SIPI Loudness

Experimental Setup and Pilot Study for Loudness Perception in Cochlea-Implant Listeners with Short-Interpulse-Interval (SIPI) and Enhanced (ENH) Pulse Stimulation

Project Report

Supported by FEMTech

Lisa Frohmann

Reviewer: O.Univ.Prof. Mag.art. DI Dr.techn. Robert Höldrich
Supervisors: Doz. Dr. Bernhard Laback, DI Martin Lindenbeck

June, 2022

ÖAW

Acoustics Research Institute

kunst uni graz

Institute for Electronic Music and Acoustics

Abstract

Cochlear implants (CIs) provide a sense of hearing to deaf/hard-of-hearing people by electrically stimulating the auditory nerve via electrode arrays placed in the cochlea. Typical stimulation strategies yield amplitude-modulated pulstrains that are generated from extracting the envelope of the acoustic signal to be processed. Those strategies, while maintaining viable speech intelligibility, discard much needed acoustic information that is stored in the fine structure of the signal, such as cues about pitch and ITD (interaural time differences).

New stimulation strategies arise that counter those problems by modifying the pulse trains with additional SIPI (short-interpulse-intervals) pulses, which already yield promising results at the Acoustic Research Institute (ARI). Despite all improvements, additional pulses might also cause unwanted loudness differences that need to be taken into account regarding a possible real-life implementation of a new stimulation technique.

In this project (Toningenieursprojekt) called SIPILoudness, an experiment for the ARI Software ExpSuite is implemented to test two kinds of electrical stimuli for possible loudness differences: stimuli with SIPI pulses and so-called ENH (enhanced) stimuli with a pulse with an amplitude increase up to 3dB at each carrier pulse at modulation peak, compared to a reference without SIPI or ENH. A pilot study with one unilaterally implanted listener is conducted to proof software functionality. The results are processed and discussed.

Zusammenfassung

Die implantierte Hörprothese, das Cochlearimplantat (CI), ermöglicht tauben oder hochgradig hörgeschädigten Personen durch elektrische Stimulation des Hörnervs über ein Elektrodenarray in der Hörschnecke eine gewisse Form von Sprachverständlichkeit (wieder) zu erlangen. Signalprozessoren, welche die akustischen Signale aufbereiten und in elektrische Signale umwandeln, bedienen sich verschiedener Stimulationsarten, welche unentwegt weiterentwickelt werden, um das elektrische Hören für die Betroffenen immer weiter zu verbessern.

Das Prinzip der Stimulation beruht im Allgemeinen auf periodischen Pulsketten mit hoher Pulsrate, welche mit dem Verlauf der Einhüllenden des akustischen Signals amplitudenmoduliert werden. Darunter leidet allerdings die Übertragung von zeitlicher Information wie die ITD (interaural time difference) oder die Wahrnehmung von Tonhöhe, da die dafür verantwortliche Feinstruktur eines akustischen Signals zu Gunsten besserer Sprachverständlichkeit (bedeutet hohe Trägerpulsraten) vernachlässigt wird.
Ein Forschungsansatz am Institut für Schallforschung (ISF), der bei Erhaltung der Sprachverständlichkeit Vorteile in besagten Bereichen verspricht, ist der SIPI-Ansatz (engl.: short interpulse intervals) - das deterministische Einfügen von zusätzlichen Pulsen mit kurzen Interpuls-Intervallen.

Trotz der Vorteile könnten zusätzliche Pulse allerdings auch zu einem veränderten Lautheitseindruck führen und somit die Implementierbarkeit der Strategie in Frage stellen. Das vorliegende Toningenieurs-Projekt SIPILoudness zielt darauf ab, Unterschiede in der Lautheitswahrnehmung von elektrischen Stimuli zu untersuchen, welche mit SIPI-Pulsen sowie mit zum Trägerpuls unverschobenen und verstärkten Pulsen (ENH, enhanced) erzeugt werden. Referenzstimuli enhalten weder SIPI noch ENH Pulse. Über das Framework ExpSuite wird eine Applikation für ein Pilotexperiment implementiert und von einer einseitig CI-implantierten Testperson auf ihre Funktionalität getestet. Die Ergebnisse des Versuchs werden aufbereitet und diskutiert.

Contents

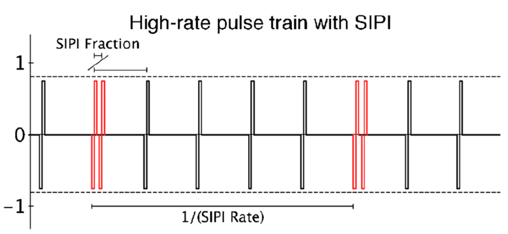
1 Introduction

The cochlear implant (CI) is a neuroprosthesis which can restore a sense of hearing in people with sensorineural hearing loss. Provided that the auditory nerve is still intact, an array of electrodes is inserted in the cochlea enabling its stimulation by a high rate carrier pulse train that samples the envelope of an acoustic sound signal. The latter is captured by a microphones and processed by a clinical signal processor that is usally worn behind the ear or directly on the spot where a magnetic coil transfers the signal through the skin to a receiving coil connected to the electrode array.

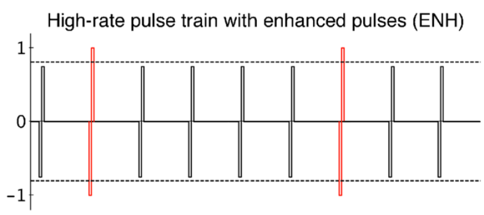
CIs provide so-called electric hearing - for most users that means a good understanding of speech in quiet situations with a small number of speakers with clear articulation. Stimulation strategies focus on the amplitude envelope of a sound that is important for speech signals but discard the temporal fine structure of the signal. However, this fine structure is exploited in hearing to gain crucial time information like interaural time differences (ITDs) or sensation of pitch that is caused by varying signal rates. Therefore, CI users have no access to these characteristics and lack sensitivity, which is causing impairment in everyday communication.
Research has its focus on new stimulation techniques that try to integrate cues for the auditory system to improve ITD and pitch sensitivity. One of them is the so-called SIPI approach. It is at the center of studies by the team of the Hearing Cluster of the Acoustic Research Institute of the Austrian Academy of Science, Vienna (ARI). As a starting point, the study of Laback and Majdak (2008) with jittered signals revealed that the occasional occourance of short interpulse intervals (SIPI) in a jittered pulse train enhances ITD sensitivity. Building on their findings, Srinivasan and Lindenbeck did particular studies on signals that contain deterministic insertions of SIPI pulses.

Srinivasan et al. (2018) did extensive experiments on ITD sensitivity enhancement with unmodulated SIPI stimuli (see Fig.1.1(a)), varying parameters like the rates at which extra pulses were presented and the exact locations within a stimulus. They reported that improvement of ITD sensitivity tended to increase with decreasing rate of SIPIs and with a decreasing SIPI fraction (fraction of interval of two adjacent pulses by which the location of the extra pulse is shifted), see Fig. 1.2.

They also presented the listeners with stimuli with enhanced (ENH) pulses (Fig. 1.1(b)) that represent an alternative to SIPIs, offering the short time energy at one pulse instead of two consecutive ones.



(a) Exemplary SIPI stimulus



(b) Exemplary ENH stimulus

Figure 1.1 – Stimuli by Srinivasan et al. (2018)

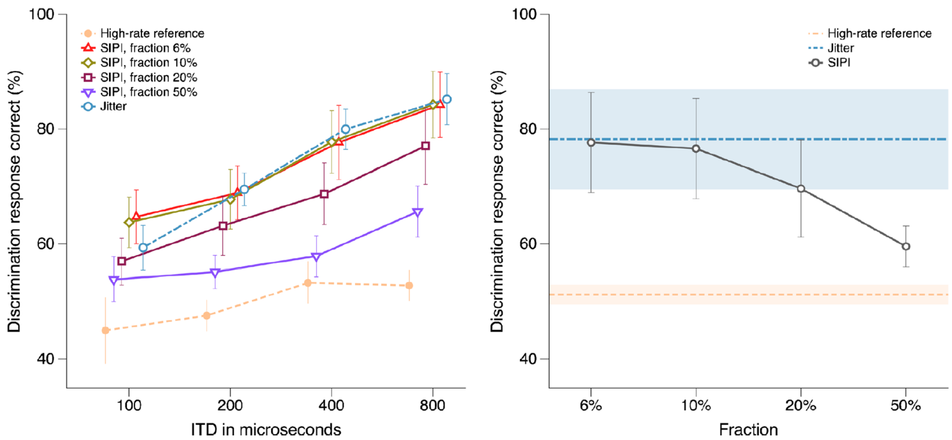


Fig. 3. (Color online): Results of experiment 1; left panel: left/right discrimination performance across different ITDs, for the four SIPI fractions (empty symbols connected with solid lines), the high-rate reference condition (filled circles), and the jitter condition (empty circles). Results are pooled over SIPI rates. Right panel: performance as a function of SIPI fractions, pooled over SIPI rates and ITDs (empty circles). The dash-dotted and dashed horizontal lines show performance in the jitter and reference conditions, respectively. Error bars around points or lines represent 95 % confidence intervals

Figure 1.2 – Results of experiments by Srinivasan et al. (2018), part 1

Results in Fig.1.3 yield a comparable ITD sensitivity improvement for the SIPI conditions and an enhancement of $+3dB$.

In the following study (Srinivasan et al. 2020), AM signals are introduced to mimic F0 modulation of voiced speech which are presented with additional SIPI pulses. In this study variations of the parameters modulation depth (MD), SIPI phase (phase of a SIPI pulse insertion) and AM rate were tested. They report a benefit for ITD sensitivity especially for SIPI insertion at the modulation peak of signals and throughout the range of tested MDs .

Lindenbeck et al. (2020) tested SIPI insertion into high rate amplitude modulated pulse trains. Insertion of the pulses at the same rate as the rate of the temporal envelope enhanced pitch sensitivity for the tested F0s and were even more prominent at lower MDs. They also listed a few advantages that the SIPI approach could have to existing strategies: Apart from one extra pulse, the envelope shape is not further modified and the temporal envelope of speech is preserved. For encoding more than one source concurrently, the timing of the SIPIs across channels is very flexible. Additionally, when a SIPI pulse is inserted at the envelope peak, the information that is provided by the SIPI pulse is explicitly conncted to the acoustic information of the sound.

However, besides all the promising positive effects of SIPI or ENH pulse insertion on ITD that were reported above, the influence of the modifications on the perceived loudness of stimuli has to be considered and evaluated. Srinivasan et al. (2018) already reported that while unmodulated SIPI stimuli with smaller SIPI fractions appear not to only a little louder than no-SIPI references, ENH signals of 3dB are perceived clearly louder than reference and SIPIs.

From a technical perspective, insertion of SIPIs would not alter the peak amplitudes of modulated

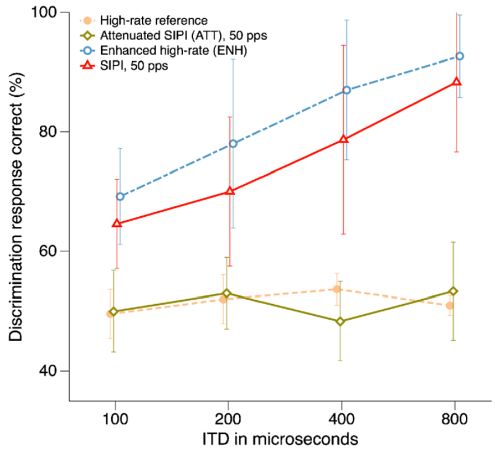


Fig. 6. (Color online): Results of experiment 2; left/right discrimination performance across different ITDs for the ENH (empty circles), the ATT (empty diamonds), the reference (filled circles), and the SIPI (empty triangles) conditions. Error bars around points or lines represent 95 % confidence intervals

Figure 1.3 – Results of experiments by Srinivasan et al. (2018), part 2

signals. However, further enhancement of the peak pulse can cause problems, because the pain threshold for stimulation limits the dynamic range and additional enhanced pulses might reach far too high beyond said threshold.

Independent of how much louder the newly introduced stimuli are perceived, the difference has to be considered and accounted for by the developers of clinical CI processors. Open questions about the feasibility of these approaches remain to be answered. At this point, more information about the loudness perception of SIPI vs. ENH signals has to be gathered concerning amplitude modulated signals to add to the bigger picture.

Report Structure. This report elaborates the design and realization of the experimental software **ExpSuite SIPILoudness** that allows the balancing of the loudness of stimuli to one another. Chapter 2 describes the generation of amplitude modulated stimuli containing SIPI pulses and enhanced pulses, provides information about the used task and procedure as well as their implementation in the application.
Chapter 3 covers the pilot study that is carried out to prove software functionality and displays the results of the loudness balancing task. (It is noted however, that an extensive perceptual evaluation is not part of this project.)
An evaluation of the data collected in the pilot study is done in Chapter 4 to discuss the efficiency of the procedure in terms of time costs and accuracy of the results.

2 SIPI Loudness Experiment: Design and Realization

This project aims at developing a test application, which consists of generation of test stimuli in Matlab, determination and modification of parameters for the testing GUI as well as visualization and interpretation of the acquired datasets. In the following sections the design of the used stimuli is presented, task and procedure characteristics are discussed, and the structure of the app and its framework are described.

2.1 Stimuli

The works of Srinivasan et al. (2018) and Lindenbeck (2017) form the basis for the generation of the modulated stimuli in **SIPILoudness**. Modifications are made so that both SIPI and ENH pulses can be inserted at the modulation peak or at the respective SIPI position which is the peak position plus the fraction of the next interval (SIPI fraction). Each stimulus consists of an onset ramp, a steady state and an offset ramp. The following sections introduce the main aspects of the stimulus composition. The displayed equations are taken from Chapter 2.1 of Lindenbeck (2017) and are slighty modified to fit the needs of this project.

2.1.1 Amplitude Modulation

In general, signals are produced that contain an absolute valued sinusoidal AM with a constant modulation frequency f_{AM}. Note that a full wave rectification doubles f_{AM}, the frequency of the sine is divided by 2 (π instead of 2π). The AM signal is written as

$$\mathrm{AM} = 1 + \mathrm{MD} \cdot (2 \cdot |\sin[\pi \cdot f_{\mathrm{AM}} \cdot t - \phi_0]| - 1) \tag{2.1}$$

where ϕ_0 is the starting phase that is needed in the process of building the whole stimulus and MD is the modulation depth with values between 0 and 1.

2.1.2 Composition of the Stimuli

Onset and Offset Ramps. Rapid on- or offsets could confound the data that is collected due to the auditory system reacting to sudden changes of level. Therefore linear ramps are introduced to each stimulus. They have a fixed length of 150 ms each and lead the signal amplitude from the measured threshold (THR) to the comfortable level (CL) of the participant. The CL should be well beyond the maximum comfortable level (MCL) so that the modulations of the signals do not scrape the uncomfortable upper limit of the dynamic range (DR). It is further noted that ramps never contain SIPI or ENH pulses.

Steady State. The actual steady state signal of a stimulus is obtained by multiplying the AM signal with the desired presentation amplitude

$$\mathrm{f(t)} = \mathrm{AMP} \cdot \mathrm{AM} = \mathrm{AMP} + (\mathrm{AMP} - \mathrm{THR}) \cdot \mathrm{MD} \cdot (2 \cdot |\sin[\pi \cdot f_{\mathrm{AM}} \cdot t - \phi_0]| - 1) \tag{2.2}$$

The length of the steady state is set to be $300ms$. Not for every value for f_{AM} it will be possible to fit an integer number of periods in that window - some will be longer, some shorter. To avoid that the AM specific length of a stimulus adds a length cue in the task, the maximum fitting number of periods in $300ms$ is accepted or one period is added randomly, yielding an average steady state duration of $300ms$.

The starting phase ϕ_0 should make sure the connection between the ramps and the steady state is as smooth as possible and big jumps in amplitude can be avoided. A minimum starting phase of 30° is introduced as a requirement. As the phases of the AM signal are quantized by the carrier pulse train, the first phase that is closest to the 30° is chosen as a best fit to connect to the onset ramp. This partial AM period is also mirrored in the last AM period to connect to the offset ramp in the same fashion.

The starting peak position of the SIPI/ENH insertion is fixed at 90° of the AM period and is called the SIPI phase. The SIPI fraction is the additional fraction of a carrier pulse interval in % that is added to the peak position and determines the final location of the SIPI pulse, which restricts the testable values for the AM periods to ones that are integer multiples of the carrier pulse period (CPP).

To ensure the insertion of ENH/SIPI pulses at the peak of the modulation, it must be checked if the peak can be sampled by the pulses of the carrier with carrier frequency f_c. That is only the case for a even number of pulses in one AM period. Thus, for odd numbers of pulses per AM peroid, the phases for all pulses are first shifted by half an interval between ajdacent pulses

$$\phi(k) = \begin{cases} k \cdot \frac{\pi}{i}, & i \text{ even} \\ \left(k + \frac{1}{2}\right) \cdot \frac{\pi}{i}, & i \text{ odd} \end{cases}, \quad i = \frac{f_c}{\mathrm{F0}} \in \mathbb{N}. \tag{2.3}$$

After that the starting phase ϕ_0 is determined like described in the paragraphs above.

$$\phi_0 = \underset{k}{\operatorname{argmin}}\ \phi(k) := \left\{k \in \mathbb{N} \middle| \phi(k) \geq 30°\right\} \tag{2.4}$$

The insertion function can now be written as

$$\mathrm{P}_{SIPI,ENH}(t,k) = \begin{cases} 0, & \text{no SIPI/ENH pulses} \\ \delta(t - [k/f_{AM} + \text{SIPI fraction}_{\%}/(100 \cdot f_c)]), & \text{SIPI pulse insertion} \\ \delta(t - [k/f_{AM}]) \cdot 10^{\frac{ENH_{dB}}{20}}, & \text{ENH pulse insertion} \end{cases} \tag{2.5}$$

with

$$\delta(t) = \begin{cases} 1, & t = 0 \\ 0, & \text{else} \end{cases} \tag{2.6}$$

In a final step, the carrier pulse train quantizises the AM steady state signal including modified or extra pulses to

$$f_Q(t) = f(t) \cdot \sum_{k=0}^{k_{max}} \underbrace{\delta(t - k/f_c)}_{\text{carrier pulses}} + \mathrm{P}_{SIPI,ENH}(t,k), \quad k_{max} = t_{max} \cdot f_c \tag{2.7}$$

$$\tag{2.8}$$

where t_{max} is the maximum length of stimulus (ramps and steady state with or without one additional AM period).

Overall Stimuli. Figures 2.1, 2.2 and 2.3 depict the reference, SIPI and ENH stimuli including ramps. In the ENH case, additional lines show an examplary THR (dashed line), the middle amplitude (dotted line) and the peak amplitude (solid line) of the modulated signals. Note that the real carrier pulse and AM periods are a lot smaller than could be depicted and the figures show a simplified version of the presented signals.

The ExpSuite framework (which will be introduced in detail in Sec.2.3) converts the generated signals into a biphasic pulse train with a phase duration of 26.7μs and forwards it to the stimulation interface that connects the participants' implants.

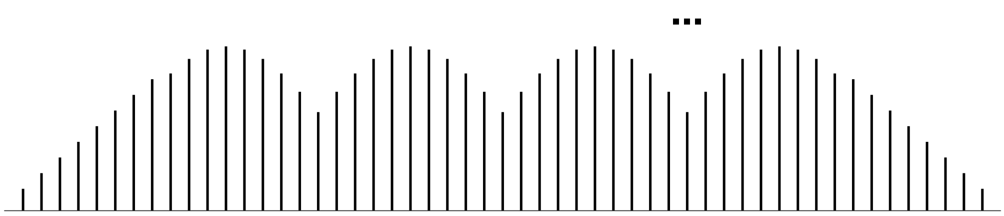


Figure 2.1 – Reference Stimulus

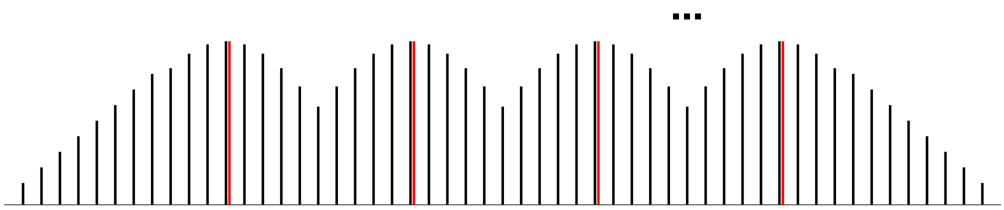


Figure 2.2 – SIPI Stimulus

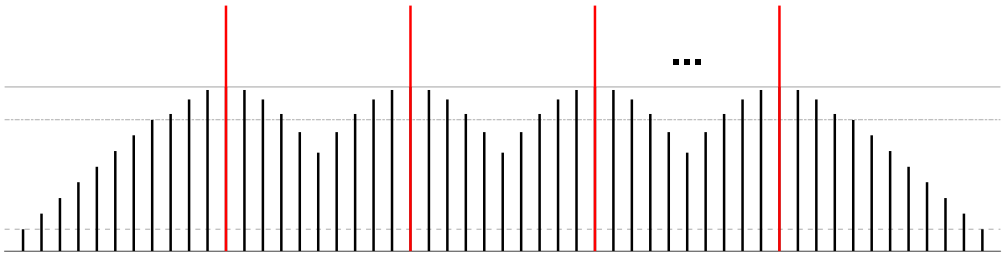


Figure 2.3 – ENH Stimulus

2.2 Task and Procedures

2I-2AFC task.

The loudness balancing used in **SIPILoudness** is a 2I-2AFC task (McNicol 1972, 40-45). That means that there are two consecutive intervals and in each one stimulus is presented. In Signal Detection Theory it is defined that one interval contains the noise, while in the other signal and noise are presented. The signal stands for stimulus characteristics that the test subject is meant to detect in the task. 2AFC (two alternatives forced choice) states that only two answers to the question of the trial are possible and that the listener has to choose one of them.

In the case of the loudness balancing, the signal of interest is the loudness difference between reference and target. The question displayed on the experiment screen says *Welcher Ton war lauter? (Which tone was louder?)* and the possible answers were *Ton 1 war lauter (Tone 1 was louder)* or *Ton 2 war lauter (Tone 2 was louder)*, corresponding to the left and right buttons on the backside of the Logitech WingMan Gamepad. Handling that kind of controller is very intuitive and ergonomic over a longer period of time and keeps additional cognitive load small. Fig.2.4 shows the instruction sheet that was given to the participant before the task. It also indicates the yellow button for starting the experiment and starting the presentation of a new trial. Pushing the button for each trial can also be omitted and the new trial starts automatically. The latter option was chosen in the pilot study.

(a) Experiment screen.

(b) Controller instructions.

Figure 2.4 – Experiment screens and controller instruction for the adaptive loudness balancing task.

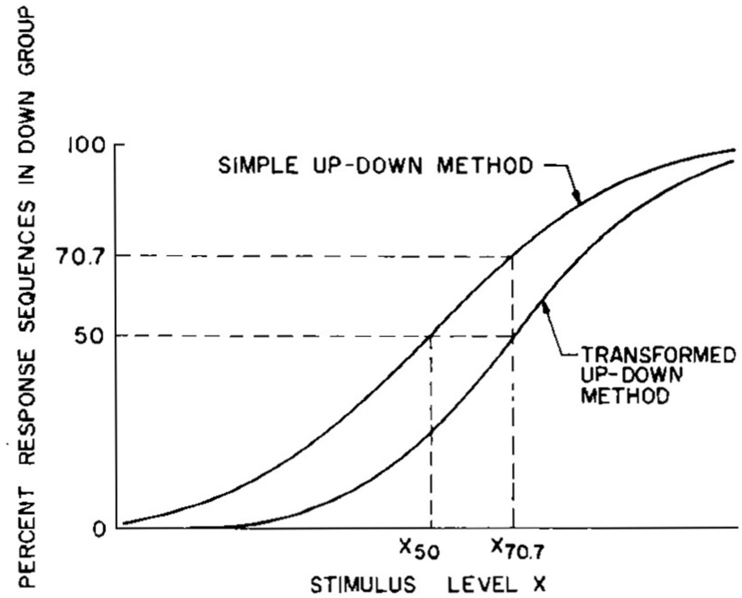
The fields 1 and 2 on the screen are highlighted slightly before and during the playback of each interval. On the bottom of the screen the listener can see their progress in the task in %.

2.2.1 Adaptive Loudness Balancing

The procedure used for the loudness balancing is called a transformed up-down method (Levitt 1971). The original up-down staircase method works in a way that the stimulus level is increased or decreased in discrete steps depending on the listener having detected the signal in the previous trial or not. In general, this method adaptively measures the stimulus level x that is needed to reach a certain performance level along the psychometric funtion (see fig. 2.5(a))

Usually, trials start at a level where there is a high chance of detection and are lowered for a correct answer. When giving an incorrect answer, the turning point, a so-called turnaround (TA) is reached and levels are rising again. These steps are repeated until a demanded number of TAs is reached to end the run. For the simple 1up-1down method, the procedure converges at the 50% point x_{50}. This is the point where the listener is guessing the answer.

As stated in the preceeding section 2.2 we are using a 2AFC task, where chance level is at 50%. It is necessary that the convergence point of the procedure is greater than the chance level of a AFC task to assure convergence. For **SIPILoudness** a 3down-1up method is selected where stimulus levels are only lowered for 3 consecutive correct (*target louder*) answers and increased for a single incorrect (*reference louder*) answer. This transformes the rules of the simple up-down method such that the performance yields the stimulus level where 79% of the given answers are correct. Vice versa, the 3up-1down staircase yields the 21% point. For each target condition, both procedures are executed and the mean of their obtained thresholds show the 50% point, where we find equal loudness of reference and target stimulus and the listener is not guessing their answer (see Jesteadt (1980) for details).
Fig.2.5(a) from Levitt (1971) illustrates the psychometric funtion for a simple up-down method and the transformed response curves for an exemplary 2down-1up method. In this case the transformed strategy converges at the 50% point, which corresponds to the 70.7% point on the listeners's psycho-



(a) Response curves for a simple up-down and a transformed 2down-1up method

TRANSFORMED UP-DOWN METHODS

TABLE I. Response groupings for transformed up–down strategies. Several simple response groupings are shown. Entry 1 corresponds to the simple up–down procedure. Entry 2 corresponds to the method used by Zwislocki *et al.* (1968) and Heinemann (1961). Entries 2 and 3, and 5 and 6, with random interleaving, were used by Levitt (1964). Entry 7 is typical of the BUDTIF procedure proposed by Campbell (1963). Entry 8 was used by Levitt and Rabiner (1967).

	Response sequences		Response groupings	
Entry	UP group increase level after:	DOWN group decrease level after:	Probability of a sequence from DOWN group $= P[\text{DOWN}]$	Probability of positive response at convergence
1	−	+	$P(X)$	$P(X)=0.5$
2	+ − or −	+ +	$[P(X)]^2$	$P(X)=0.707$
3	− −	− + or +	$[1-P(X)]P(X)+P(X)$	$P(X)=0.293$
4	+ + − or + − or −	+ + +	$[P(X)]^3$	$P(X)=0.794$
5	+ + + − or + + − or + − or −	+ + + +	$[P(X)]^4$	$P(X)=0.841$
6	− − − −	− − − + or − − + or − + or +	$1-[1-P(X)]^4$	$P(X)=0.159$
7	Any group of 4 responses with 1 or more negative responses	+ + + +	$[P(X)]^4$	$P(X)=0.841$
8	− − − + − + − −	+ + + − + − + +	$[P(X)]^2[3-2P(X)]$	$P(X)=0.5$

(b) Table for transformed up-down methods

Figure 2.5 – Details of trasformed up-down methods, taken from Levitt (1971)

metric funtion. The table in Fig.2.5(b) shows response sequences and probabilities at convergence for various transformations.

Implementation in SIPILoudness. The participant has to balance the target stimulus to be perceived as loud as the reference, the latter being stimulated at their CLs. The starting amplitude for the target is calculated with $p = 20\%$ to

$$\mathrm{AMP_{Start}} = \left\{ \begin{array}{ll} \mathrm{CL} + (\mathrm{DR}) \cdot \frac{p}{100} & \text{for } \beta = 0 \\ \mathrm{CL} - (\mathrm{DR}) \cdot \frac{p}{100} & \text{for } \beta = 1 \end{array} \right\}, \quad \beta \in \{0; 1\} \tag{2.9}$$

where β is called decision rule. $\beta = 0$ states a down-staircase starting from the higher amplitude, while $\beta = 1$ describes an up-staircase starting with the lower amplitude. Figures 2.6(a) and 2.6(b) show examplary plots of a down and an upward staircase from the pilot data (see Chapter 3).

The implementation in the **SIPILoudness** app offers a few parameters that enable a faster convergence of the procedure. The step size μ by which the target amplitudes are altered is modified after each TA by the factor to decrease step size d. The following equations show the relation between step size μ, the value of the actual step in current units μ_{cu} and how the decision rule β is incorporated to determine the next amplitude value (adapted from Lindenbeck (2017)).
In **SIPILoudness** the initial step size is 10% of the DR and the minimum step size is 2% of the DR. If the target amplitude reaches a value higher than MCL twice in a row, the run is stopped and labeled as not finished. Furthermore, amplitudes can never go below THR.

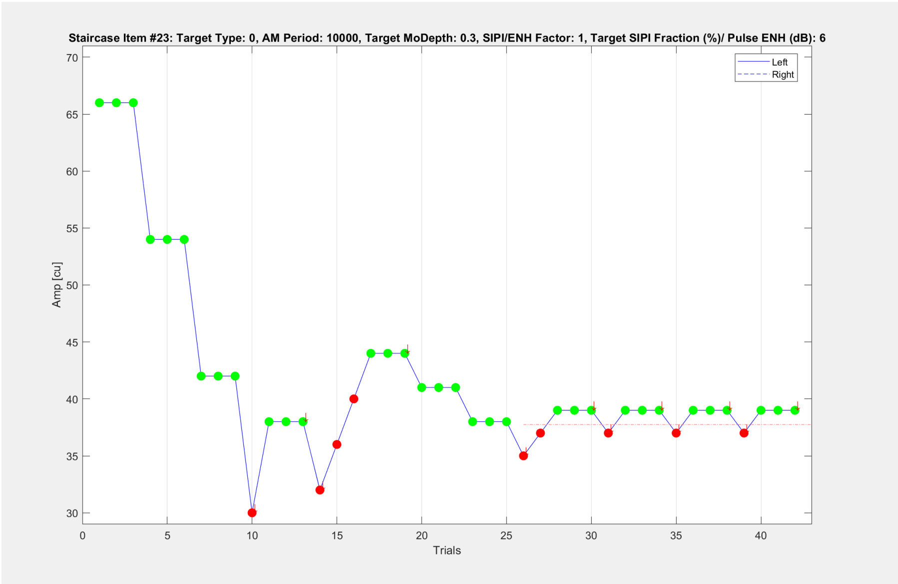
$$\mu = \max\{\mu_0 \cdot d^{N_T}; \mu_{min}\} \quad \textit{in } \% \text{ of DR} \tag{2.10}$$

$$\mu_{cu} = \left\lfloor (\mathrm{MCL} - \mathrm{THR}) \cdot \frac{\mu}{100} \right\rceil \quad \textit{in } \mathrm{cu} \tag{2.11}$$

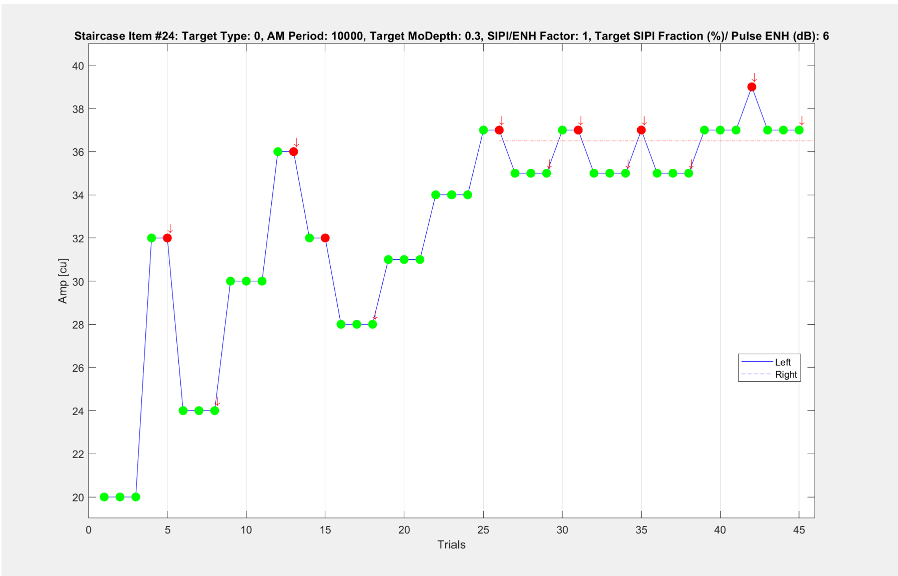
$$\mathrm{AMP_{k+1}} = \left\{ \begin{array}{ll} \mathrm{AMP_k} - (2 \cdot \beta - 1) \cdot \mu_{cu} & [\beta = 0 \text{ and ref louder}] \textit{ or } [\beta = 1 \text{ and tar louder}] \\ \mathrm{AMP_k} + \left\{ \begin{array}{ll} 0, & l < n \\ (2 \cdot \beta - 1) \cdot \mu_{cu}, & l = n \end{array} \right\} & [\beta = 0 \text{ and tar louder}] \textit{ or } [\beta = 1 \text{ and ref louder}] \end{array} \right. \tag{2.12}$$

$$k \in \{1, ..., K\} \qquad \mathrm{AMP_{k+1}} = \max\{\mathrm{AMP_{k+1}}; \mathrm{THR}\}$$

μ	...	Step size	μ_{cu}	...	Step in current units
μ_0	...	Initial step size	μ_{min}	...	Minimum step size
d	...	Factor to decrease step size	N_T	...	# of current turnaround
β	...	Decision rule	l	...	# of correct responses in a row
n	...	n-down 1-up method	K	...	total # of responses in a run



(a) 3down-1up staircase.



(b) 3up-1down staircase.

Figure 2.6 – Staircase example from pilot: SIPI fraction 6%, 100Hz AM rate and 0.3 MD (plottet from ExpSuite, right ear signal only).

2.3 ExpSuite

Over the last decade, the open source software *ExpSuite* for psychoacoustical experiments (Mihocic 2014) was (and still is) developed at the ARI. It consists of a framework containing functions and various experimental designs and over 40 applications for different studies that were and are performed at the institute. While the framework is taken care of by the lab technician, the experimentator is welcome to develop their own application fitting the needs of their planned experiments. For this project, the application **SIPILoudness** was realized. The application **ALBforPitchSIPI** functioned as a basis and was modified over a course of six weeks.

Fig.2.7 shows the overall structure of the software. The framework and applications are programmed in VisualBasic.NET. Stimulus generation in generally executed in Matlab. Access to computational software is guaranteed via the framework. Pure Data (Pd) can be used for acoustic stimulation for normal hearing listeners. For the CI experiments, a connection exists to the Research Inferface Box (RIB). RIB2 (developed by the Department of Ion Physics and Applied Physics at the University of Innsbruck, Innsbruck, Austria) faciliates bypassing the listeners' own CI processors and therefore directly stimulate the user at selected electrodes of the implant. The interface offers a magnetic coil for each ear.

The framework takes care of the appearance of the GUI (including all visual presentation and feedback), structuring events when pushing buttons, administer global configurations, ensuring connection to interfaces and hardware and offering predefined procedures (e.g. AFC). Via the application, stim-

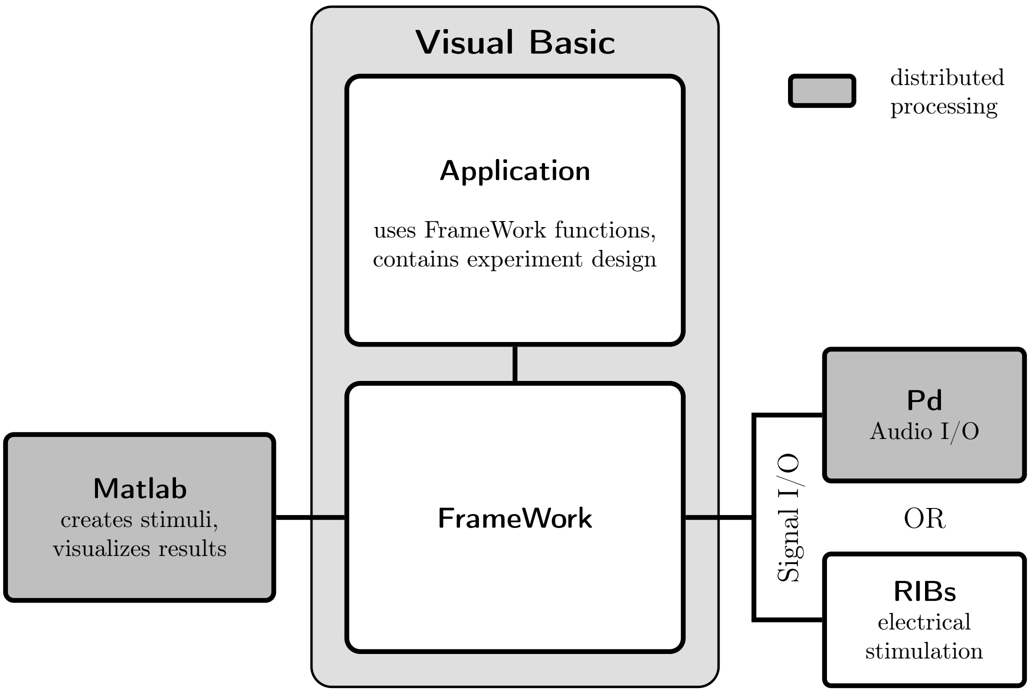


Figure 2.7 – ExpSuite software strucure, modified from Mihocic (2014), taken from **?**.

uli are created and played back, variables and constants of the experimental setup are defined and so-called item lists are created, the latter listing all possible combinations of stimulus conditions that are meant for presentation in the task.

Framework functions for CI experiments. When electrical stimulation via RIB or RIB2 is chosen, the software must be fed with listener-specific fitting files, that are generated in the so-called Fitt4fun section. One file contains information about the type of implant and the identified THR, MCL and CL for one ear at each electrode. Only electrodes selected in the file can be stimulated in the tasks afterwards.

2.3.1 SIPILoudness Application

As already stated in previous sections, the SIPILoudness application offers the loudness balancing where levels of various target stimuli are meant to be adjusted to sound as loud as a reference signal. Modifications were necessary in the means of creating an item list that contains all of the needed values. Fig.2.8 shows the main window of the app with a dummy item list loaded. It does contain the variables chosen for the task, but not yet the answers and their amplitudes, nor calculated THRs or STDs. Such lists are automatically created by the `Create list` button, provided that values and ranges of the variables as well as necessary constants are entered in the respective tabs of the `Settings` window (see Fig.2.9(a) and 2.9(b)). All variables are required to have a certain format, be linked to other variables if needed and not exeed a certain range.

In the `Constants` tab, the overall length of stimuli and length of ramps are determined. The required number of TAs to complete a run as well as TAs for calculation are stated, as well as stepsize parameters and the type of transformed method (x-down 1up). All information about the reference (MD, SIPI fraction, AM Period) can be found here, also the Carrier Pulse Period and SIPI Phase (phase at which a SIPI pulse is inserted) for all stimuli.

In the same way as already used in previous ARI studies, every condition is asked twice, an up-staircase and a down-staircase for each condition which only differ in the value for the decision rule β. In case the participant comprehends the nature of the adaptive task, this might cause loss or jumps of focus. To minimize this bias effect, the option of interleaved item representation is available for the adaptive procedure as proposed by Jesteadt (1980). The Number of Interleaved Items is entered in the `Constants` tab. For a number of two, the runs of the two related conditions with $\beta = 0$ and $\beta = 1$ are interleaved randomly.

In the lower right corner of the main window, one can choose to execute fast calculations for average THRs for each condition and STDs, show a plot of the actual staircase of a selected item and initiate a re-calculation of the THR using more or less TAs for calculation than stated in the `Constants` settings after the data collection is finished.

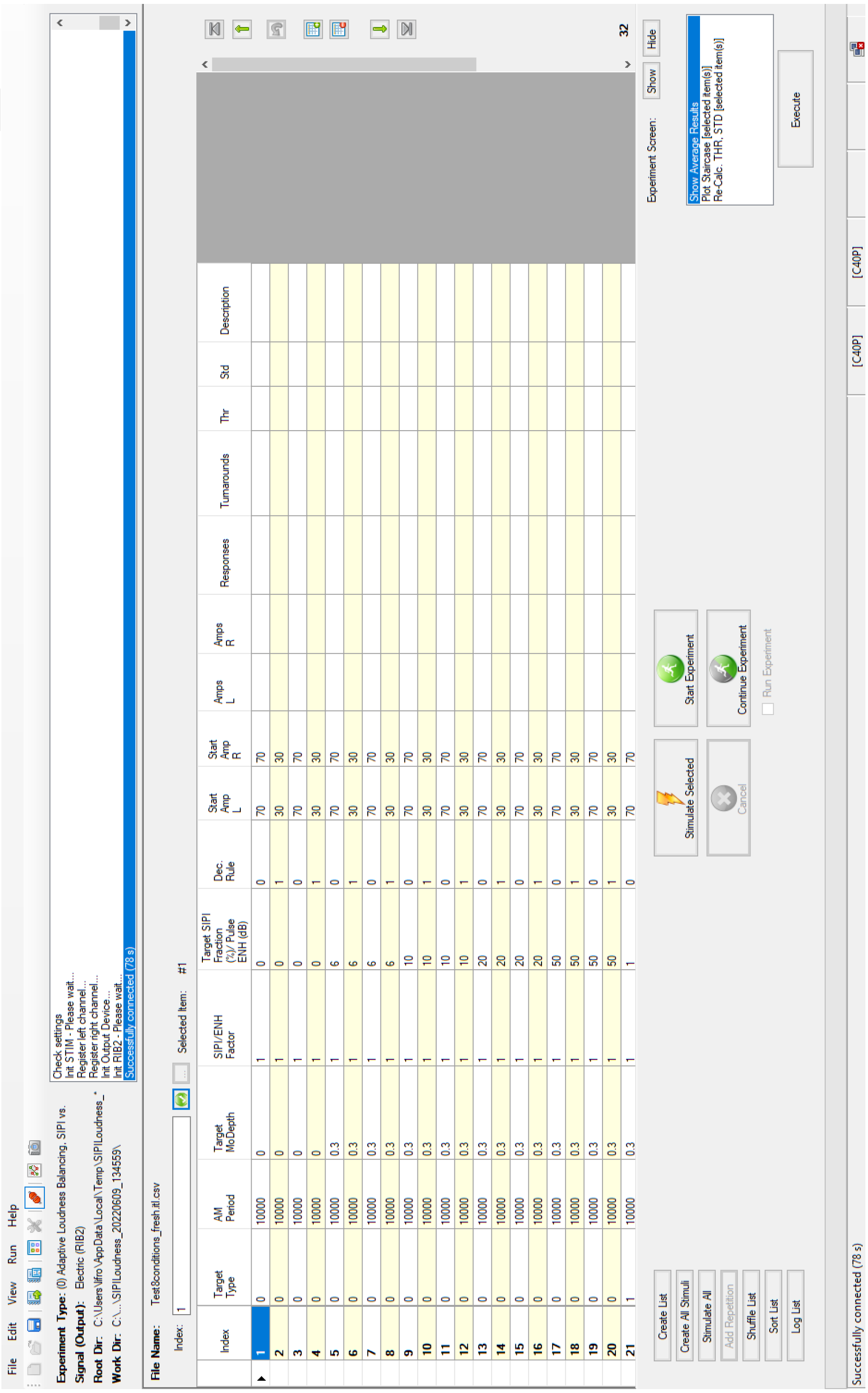
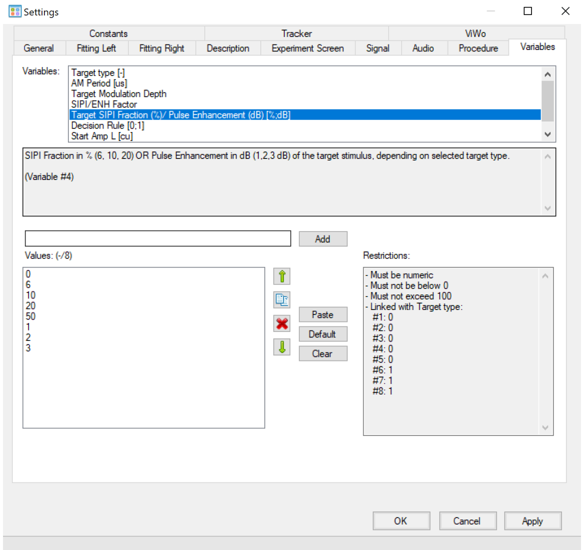
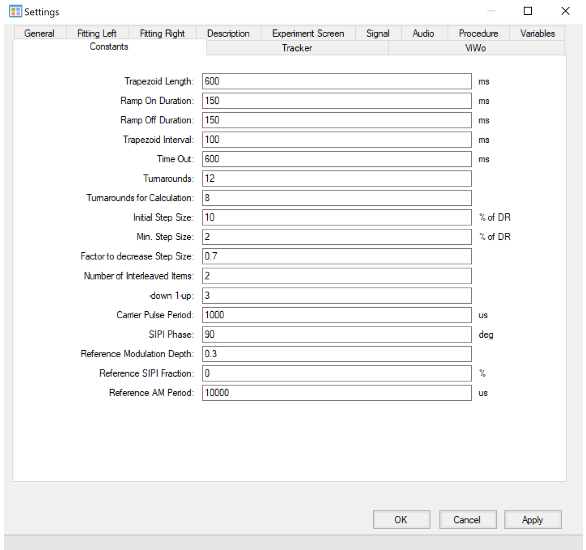


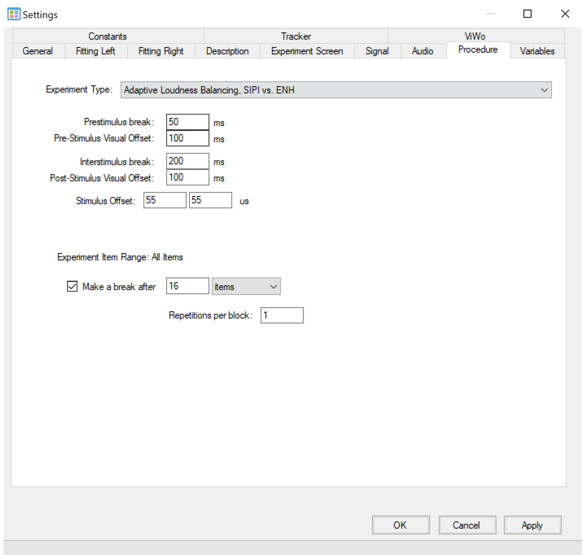
Figure 2.8 – SIPILoudness-ExpSuite main window.



(a) Variables.



(b) Constants.



(c) Procedure.

Figure 2.9 – SIPILoudness-ExpSuite: Pilot settings.

In the `procedure` tab 2.9(c) parameters concerning visual representation in the task can be stated. Repetition per block determines to the number of runs of an adaptive procedure, breaks for the listener can be scheduled after a certain amount of items or percent of all items of the current list.

3 Pilot Study

The pilot study checks for the functionality of the implemented experiment application. It is conducted unilaterally. The adaptive loudness balancing task should reveal the perceived loudness relations between the SIPI and ENH test stimuli compared to the reference stimulus.

3.1 Setup

3.1.1 Participant

Due to time restictions in the project schedule, only one subject participated in the pilot study. We want to gratefully acknowledge that all test sets were patiently completed by the participant within the duration of one session. Table 3.1 shows detailed subject information. As the subject had already participated in the related SIPI study (Srinivasan et al. 2018), the electrode used for testing was chosen to be identical to the one used that last time.

ID	Gender	Age at testing	Implanted since (right ear)	Implant type	Electrode used	Deafness onset
CI116	male	56 years	2018	Synchrony [1]	8	2017 [2]

Table 3.1 – Details about the participant of the pilot study.

Before testing, the fitting procedure was performed in two steps. First the ExpSuite **Fitt4fun** fitting software identified the listeners treshold THR_1 and maximum comfort level MCL_1 for an unmodulated test signal in the subject specific range for clinical units (in this case range 3 is selected [3]). The fitting was conducted by the experimenter's stepwise manual stimulation and the subsequent subject feedback by indication of the perceived comfort level on a printed scale.
The second fitting with the ExpSuite **LevelDancer** application checked for THR_2 and MCL_2 when stimulating with a modulated signal (MD = 0.3). THR_1 and THR_2 should be similar to each other. As the measured MCL_2 will be lower than the actual MCL that is the peak of the modulation, the latter has to be calculated as follows:

$$\mathrm{MCL_{calc}} = (\mathrm{MCL}_2 - \mathrm{THR}_2) \cdot 0.3 + \mathrm{MCL}_2 \tag{3.1}$$

$$= (97 - 7) \cdot 0.3 + 97 = 124\ cu \tag{3.2}$$

In the final step of the fitting, the comfort level (CL) was determined which represents the amplitude for the reference signal and the starting value for the calculation of the starting amplitudes of the

1. Manufactured by Med-EL GmbH in Innsbruck, Austria
2. Sudden hearing loss due to noise exposure in the working place (metalworker), hearing aid on left ear, little to none remaining hearing without aids
3. Ranges 1, 2 and 3 are available and span each from 0 to 127 current units. Multiplication with implant and range specific constants yield the stimulation amplitude in μs. In this case the constant factor is 9.45.

target stimuli of the task.

THR_1	MCL_1	THR_2 LevelDancer	MCL_2 LevelDancer	MCL calculated	CL	Dynamic range
cu						
18	86	7	97	124	47	117

Table 3.2 – Thresholds and MCLs from the fitting procedure.

3.1.2 Conditions

The participant was presented with three test sets. The carrier pulse period was set to $1000\mu s$ (i.e. a rate of 1000 pulses per second). All variations of AM period, modulation depth (MD) and SIPI fraction or enhancement of the peak pulse are documented in Tab. 3.3. In Set 1 SIPI fractions and ENH were varied, in set 2 the AM period was changed for $3dB$ and 6% conditions. Set 3 offers three

	Target type SIPI=0 ENH=1	AM period (AM rate) μs (Hz)	Modulation depth	SIPI fraction or ENH % / dB	THR cu	STD cu
ref	0	10000 (100)	0.3	0	43	
			Set 1			
1	0	10000 (100)	0	0	50.25	0.74
2	0	10000 (100)	0.3	6	36.88	1.83
3	0	10000 (100)	0.3	10	37.94	0.94
4	0	10000 (100)	0.3	20	39.12	2.32
5	0	10000 (100)	0.3	50	43.31	0.85
6	1	10000 (100)	0.3	1	39.88	0.92
7	1	10000 (100)	0.3	2	34.94	1.28
8	1	10000 (100)	0.3	3	30.94	0.31
			Set 2			
9	0	2000 (500)	0.3	6	36.81	1.39
10	0	4000 (250)	0.3	6	37.63	1.45
11	0	20000 (50)	0.3	6	37.38	3.43
12	1	2000 (500)	0.3	3	30.13	0.97
13	1	4000 (250)	0.3	3	31.81	0.55
14	1	20000 (50)	0.3	3	28.88	1.33
			Set 3			
15	0	4000 (250)	0	20	51.5	0.46
16	0	4000 (250)	0.3	20	39.94	0.80
17	0	4000 (250)	0.7	20	32.44	1.26

Table 3.3 – Conditions for the adaptive loudness balancing task including result THRs (the balanced amplitudes of the target) and STD: reference and targets.

modulation depths for a 20% SIPI item. Each condition item was presented four times, two upward (decision rule = 0) and two downward staircase runs (decision rule = 1). Mean THR (i.e. the balanced amplitude of the target) and STD are calculated over the results for mentioned four items. The target type is coded in 0 for SIPI and 1 for ENH signals.

3.2 Results

In this section the results of the adaptive loudness balancing tasks are displayed which were stored by the application as result.csv files. All figures show the amplitude difference of the target stimulus from the reference in cu. The peak amplitudes of the modulated signals are calculated in the same fashion as in (3.2). For the participant's THR and CL from Tab. 3.2 we obtain the peak amplitude of the reference

$$\begin{aligned} \mathrm{Peak}_{\mathrm{ref}} &= (\mathrm{CL} - \mathrm{THR}_2) \cdot 0.3 + \mathrm{CL} \\ &= \mathrm{round}((43 - 7) \cdot 0.3 + 43) = 54cu \end{aligned} \tag{3.3}$$

and the peak amplitudes for all target conditions with THR_{SCM} being the staircase condition mean over all repetitions per condition

$$\mathrm{Peak}_{\mathrm{target}} = (\mathrm{THR}_{\mathrm{SCM}} - \mathrm{THR}_2) \cdot 0.3 + \mathrm{THR}_{\mathrm{SCM}}. \tag{3.4}$$

For the original middle amplitude THRs and STDs see the last columns in 3.3. The 95% confidence interval is plottet additionally to the mean. It is computed in Matlab using the Student's t inverse cumulative distribution function with the mean $\bar{x}$, the STD σ, the significance level $\alpha = 0.05$ and $N = 4$ degrees of freedom.

$$C_{95\%} = \left[\bar{x} - t_{(1-\frac{\alpha}{2},N-1)} \cdot \frac{\sigma}{\sqrt{N}}; \bar{x} + t_{(1-\frac{\alpha}{2},N-1)} \cdot \frac{\sigma}{\sqrt{N}}\right] \tag{3.5}$$

Fig. 3.1 displays the results for test set 1. For the SIPI conditions we see that the smaller SIPI fractions need a lesser amplitude (starting at -8 cu) than the larger fractions to be perceived as loud as the no-SIPI reference. 50% fraction even lands on the same amplitude as the reference, so the extra pulse in the middle of the interval doesn't seem to elicit a notable difference in loudness. Confidence intervals spread wider than +/-2cu on average and overlap for all conditions except the 50%.

Srinivasan et al. (2018) tested $+3dB$ enhancements to improve ITD sensitivity with all their subjects. One participant was available for additionally testing $1dB$ and $2dB$, which also lead to improvement, increasing with dB value. Thus, it is of interest to test stimuli with all three options.
For a larger ENH of 3dB, stimuli have to have an amplitude of even 16cu lower than the reference to be perceived as equally loud. This corresponds to decreasing the amplitude of the target by about

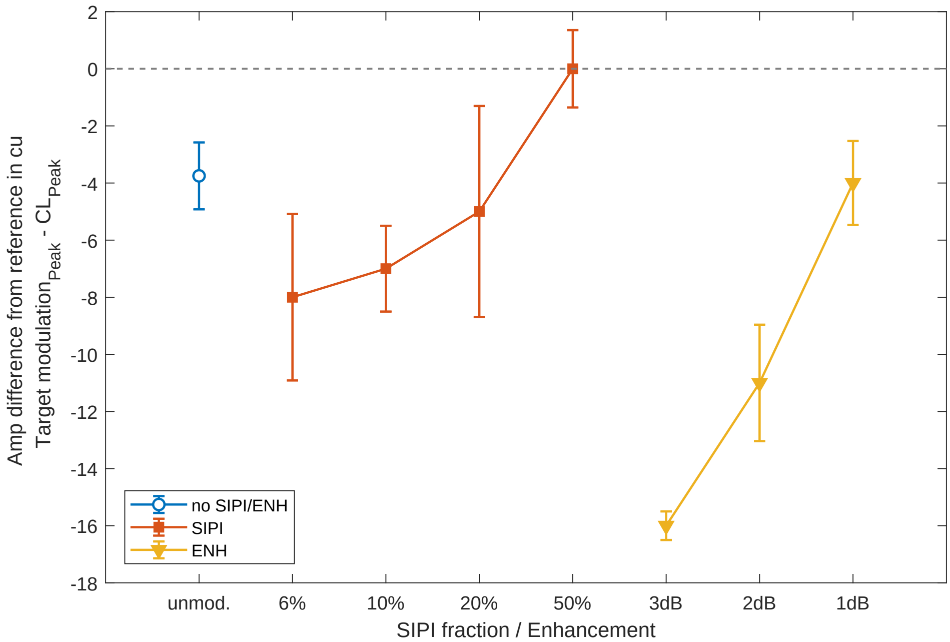


Figure 3.1 – Set 1: Difference from reference for SIPI fractions 6, 10, 20 , 50 % and 3, 2, 1dB ENH @ 100Hz AM rate and MD = 0.3

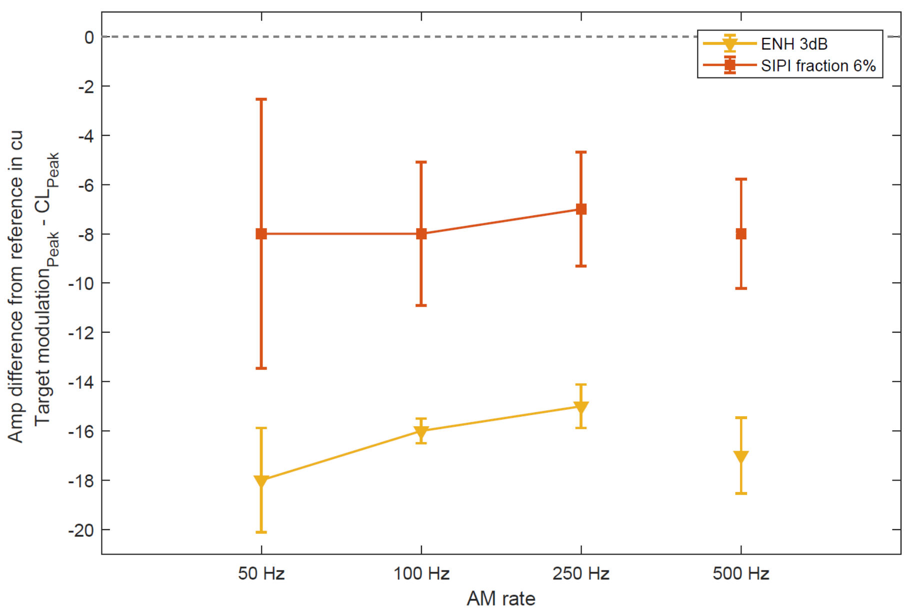


Figure 3.2 – Set 2: Difference from reference for 50, 100, 250, 500Hz AM rate @ SIPI fraction 6%, 3dB ENH and MD = 0.3

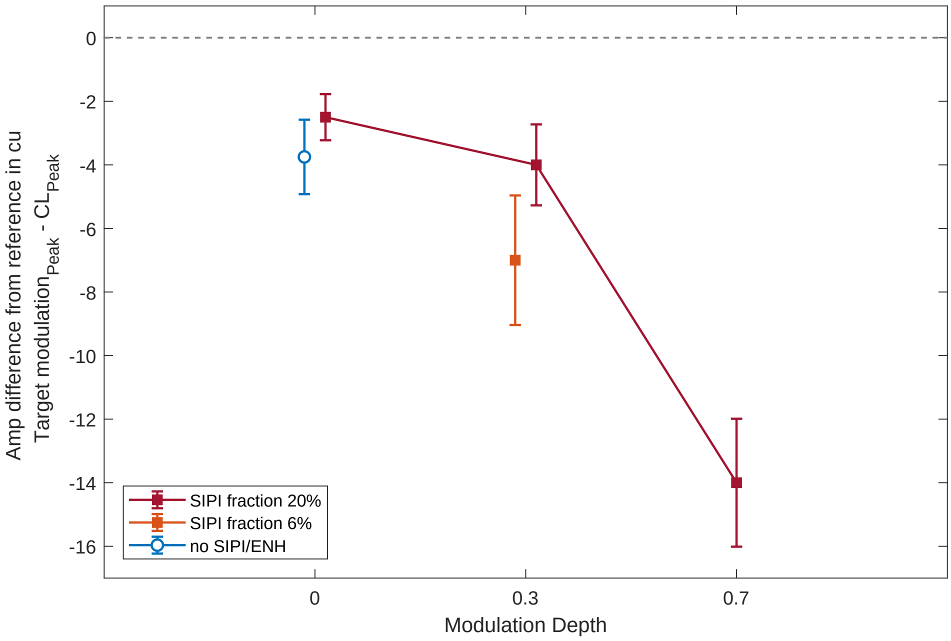


Figure 3.3 – Set 3: Difference from reference for 0, 0.3, 0.7 MD @ SIPI fraction 20% and 250Hz AM rate

3dB in relation to the reference. For the 2dB ENH it is about 1.9dB and for 1dB ENH one would need a 0.6dB decrease. The 1dB ENH is in the same range as the unmodulated stimulus without modified peak pulses. The loudness perception for 3dB ENH differs substantially from all other conditions.

The results of the second test set is shown in Fig.3.2. Because Srinivasan et al. (2018) have shown that a SIPI fraction of 6% and 3dB ENH yield comparable benefits for ITD perception tasks, the AM rate is varied in this part of the study. The SIPILoudness app takes AM period values in μs, but is displayed here in AM rate in Hz for better readability.
Although the amplitude differences of the two target types differ significantly throughout all rates, there is no effect of the AM rate spotted within these groups. It is noted, that the confidence intervals for ENH stimuli are always smaller than for SIPI stimuli, especially for 100 and 250Hz. There seems to be a high reproducability of very similar THRs for ENH pulse staircase runs. Additionally, in electric hearing the carrier pulse rate should be about four times higher than the highest AM rate to make sure the response of the neuron to the signal is not distorted (Wilson et al. 1997). That already might be the case for the 500Hz conditions.

The last test set aims at the effect of MD. The 20% SIPI fraction stimulus is presented with MDs of 0, 0.3 and 0.7. The AM rate is set to be 250Hz for this task, because it still fullfills the requirement by Wilson et al. (1997) stated before. The 20% fraction was selected, because it seemed to be the upper limit where the ITD discrimination performance (Srinivasan et al. 2018) was still comparable to the performance with jittered signals (see 1.2).

The results are plottet in Fig.3.3 with the unmodulated, unmodified condition and the 6% condition for additional comparison. Stimuli with the largest MD of 0.7 seem to elicit the same loudness as the reference only when target amplitude is set about 14cu lower than reference, which lies in the range of the 3 and 2dB stimuli from test set 1 (see Fig.3.1).

4 Evaluation of Staircase Data

Adaptive staircase procedures as the loudness balancing task are part of most experimental designs in the CI field and are used in the pretest-phase of many of the studies at the Acoustic Research Institute. In most cases, participants take part in experiments for only a short period of time, usually a few days where they get compensated for their time and energy spent. It is therefore desireable to keep pretests short while preserving the accuracy of the obtained data as a basis for the following experiments.

One way to reduce the time spent for the loudness balancing procedure could be to supply the subject with stimulus types that might enable a more rapid decision-making about the loudness relations - if there is a choice of type and an interpretable difference in performance. Another possibility could be the reduction of turnarounds that are required to complete a staircase run successfully, which is still in the power of the instructor. The following sections will focus on the evaluation of the data obtained in the pilot test to shed an informative light on potential improvements. It should be noted that all conclusions are only drawn from one participant's dataset and and can therefore not be stated as true for every CI user.

4.1 Average Trial Count for Staircase Run

A first and intuitive step is to compare the average length of staircase runs of the different stimulus types. Each answer given by the participant, each trial, is counted. The sum of counts is divided by the number of items of the respective stimulus type, as depicted in 4.1 for the main categories. Results show slightly below 45 trials needed to close one run on average, indepentend of signal type. With two stimuli of 600ms each and an interstimulus break of 200ms and time for offsets and answering, one could estimate up to approx. 3 sec per trial, leading to 2min 15sec per run and 2h 33min for the

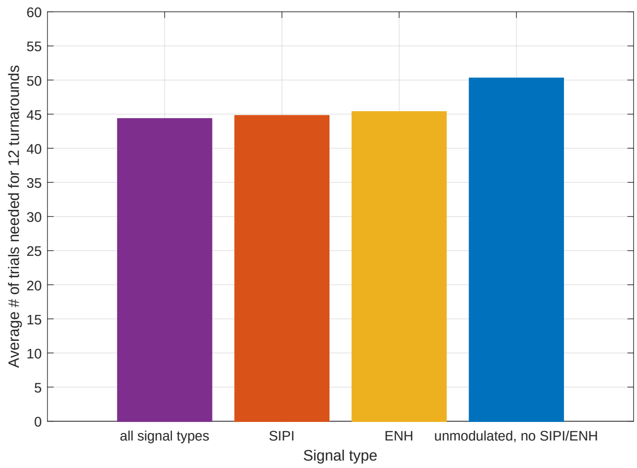


Figure 4.1 – Average length of staircase run per signal type

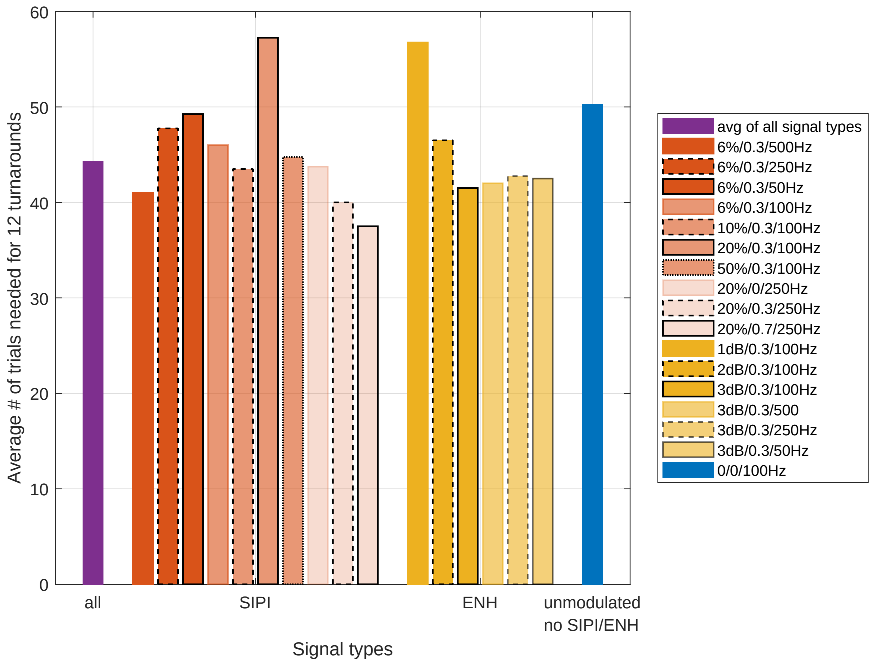


Figure 4.2 – Average length of staircase run per condition

whole pilot study.
Fig.4.2 goes into detail about the different conditions in the signal types. We see variation within the SIPI group, with the shortest run for the biggest modulation depth at 20 % SIPI fraction and longest run for a 20 % SIPI fraction at 100Hz AM and 0.3 MD. ENH runs are of same length for 3dB and are longer for smaller enhancements.

4.2 Occurrence of Turnarounds relative to Length of Staircase Run

In a next step we look into the progress behaviour in the staircase run. It might be interesting to know when the 12 demanded turnarounds (TA) are reached. This progress is calculated by first retrieving the indices of TAs for each run and seeking the mean of the four items that belong together. After that, the mean is divided by the average length of the respective condition yielding the progress along a run in %. Figures 4.4 and 4.5 are plots of progress bars for all the SIPI and ENH conditions, whereas 4.3 displays an overview for the pooled signal types. The overall development appears to be rather linear, for neither of the conditions one can see a exponential curvature. The latter could be assumed if the test subject was very sensitive and determined in their decision with small oscillations of the staircase in the run.
Focusing on the occurence of the first turnaround, we can observe that there is little variation within the ENH (around 15%) but more within the SIPI conditions. Intuitively, reaching the 12th TA would mean a progress of 100% - due to rounding errors from calculation of mean and the integer nature of an index, some conditions do not reach 100% at the last TA.

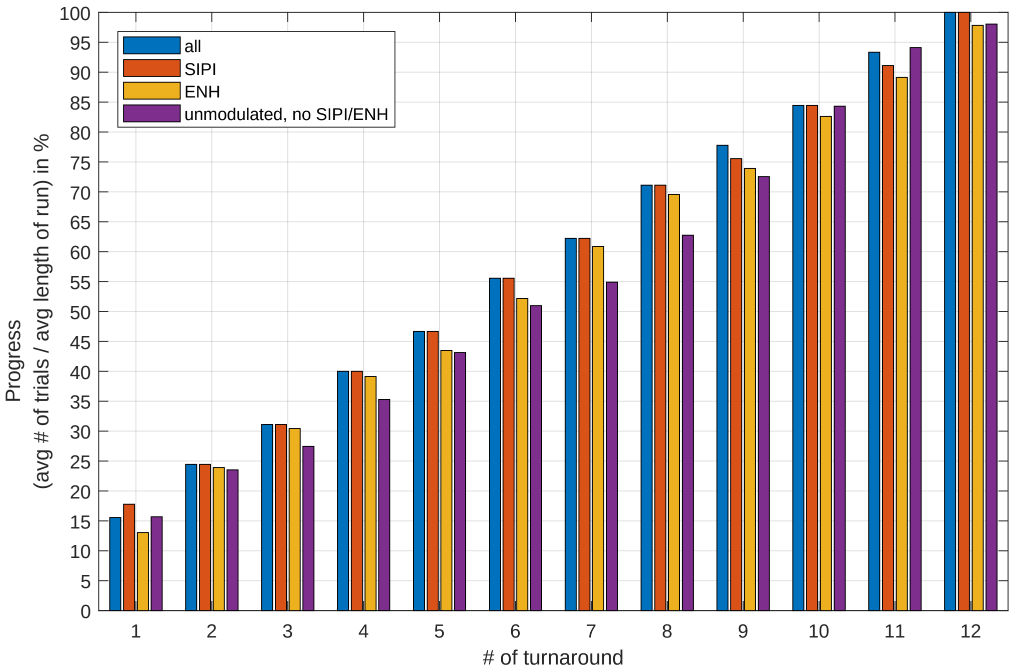


Figure 4.3 – Occurrence of turnarounds relative to length of staircase run

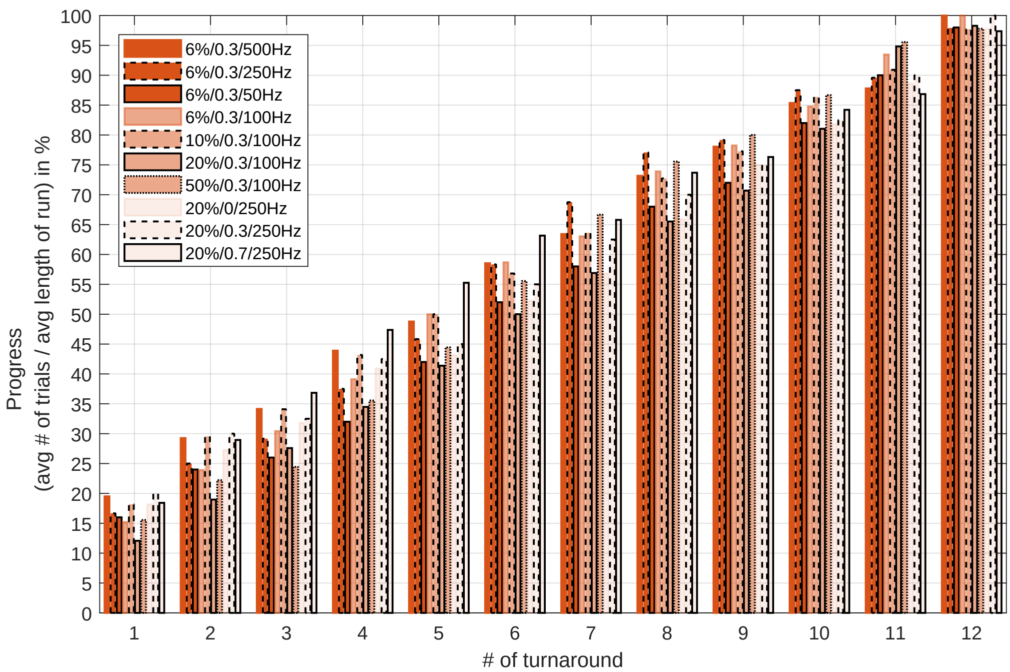


Figure 4.4 – SIPI conditions: Occurrence of turnarounds relative to length of staircase run

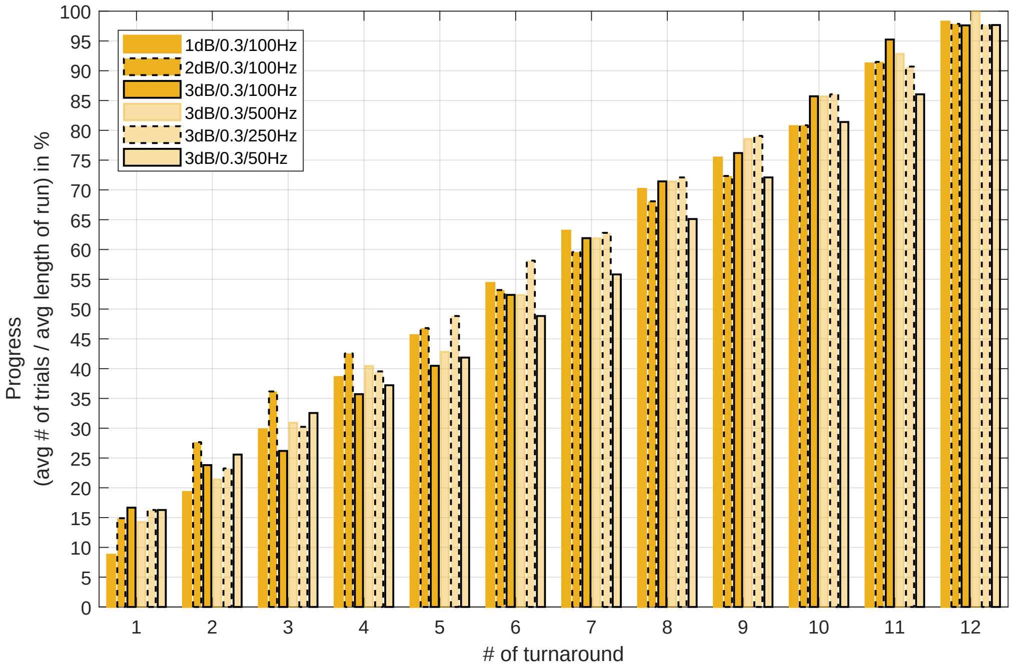


Figure 4.5 – ENH conditions: Occurrence of turnarounds relative to length of staircase run

4.3 Recalculation of Thresholds for varying Selected Number of Turnarounds

The most efficient way to reduce the time costs of the adaptive procedure is reduction of required TAs. However, is is assumed that the shortening directly lowers the precision of the obtained THRs (i.e. the balanced amplitude of the target). To check whether this might be true in case of the pilot, a re-calculation of THRs was conducted for smaller datasets, that cut runs of at 8, 9, 10 and 11 TAs. The number of TAs for THR calculation itself remained 8 for this examination.
In a first step the THRs for all five selectable numbers of TAs were obtained by calculating the mean of the last 8 TA amplitudes. This was done for all 68 items. Then the mean of THRs of every 4 associated items was computed for the original 12-TA-option. This value we state as our accessible optimum, so we can deduct it from each of the respective THRs. The deviation $Delta_{THR-THRmean}$ is gained. In Figures 4.6 and 4.7 the mean $Delta_{THR-THRmean}$ and the 95%-confidence interval ($CInt_{95\%}$) are plottet for each condition. The unmodulated, no-SIPI/ENH condition is shown in each for direct comparison and yields the smallest intervals.

In order to accept fewer than 12 TAs, one should insist on a reasonable boundary for the *Delta* displayed. In the CI the DR is quantized by steps of 1 current unit, implying a technical limit. It can be argued that a deviation of the mean that stays in the category $\pm 1cu$ is definetly acceptable because it is also minimal. Nevertheless, this difference might still be heard by CI users. The unsatisfactory coarseness of the cu steps is still a technical deficiency in CIs that should not be forgotten.

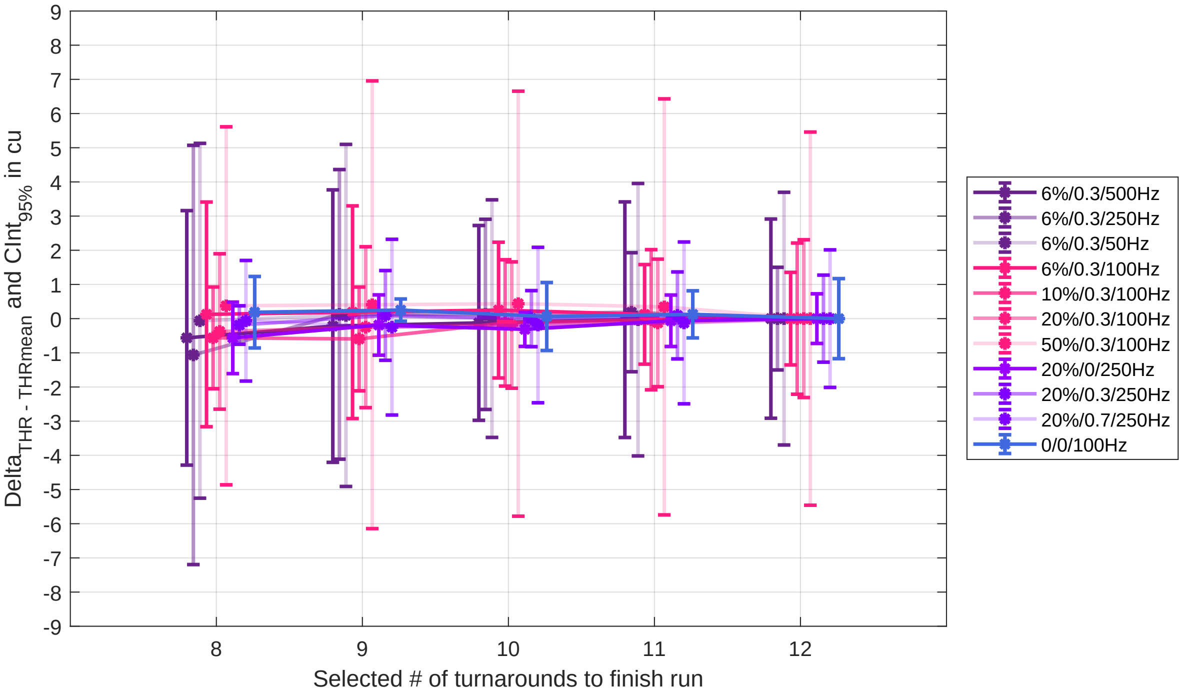


Figure 4.6 – THR recalculation for varying # of turnarounds: SIPI conditions

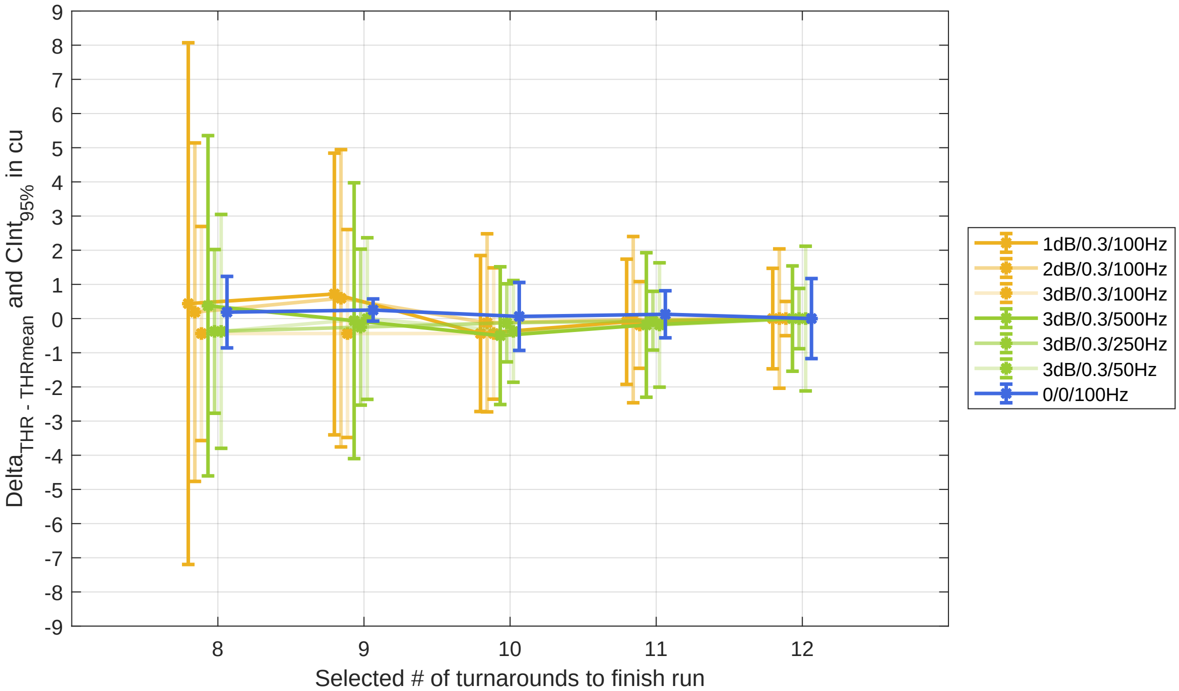


Figure 4.7 – THR recalculation for varying # of turnarounds: ENH conditions

Looking closely into the different conditions for SIPI pulses, one could state that all of the means stay within the acceptable limit for 9 TAs, but regarding the CInts, great differences can be spotted. It is assumed that condition 20%/0/250Hz can be reduced to 9 TAs within the acceptable limit, showing small confidence intervals. Slightly larger boundaries of the CInts can be spotted for conditions 20%/0.3/250Hz, 6%, 10%, 20%/0.3/100Hz and 6%/0.3/250Hz. The CInts show a large variation of individual values especially for 6% and 50% SIPI fractions. Also the MD seems to bring a factor of uncertainty to the task, as the intervals grows bigger for a larger MD in the 20% conditions.

As stated above, the unmodulated no-SIPI/ENH condition shows the smallest overall CInts and it can be argued that the required number of TAs can be therefore reduced. The means of ENH stimuli stay within the interval of $\pm 1cu$ up to the reduction to 8 TAs. What meets the eye is that for all the ENH conditions the spread of the CInts decreases with growing TA number, implying a greater certainty of the participant in balancing ENH stimuli. Condition 3dB/0.3/250Hz shows very small spreads up to 10 TAs, but also all the other variants display a very compact picture of CInts of $\pm 2cu$ in the range from 10-12 TAs.

In other words, accepting intervals of a larger size would allow reduction of the TAs to 10, which could reduce time costs by $\frac{1}{6}$ of the original 2min 15sec per run, yielding about 1min 52sec. For lowering to 9 TA this would even lead to reduction to $\frac{3}{4}$ (or to $\frac{2}{3}$ for of 8 TAs) of the original time spent. However, this isn't necessarily a clear and right decision to make. It has to be stated again, that the data only shows the answering behaviour for one person with an astonishing stability of the means. This might not be the case for other listeners and therefore the spread of the confidence interval is of greater importance.

5 Conclusion and Outlook

Software functionality and pilot results. The primary goal of this project was the implementation of the experimental software and the testing of functionality in a pilot experiment. No technical problems occoured during the experiment and functionality was proven. The participant expressed no complaints about the fitting process or the task itself and they are willing to take part in similiar experiments in the future. The results of the balancing task revealed the assumed loudness differences for SIPI and ENH stimuli from the reference, which confirms suitability of the task and procedure to detect loudness effects.

Evaluation of Adaptive Staircases. The postprocessing of the obtained staircase data showed details about the answering behaviour and possible improvements regarding the choice of forced numbers of TAs to complete the adaptive procedure. Though these assumptions are based on the results of only one person, they introduce parameters that express and visualize the recalculations of THRs in a way that enables some judgement about acceptable shortenings. Even small reductions could spare a considerable amount of the time costs for the loudness balancing that remains a necessary part of the pretests for numerous CI studies.

Outlook on Improvements. To make access to the data in the result .csv-tables easier for processing in Matlab, more data like the index of the turnaround could be directly logged. Also the recalculation of THRs with all the mentioned options could be integrated in the SIPILoudness app to get a glimpse at the answering behaviour directly after tasks are completed.

The evaluation of the staircase procedure could be done for a larger dataset of more participants, to get more accurate answers about average run durations, TA locations and necessary TA numbers and see if there are small or large deviations from the pilot data.

References

Jesteadt, W. (1980), 'An adaptive procedure for subjective judgments', *Perception & Psychophysics* .

Laback, B. and Majdak, P. (2008), 'Binaural jitter improves interaural time-difference sensitivity of cochlear implantees at high pulse rates', *Proceedings of the National Academy of Sciences* **105**(2), 814–817.

Levitt, H. (1971), 'Transformed Up-Down Methods in Psychoacoustics', *The Journal of the Acoustical Society of America* **49**(2), 467–477.

Lindenbeck, M. J. (2017), Pitch SIPI. Experiments for Rate Pitch Perception in Cochlear-Implant Listeners with Short-Interpulse-Interval (SIPI) Stimulation, Technical report, Acoustic Research Institute, Vienna.

Lindenbeck, M. J., Laback, B., Majdak, P. and Srinivasan, S. (2020), 'Temporal-pitch sensitivity in electric hearing with amplitude modulation and inserted pulses with short inter-pulse intervals', *The Journal of the Acoustical Society of America* **147**(2), 777–793.

McNicol, D. (1972), *A Primer of Signal Detection Theory*, Lawrence Erlbaum Associates, Inc.

Mihocic, M. (2014), ExpSuite: Open Source-Framework zur Durchführung psychoakustischer Experimente und akustischer Messungen, Master's thesis, Fachhochschule Technikum Wien.

Srinivasan, S., Laback, B., Majdak, P. and Arnoldner, C. (2020), 'Improving interaural time difference sensitivity using short inter-pulse intervals with amplitude-modulated pulse trains in bilateral cochlear implants', *Journal of the Association for Research in Otolaryngology* **21**(1), 105–120.

Srinivasan, S., Laback, B., Majdak, P. and Delgutte, B. (2018), 'Introducing short interpulse intervals in high-rate pulse trains enhances binaural timing sensitivity in electric hearing', *Journal of the Association for Research in Otolaryngology* **19**(3), 301–315.

Wilson, B. S., Finley, C. C., Lawson, D. T. and Zerbi, M. (1997), 'Temporal representations with cochlear implants', *The American Journal of Otology* **18**(6 Suppl), S30–4.

List of Figures

List of Tables