby the monaural representations consist of the output of the adaptation loops, which is low-pass filtered by a double-sided exponential window with a time constant of 100 ms. There is only one binaural input per auditory filter. Therefore the values for $\alpha$ and $\tau$ are restricted to one value only where the binaural processing has a minimum point in the EI activity for a masker-alone interval.

The central processor develops a template $\bar{E}(i)[n]$. This template contains the mean of all processed masker alone intervals $\beta$:

$$\bar{E}(i)[n] = \frac{1}{\beta}\sum_{\beta} E''(i)[n] \quad (10)$$

The central processor estimates the variance:

$$\sigma^2(i)[n] = \sigma_N^2 + \frac{1}{\beta}\sum_{\beta}(E''(i)[n])^2 - \bar{E}^2(i)[n] \tag{11}$$

with ($\sigma_N^2 = 1$) being the variance of the internal noise .

Furthermore the mean distance between the template $\bar{E}(i)[n]$ and the average signal interval is computed:

$$\mu(i)[n] = \frac{1}{\mu}\sum_{\mu}(E''(i)[n]) - \bar{E}(i)[n] \tag{12}$$

In the end the weighted difference between template and actual representation:

$$U = \sum_i \sum_n \frac{\mu(i)[n]}{\sigma^2(i)[n]}(E''(i)[n] - \bar{E}(i)[n]) + N_U \tag{13}$$

with the variance of $N_U$:

$$\sigma_{NU}^2 = \sigma_N^2 \sum_i \sum_n \frac{\mu^2(i)[n]}{\sigma^4(i)[n]} \tag{14}$$

The central processor will choose the interval with the highest U value.

## 3.2 Implementation of the Breebaart Model

### 3.2.1 Peripheral Processor

The Maltab functions for the elements of the peripheral processor can be run separately or one can call a function and get the output of the periphal processor as a whole[1].

**Outer and Middle Ear**

This filter is new implemented for AMT[2] as described in 3.1.1.

**Cochlear with Basilar Membrane**

Due to a lack of declaration in [Breebaart et al., 2001a] for the double spacing of the filters and the discrepancy with [Breebaart, 2001][page 92] the standard spacing of 1 ERB is used. In the paper no start and stop frequencies are given. Therefore the lowest frequency in the filterbank is 80 Hz and the highest 8000 Hz as it has been already implemented in AMT. So in the end there are 31 filters. This auditory filterbank has already been implemented in the AMT and can be applied[3].

[1] use breebaart2001preproc.m in AMT
[2] use breebaart2001outmiddlefilter.m in AMT
[3] use auditoryfilterbank.m in AMT

**Inner Hair Cells**
The model of the inner hair cells has already been implemented in the AMT and can be applied with the '$methodname$' of $'ihc\_breebaart'$ to use the correct parameters[4].

**Absolute Threshold and Auditory Nerve**
In the AMT the absolute threshold is implemented as in [Dau et al., 1996]. The minimal value at the input of the adaptation stage is limited by a constant value. The model of the adaptation loops has already been implemented in the AMT and can be applied with a flag of $'adt\_breebaart'$ to use the right parameters[5].

**Output of the Peripheral Processor**
An example of the output of the peripheral processor is shown in fig. 12 - 13. Fig. 12 presents the output of the peripheral processor by Breebaart (cf. [Breebaart et al., 2001a] fig. 2). Fig. 13 shows the output of our implementation of the model[6] of the peripheral processor for a 500 Hz and 4000 Hz tone of 100 ms duration with an overall level of 70 dB SPL (The rms value of 1 corresponds to a range of 100 dB). For the other 100 ms zeros were added to show the fade out as well.

The outputs have the same shape but the values are a bit different. This can be due to some of the changes to the original model. Also fig. 12 shows the output for filters centered at 500 and 4000 Hz and fig. 13 only uses filters from the basilar membrane filterbank which are nearest to 500 and 4000 Hz.

A strong overshoot can be seen at the beginning. In this implementation the overshoot is not limited. In [Dau et al., 1997] a limitation of the onset response is applied.

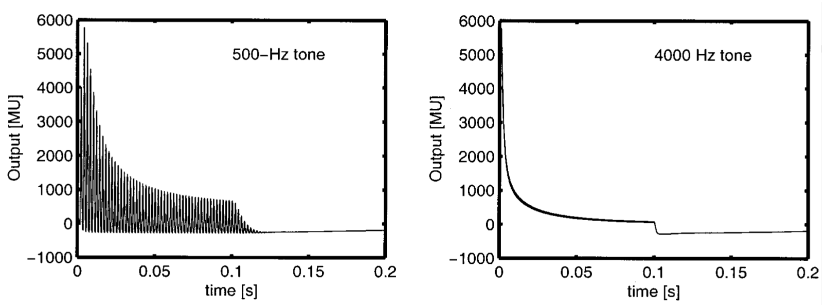


Figure 12: Output of the peripheral procesor [Breebaart et al., 2001a, Figure 2]

[4]use ihcenvelope.m in AMT
[5]use adaptloop.m in AMT
[6]use exp_breebaart2001('afig2') in AMT

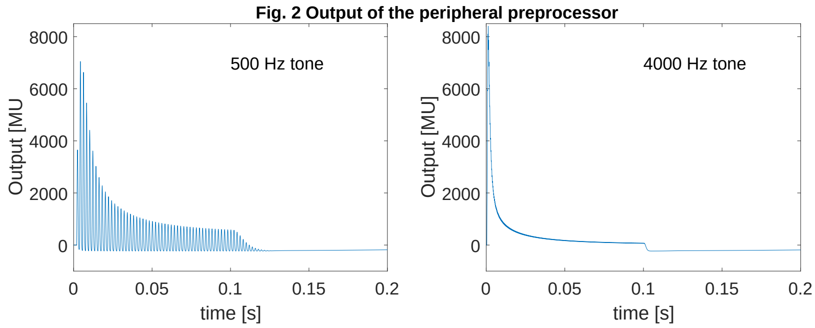


Figure 13: Output of the model of our peripheral processor

### 3.2.2 Binaural Processor

The model of the binaural processor has already been implemented in the AMT[7]. The only change to the binaural processor described in [Breebaart et al., 2001a] is that the weighting function is implemented as in [Larsen, 2010]:

$$p(\tau) = exp(\frac{-|\tau|}{0.0022}) \tag{15}$$

with $\tau$ in seconds.

The constants a and b of equation (8) are set to 0.1 and 0.00002.

#### Output of the Binaural Processor

An example for the output of the binaural processor is shown in fig. 14 - 15. Fig. 14 presents an output of the binaural processor by Breebaart (cf. [Breebaart et al., 2001a] fig. 2). Fig. 15 (left pattern) shows the output of our implementation of the model of the binaural processor[8] for an idealized EI-activity ($p(\tau) = 0$) for a wideband diotic noise with a frequency range of 0 to 4000 Hz and an overall level of 70 dB SPL. The rms value of 1 corresponds to a range of 100 dB. The right pannel shows, how the pattern changes, if a 500 Hz interaurally out-of-phase signal $S\pi$ is added with a level of 50 dB.

The outputs have kind of the same shape but the values are different. This can be due to the different output of the peripheral processor and due to lack of knowledge of the time step shown in the picture. For our plot we use the mean value for a time of 0 to 100 ms, whereas the signal and noise have a duration of 1 second.

[7] use eicell.m in AMT
[8] use exp_breebaart2001('afig6') in AMT

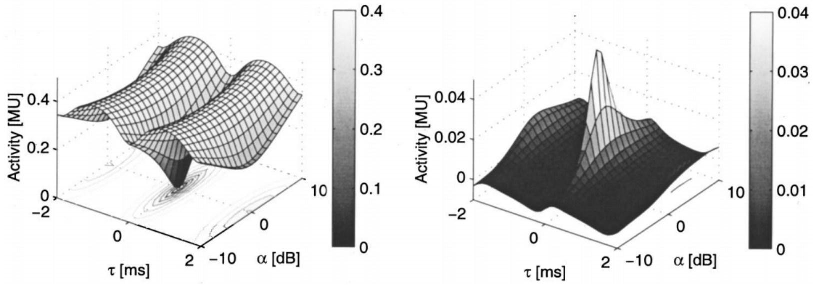


Figure 14: Output of the binaural processor [Breebaart et al., 2001a, Figure 6] Left: Idealized EI-activity for a wideband diotic noise, Right: Change in the activity pattern of the left panel if an out-of-phase signal $S\pi$ is added

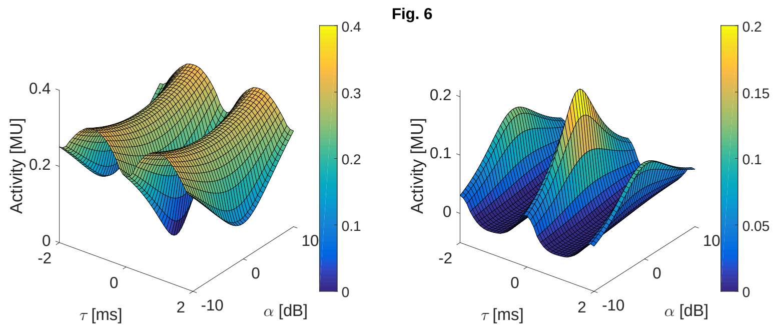


Figure 15: Output of the model of our peripheral processor [Breebaart et al., 2001a, Figure 6]

### 3.2.3 Central Processor

The central processor is new implemented as described in 3.1.3[9]. Due to lack of given information about the combination of the monaural and binaural decision and the multiplication factor for the monaural inputs, I experimented a lot with the factor and combinations and ended up with a factor of 0.0003. This factor is needed because otherwise the monaural inputs would be way to big compared to the binaural input and therefore would dominate the decision.

In the central processor separate templates and decisions are generated for binaural, monaural left and monaural right channels. An input parameter contains a string. Its content determines which templates and U values are generated and used for the decision:

- 'l'
  use left mono channel

[9]use breebaart2001centralproc.m in AMT

- 'r'
  use right mono channel
- 'b'
  use binaural channel, but only if the binaural representation yields a non-zero decision distance
- 'B'
  use binaural channel in any case

In the end the mean of the generated U values is calculated and the interval with the highest U value is used as decision interval.

## 3.3 Evaluation of the Implementation of the Model

In the following chapter results of binaural signal detection (masking noise + sine signal) with our implementation of the model and the AFC procedure are shown. The initialization of parameters for the experiment is performed as in the examples in chapter 2.3.2. Tests showed that the starting level of the experimental variable should be 20 dB higher as the lowest S/N threshold at 125 Hz. This factor is than added to the overall masker level and the result is used as the starting level of the experimental variable.

### 3.3.1 $N_0S_\pi$ Thresholds

First, an interaurally out-of-phase signal ($S_\pi$) is masked by an interaurally in-phase noise ($N_0$) with variable bandwidth ($= N_0S_\pi$). Figure 16 shows the signal to noise thresholds[10]. The bandwidth is varied between 5 Hz and twice the center frequency. The overall masker level is kept constant at 65 dB SPL. The maskers have a duration of 400 ms and the sinusoidal signals of 300 ms. The sinusoids have a center frequency equal to the center frequency of the noise masker. The starting level of the epxerimental variable (level of sinusoid) is 65 dB. Both signal and masker are gated with 50 ms raised cosine ramps. To get one point in this figure, six runs are performed and their mean and standard deviation are plotted. For the next point the masker bandwidth is changed. By changing the center frequency results for another subfigure are generated. Our results (blue) are similar to the model output of Breebaart (red) for most frequencies. At higher frequencies and small bandwidths our model output results in a lower S/N threshold. Therefore we tested different monaural and binaural combinations too. Figure 17 shows the results[11]. According to this plot the use of the binaural input only matches best with the results of Breebaart. The use of the monaural inputs only (lr) leads to a completely false threshold. This is explainable by the fact that the detectable signal has a phase difference and therefore the monaural inputs cannot contribute to the signal detection.

[10] use exp_breebaart2001('bfig3')
[11] use demo_breebaart2001('N0Spi','exact')

Furthermore the model thresholds of Breebaart stays fairly constant up to a certain bandwidth and then declining with 3 dB/octave. At 4000 Hz the difference between model threshold of Breebaart and experimental data is biggest. For more explanations the interested reader is referred to chapter C in [Breebaart et al., 2001b].

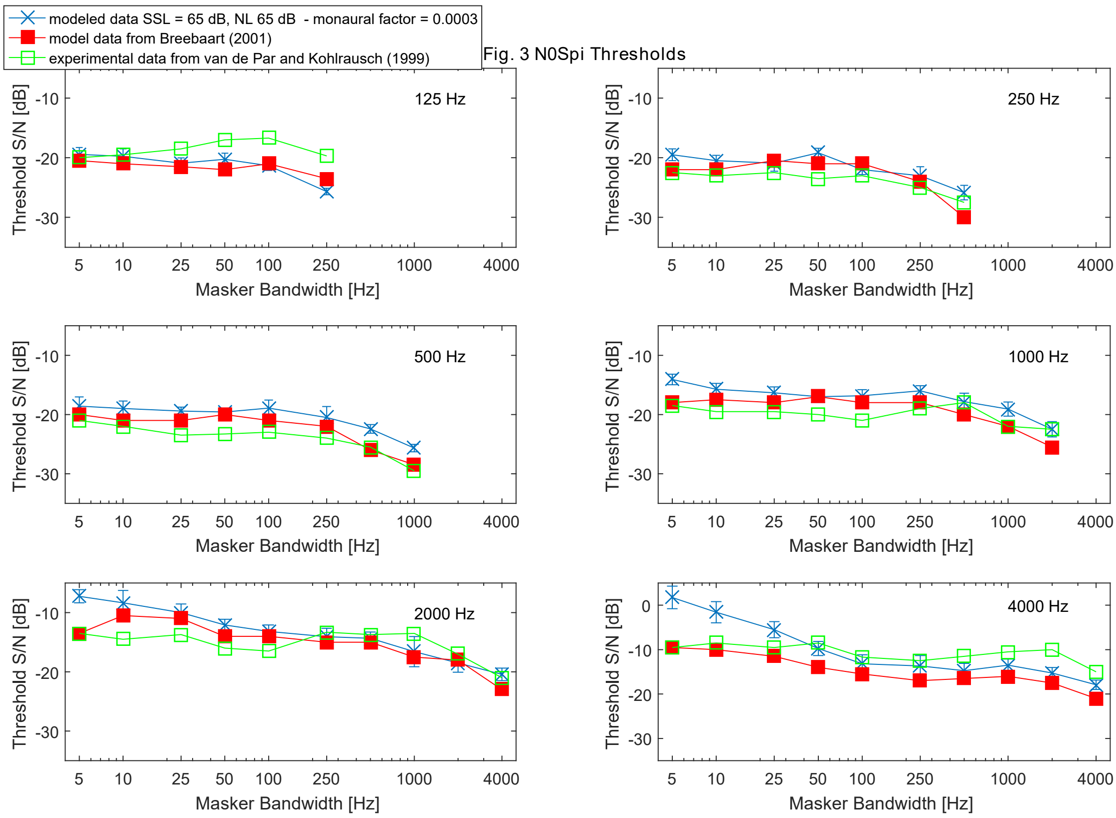


Figure 16: $N_0S_\pi$ Thresholds as a function of masker bandwidth [Breebaart et al., 2001b, Figure 3]. The filled squares are model predictions from [Breebaart et al., 2001b]. The x-symbols are model predictions calculated with our implementation of the Breebaart model for a combination of binaural and monaural decisions. The open squares are experimental data from [van de Par and Kohlrausch, 1999].

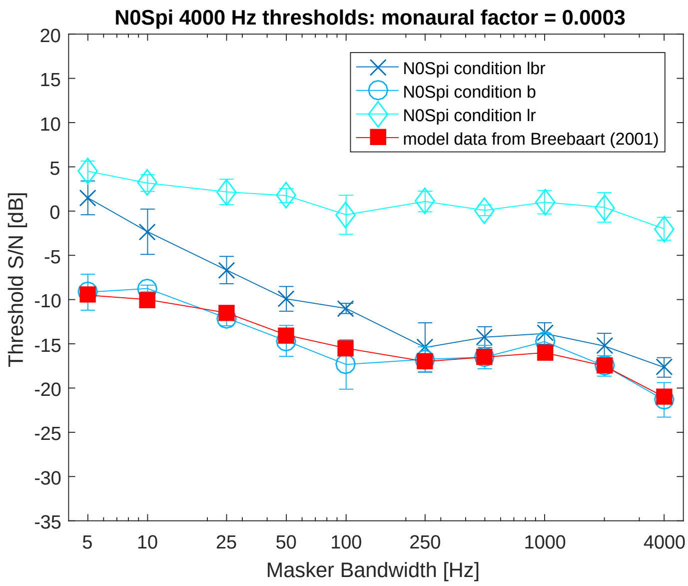


Figure 17: $N_0S_\pi$ Thresholds as a function of masker bandwidth for 4000 Hz and different combinations of monaural and binaural decisions. The filled red squares are model predictions from [Breebaart et al., 2001b]. The blue symbols are model predictions calculated with our implementation of the Breebaart model (x-symbols = combination of binaural and monaural decisions, In this case equal to left an right monaural decision, because the binaural decision in zero. circles = Decision is combination of left, binaural and right result even if the binaural result is zero.).

### 3.3.2 $N_\pi S_0$ Thresholds

Second, an interaurally in-phase signal ($S_0$) is masked by an interaurally out-of-phase noise ($N_\pi$) with variable bandwidth ($= N_\pi S_0$). Figure 18 shows the signal to noise thresholds[12]. The bandwidth is varied between 5 Hz and twice the center frequency. The overall masker level is kept constant at 70 dB SPL. The maskers have a duration of 400 ms and the sinusoidal signals of 300 ms. The sinusoids have a center frequency equal to the center frequency of the noise masker. The starting level of the experimental variable (level of sinusoid) is 85 dB. Both signal and masker are gated with 50 ms raised cosine ramps. At this condition the phase shift is compensated by an internal delay. According to [Breebaart et al., 1998] the optional internal delay equals half the period of the center frequency of the noise. After some tests I ended up with 3.9 ms for 125 Hz, 2 ms for 250 Hz, 1 ms for 500 Hz and 0.5 ms for 1000 Hz. All in all the results could be better. Unknown and not mentioned coefficients in the papers [Breebaart et al., 2001a] and [Breebaart et al., 2001b] cause problems. A modification of the monaural factor, the signal start level or the compensation delay time changes a lot. Also the combination of monaural and binaural input at the central processor plays a role again. Figure 19 shows

[12] use exp_breebaart2001('bfig6')

the results for different combinations at 1000 Hz[13]. It seems as if from 25 to 250 Hz the "lbr" condition is the best and from 500 to 2000 Hz the "b" (binaural only) condition is a good approach.

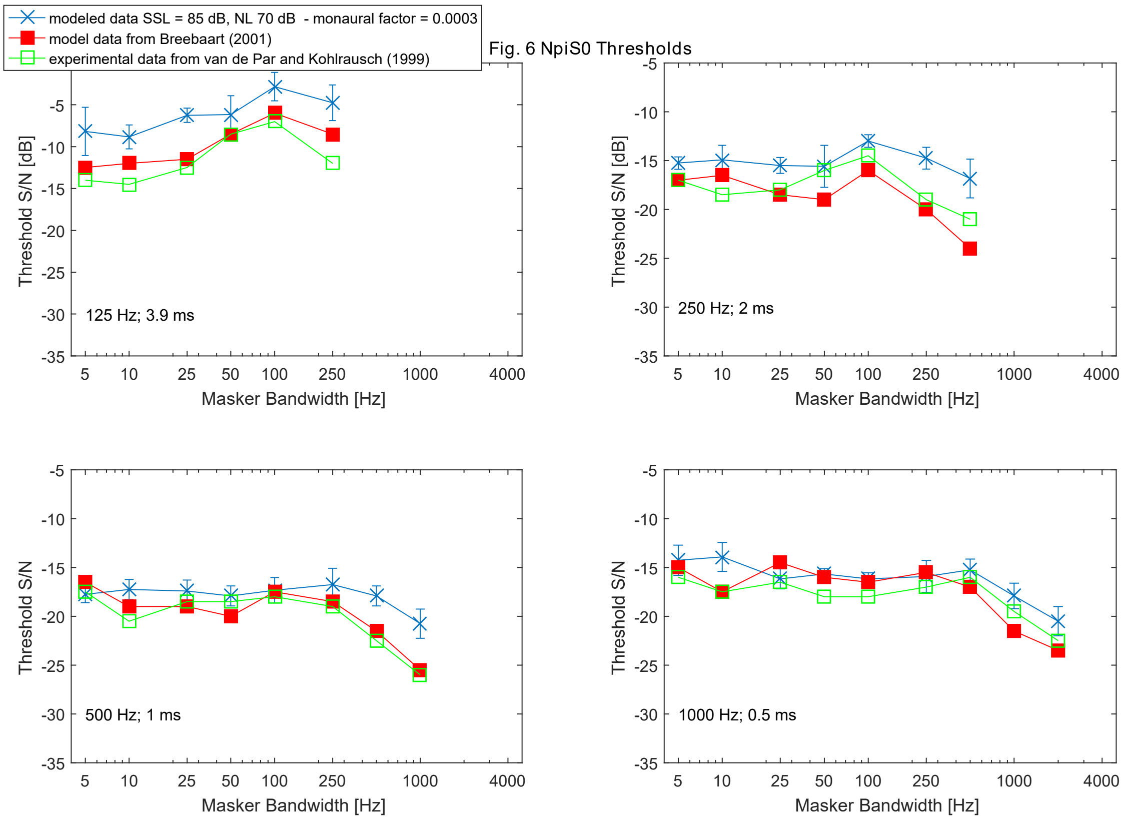


Figure 18: $N_\pi S_0$ Thresholds as a function of masker bandwidth [Breebaart et al., 2001b, Figure 6]. The filled squares are model predictions from [Breebaart et al., 2001b]. The x-symbols are model predictions calculated with our implementation of the Breebaart model for a combination of binaural and monaural decisions. The open squares are experimental data from [van de Par and Kohlrausch, 1999].

[13]use demo_breebaart2001('NpiS0','exact')

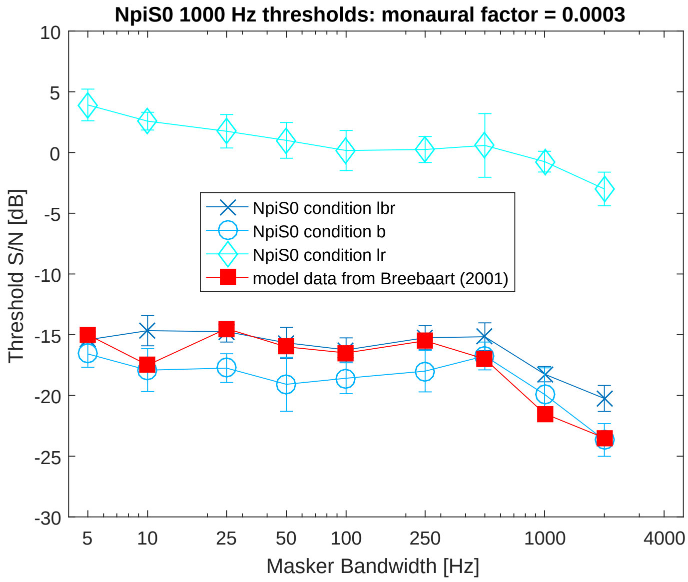


Figure 19: $N_\pi S_0$ Thresholds as a function of masker bandwidth for 1000 Hz and different combinations of monaural and binaural decisions. The filled red squares are model predictions from [Breebaart et al., 2001b]. The blue symbols are model predictions calculated with our implementation of the Breebaart model (x-symbols = combination of binaural and monaural decisions, circles = binaural decisions only, diamonds = left and right monaural decision only).

### 3.3.3 $N_0 S_0$ Thresholds

Third, an interaurally in-phase signal ($S_0$) is masked by an interaurally in-phase noise ($N_0$) with variable bandwidth (= $N_0 S_0$). Figure 20 shows the signal to noise thresholds[14]. The bandwidth is varied between 5 Hz and twice the center frequency. The overall masker level is kept constant at 70 dB SPL. The maskers have a duration of 400 ms and the sinusoidal signals of 300 ms. The sinusoids have a center frequency equal to the center frequency of the noise masker. The starting level of the experimental variable (level of sinusoid) is 90 dB. Both signal and masker are gated with 50 ms raised cosine ramps. Due to lack of appearance of this experimental procedure in [Breebaart et al., 2001b] only our model results and the experimental data from [van de Par and Kohlrausch, 1999] are shown. In this case only the monaural decision is used. The results for 125 Hz are quite okay. That is, because I used this frequency to tune and set the parameters. At other frequencies the variance is huge. Therefore the figure shows that the central processor, like it is implemented in this model, is not a good approach for monaural-only based decisions. According to [van de Par and Kohlrausch, 1999], for a narrow-band monaural condition, combining information across frequency will not lead to any improvement in signal detection. The internal noise plays a big rule too. Our central processor, however, considers all frequency channels. Also figure 21 at 4000 Hz

[14] use exp_breebaart2001('fig1_N0S0_vandepar1999')

shows no improvement by changing the combinations[15]. For the $\boldsymbol{N_0S_0}$ condition the binaural decision is always zero, because the left and the right signals are identically and therefore canceled. In this case "lbr" is equal to "lr" and in the "lBr" condition the zero result for the binaural decision influences the mean decision value. However, there is still no significant difference between these two conditions. A tuning of the monaural factor for different frequencies could lead to more accurate results, but this is not the aim of this project.

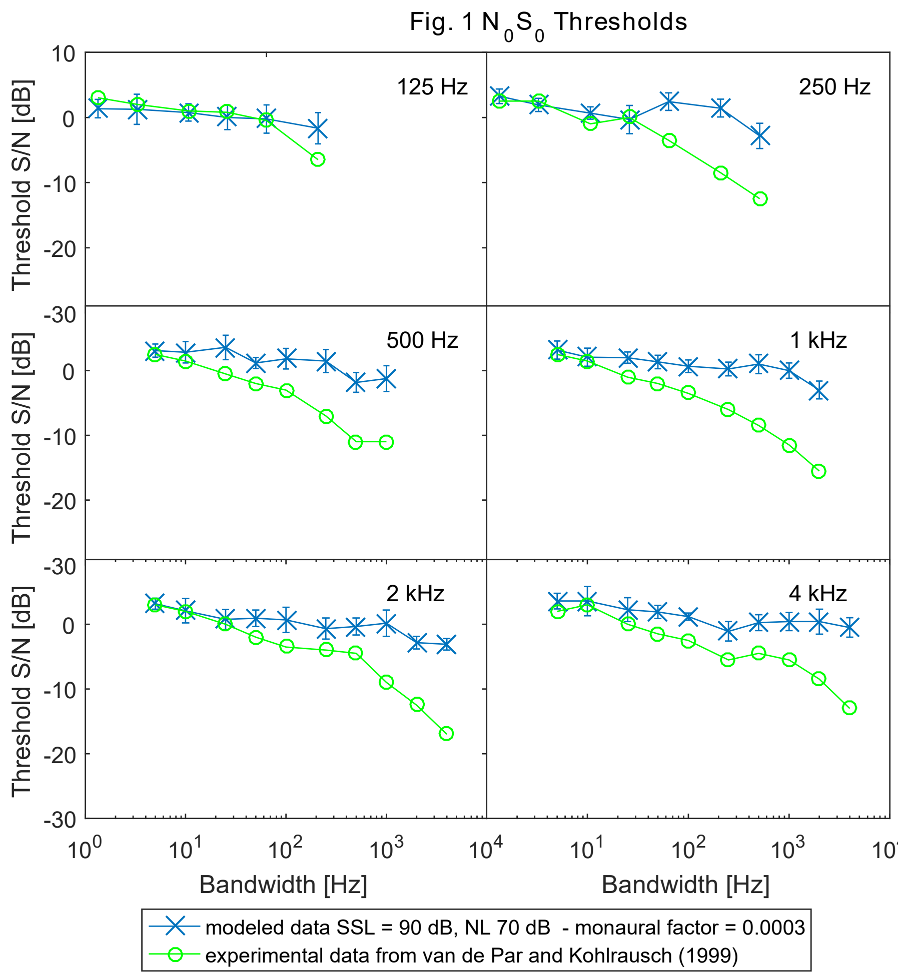


Figure 20: $\boldsymbol{N_0S_0}$ Thresholds as a function of masker bandwidth [van de Par and Kohlrausch, 1999, Figure 1]. The x-symbols are model predictions calculated with our implementation of the Breebaart model for a combination of binaural and monaural decisions. The open circles are experimental data from [van de Par and Kohlrausch, 1999].

[15] use demo_breebaart2001('N0S0','exact')

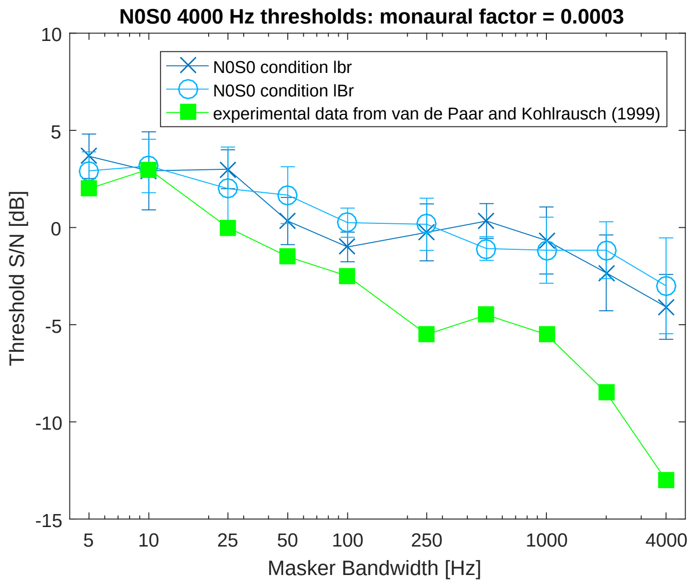


Figure 21: $N_0S_0$ Thresholds as a function of masker bandwidth for 4000 Hz and different combinations of monaural and binaural decisions. The filled green squares are experimental data adapted from [van de Par and Kohlrausch, 1999]. The blue symbols are model predictions calculated with our implementation of the Breebaart model (x-symbols = combination of binaural and monaural decisions, circles = binaural decisions only, diamonds = left and right monaural decision only).

# 4 Conclusion and Discussion

The aim of this project was to include forced-choice experiments in the Auditory Modeling Toolbox (AMT) and test the adaptive run with a model. It was done by adapting already implemented parts and evaluating new written components.

For emulating and running an experiment in AMT one has to initialize the experiment, the signal generation function, the model function as well as the decision function. The command 'run' starts the experimental trial. Furthermore figures from [Breebaart et al., 2001a] and [Breebaart et al., 2001b] can be generated (function exp_breebaart2001) and the impact of the combination of monaural and binaural decisions can be demonstrated (function demo_breebaart). The results of our model are not completely consistent with the output of the [Breebaart et al., 2001a] model and experimental results - a common problem in science.

[Davidson et al., 2009] evaluated the [Breebaart et al., 2001a] model as well. They found that it did not predict a large potion of the variability, but it is capable of predicting thresholds for a multitude of psychophysical tasks. By having a look at their code, which can be downloaded, I found out that they did not use an AFC procedure for their tests and their decision stage is different. They use a so-called decision variable to compare the outcome of different models without computing a running template.

In the future, the AFC procedure should be usable for other models. Therefore the implementation has been done as general as possible.

# References

[Breebaart, 2001] Breebaart, J. (2001). *Modeling Binaural Signal Detection*. Phd thesis, Eidhoven University of Technology.

[Breebaart et al., 1998] Breebaart, J., van de Par, S., and Kohlrausch, A. (1998). Binaural signal detection with phase-shifted and time-delayed noise maskers. *Journal of the Acoustical Society of America*, 103(4):2079–2083.

[Breebaart et al., 2001a] Breebaart, J., van de Par, S., and Kohlrausch, A. (2001a). Binaural processing model based on contralateral inhibition. I. Model structure. *Journal of the Acoustical Society of America*, 110:1074–1088.

[Breebaart et al., 2001b] Breebaart, J., van de Par, S., and Kohlrausch, A. (2001b). Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters. *Journal of the Acoustical Society of America*, 110:1089–1104.

[Dau et al., 1997] Dau, T., Kollmeier, B., and Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *Journal of the Acoustical Society of America*, 102:2892–2905.

[Dau et al., 1996] Dau, T., Pueschel, D., and Kohlrausch, A. (1996). A quantitative model of the effective signal processing in the auditory system. I. Model structure. *Journal of the Acoustical Society of America*, 99(6):3615–3622.

[Davidson et al., 2009] Davidson, S. A., Gilkey, R. H., Colburn, H. S., and Carney, L. H. (2009). An evaluation of models for diotic and dichotic detection in reproducible noises. *Journal of the Acoustical Society of America*, 126(4):1906–1925.

[Durlach, 1972] Durlach, N. I. (1972). Binaural signal detection-equalization and cancellation theory. *in Foundations of Modern Auditory Theory, edited by J. Tobias (Academic New York)*, II:369–462.

[Gelfand, 2009] Gelfand, S. A. (2009). *Hearing: An introduction to psychological and physiological acoustics*. Informa Healthcare, 5 edition.

[Glasberg and Moore, 1990] Glasberg, B. R. and Moore, B. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1-2):103–138.

[Laback, 2010] Laback, B. (Ausgabe 2010). Psychoakustik ii, schwerpunkt: Experimentelle audiologie, skriptum zur vorlesung.

[Larsen, 2010] Larsen, A. P. (2010). Model binaural signal processing in humans. Thesis, Technical University of Denmark, Department of Electrical Engineering.

[Levitt, 1971] Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49(2B):467–477.

[Patterson et al., 1987] Patterson, R., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (1987). An efficient auditory filterbank based on the gammatone function. *APU report*, 2341.

[Søndergaard and Majdak, 2013] Søndergaard, P. and Majdak, P. (2013). The auditory modeling toolbox. In Blauert, J., editor, *The Technology of Binaural Listening*, pages 33–56. Springer, Berlin, Heidelberg.

[Søndergaard et al., 2012] Søndergaard, P. L., Torrésani, B., and Balazs, P. (2012). The linear time frequency analysis toolbox. *International Journal of Wavelets, Multiresolution and Information Processing*, 10(04):1250032.

[Stangor, 2011] Stangor, C. (2011). Introduction to psychology; 4.3 hearing. `http://catalog.flatworldknowledge.com/bookhub/reader/127?e=stangor-ch04_s03`. Accessed: 2016-07-29.

[van de Par and Kohlrausch, 1999] van de Par, S. and Kohlrausch, A. (1999). Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters. *Journal of the Acoustical Society of America*, 106(4):1940–1947.