MASTER THESIS

COMPARISON OF ADAPTIVE PSYCHOPHYSICAL METHODS FOR NOCICEPTIVE THRESHOLD TRACKING: A SIMULATION AND HUMAN SUBJECT STUDY.

R.J. Doll (s1017209)

FACULTY OF ELECTRICAL ENGINEERING, MATHEMATICS AND COMPUTER SCIENCE
BIOMEDICAL SIGNALS AND SYSTEMS

EXAMINATION COMMITTEE
prof. dr. ir. P.H. Veltink
dr. ir. J.R. Buitenweg
dr. H.G.E. Meijer
dr. ir. W. Olthuis

DOCUMENT NUMBER
BSS – 12-05

UNIVERSITY OF TWENTE.

24TH AUGUST 2012

Title:

Comparison of adaptive psychophysical methods for nociceptive threshold tracking: a simulation and human subject study.

Short title:

Comparison of psychophysical methods for threshold tracking

Authors:

Robert J. Doll[a], Jan R. Buitenweg[a], Hil G.E. Meijer[b], Peter H. Veltink[a]

[a]Biomedical Signals and Systems, MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente

[b]Applied Analysis and Mathematical Physics, MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente

Corresponding author:

Robert-Jan Doll, Biomedical Signals and Systems, University of Twente, Zuidhorst, room ZH222, Drienerlolaan 5, Postbus 217, Enschede. E-mail: r.j.doll@utwente.nl

Keywords:

psychophysics, stimulus selection procedure, threshold estimation method, threshold tracking, electrocutaneous stimulation, nociception, conditioning stimulus

Possible journal (1):

Behavior research methods (IF: 2.116)

(http://www.springer.com/psychology/cognitive+psychology/journal/13428)

Possible journal (2):

Psychological methods (IF: 4.449)

(http://www.apa.org/pubs/journals/met/index.aspx

1 ABSTRACT

Tracking psychophysical thresholds over time is useful for the investigation of dynamic behavior of underlying mechanisms. For example, noxious events can activate endogenous analgesic mechanisms reflected in an increasing nociceptive threshold. Adequate psychophysical methods for threshold tracking should include an efficient stimulus selection procedure and threshold estimation method. This study compares the performance of adaptive stimulus selection procedures and estimation methods in both a simulation and a human subject study.

Monte Carlo simulations were performed to compare bias and precision of threshold estimates of three stimulus selection procedures (simple staircase, random staircase and a minimum entropy procedure) and two threshold estimation methods (logistic regression and Bayesian estimation). Logistic regression was found to result in more precise estimates than Bayesian estimates in all simulations. Moreover, estimates were more precise with the simple staircase procedure than the other procedures. However, the random staircase procedure is less sensitive to different procedure specific settings (e.g. step-size) than the other procedures.

The simple staircase and random staircase procedures, both using logistic regression, were compared in a human subject study (N=30). A cold pressor was applied as nociceptive conditioning stimulus. Electrocutaneous stimulation was used for nociceptive detection threshold tracking before, during and after the conditioning stimulus. Both procedures were able to detect habituation as well as changes induced by the conditioning stimulus, but highest precision was obtained with the random staircase procedure.

Based on these results, we recommend the random staircase procedure in combination with logistic regression for threshold tracking experiments.

2 INTRODUCTION

Tracking psychophysical thresholds over time is useful for investigation of dynamic behavior of underlying mechanisms. It is, for example, known that noxious events, such as disease (e.g. hyperalgesia), clinical intervention (Yarnitsky, Crispel et al. 2008; Wilder-Smith, Schreyer et al. 2010) or experimental conditioning stimuli (e.g. cold pressor, see Pud, Granovsky et al. (2009)) can activate endogenous analgesic mechanisms, as reflected in an increasing nociceptive threshold. Observing the dynamics of the nociceptive thresholds during a noxious event might provide more information on underlying nociceptive mechanisms for clinical applications, e.g. pre-operative detection of an increased risk of chronic pain development (Yarnitsky, Crispel et al. 2008; Wilder-Smith, Schreyer et al. 2010; Yarnitsky, Granot et al. 2012).

Stationary psychophysical thresholds can be estimated using yes-no experiments, in which multiple stimulus amplitudes and corresponding responses (perceived or not) are used to probe the psychophysical function of the subject (Gescheider 1985; Treutwein 1995; Klein 2001; Kingdom and Prins 2009). Threshold measurements therefore include (1) stimulus selection procedures for collection of stimulus-response pairs and (2) estimation methods determining the most likely threshold from the collected stimulus-response pairs.

Stimulus selection procedures can be either adaptive or non-adaptive. Non-adaptive procedures select new stimuli independent of preceding stimulus-response pairs. An example of a non-adaptive selection procedure is the *method of constant stimuli* (Simpson 1988). Stimuli are selected randomly from a set of pre-defined amplitudes, thus allowing global probing of the psychophysical function. However, due to the fixed range of pre-defined amplitudes and a relatively large number of required stimuli, this procedure is inefficient compared to adaptive procedures (Watson and Fitzhugh 1990).

Adaptive procedures select new stimuli based on preceding stimulus-response pairs. They have been demonstrated to be more efficient than non-adaptive procedures (Leek 2001). A widely used adaptive stimulus selection procedure is the *simple up-down staircase* procedure (Cornsweet 1962). The stimulus amplitude is increased after a negative response to the preceding stimulus and decreased after a positive response. In this way, probing around the threshold of the psychophysical function is achieved. However, subjects may anticipate to the sequential order of stimuli when identified as such (Levitt 1971; Ehrenstein and Ehrenstein 1999).

A more advanced stimulus selection procedure is one where stimuli are placed that will minimize the expected entropy in a posterior probability distribution and thus will provide most information about the psychophysical curve (Kontsevich and Tyler 1999; Kujala and Lukka 2006). This *minimum entropy* procedure requires a prior probability distribution, which is continuously updated with new stimulus-response pairs according to Bayes' rule. Due to this requirement, assumptions have to be made about the possible range of psychophysical parameters (e.g. threshold). Especially when thresholds need to be tracked over time, large ranges might be necessary and increase the computational complexity (Kontsevich and Tyler 1999; Kujala and Lukka 2006; Kingdom and Prins 2009).

Since responses to stimuli are either perceived or not perceived, *logistic regression* can be used to estimate the threshold (Hosmer and Lemeshow 2000). Alternatively, the results of every new obtained stimulus-response pair can be used to update the posterior probability distribution according to *Bayes'* rule (Watson and Pelli 1983; King-Smith and Rose 1997; Kontsevich and Tyler 1999; Treutwein and Strasburger 1999; Kujala and Lukka 2006).

For tracking thresholds over time, a moving time window, which includes a number of most recently collected stimulus-response pairs, can be used (von Dincklage, Hackbarth et al. 2009). The length of this time window is determined by (1) the number of stimulus-response pairs used for threshold estimation and (2) the inter-stimulus intervals, e.g. the time subjects need to indicate if stimuli are perceived. For adequate tracking of nonstationary thresholds over time, this time window should be sufficiently small, which limits the number of stimulus-response pairs available for momentary threshold estimations. However, fewer stimulus-response pairs will result in higher estimation bias and lower estimation precision. Therefore, efficient stimulus selection procedures and threshold estimation methods are crucial for successful threshold tracking.

We introduce a new stimulus selection procedure, which we think overcomes disadvantages of the above-described procedures. Stimuli are selected randomly from a small pre-defined set of amplitudes (as with the *method of constant stimuli*). All amplitudes in the set are increased with a fixed step-size after a not-perceived stimulus and decreased after a perceived stimulus (as with the *simple staircase procedure*). In this way, both the calculation complexity (as with the *minimum entropy* procedure) and stimulus predictability (as with the *simple staircase* procedure) are minimal. We refer to this new procedure as the *random staircase* procedure.

Even though comparison of performance of stimulus selection procedures and estimation methods for stationary thresholds is frequently done, performance during determination of nonstationary thresholds over longer periods of time have not been extensively studied before. Therefore, the aim of this study is to compare the performance of adaptive stimulus selection procedures and threshold estimation methods for the use in psychophysical threshold tracking experiments.

We performed Monte Carlo simulations of a modeled psychophysical experiment to compare three adaptive procedures (the simple staircase, random staircase and minimum entropy procedure) and two estimation methods (logistic regression and Bayesian estimation) on bias and precision of threshold estimates during both stationary and nonstationary thresholds. Next, two stimulus selection procedures are compared in a human subject study. Nociceptive detection thresholds were tracked using electrocutaneous stimulation. A cold pressor task was used to perturb the nociceptive system.

3 MONTE CARLO SIMULATIONS

3.1 Methods

The bias and precision of three adaptive stimulus selection procedures and two estimation methods were compared by means of Monte Carlo simulations. A stochastic psychophysical model was applied to simulate responses to stimuli. New stimuli are selected according to the stimulus selection procedures. The simulated stimulus-response pairs were used in the estimation methods to obtain the bias and precision of threshold estimates.

3.1.1 Psychophysical model

The detection probability p to stimulus amplitude x was modeled with a logistic psychometric function:

$$p(x) = \left(1 + \exp\left(\frac{\alpha - x}{\beta}\right)\right)^{-1} \text{ (eq. 1)}$$

The slope of the logistic function was fixed at β=0.05 mA. A threshold of α=0.3 mA was taken as stationary threshold. Responses to stimuli with amplitude x were set as perceived if $p(x) > \varepsilon$, with ε a random number taken from a uniform distribution between 0 and 1, and set to not-perceived otherwise. The random number generator was shuffled each time a new simulation was started.

3.1.2 Simple staircase stimulus selection

New stimuli, according to the simple staircase procedure, are determined taking into account the response to the preceding stimulus: a not-perceived stimulus results in an increment of stimulus amplitude with a fixed step-size or a perceived stimulus results in a decrement of stimulus amplitude with a fixed step-size. The initial stimulus amplitude is set to zero. A fixed step-size of 0.05, 0.1, or 0.2 mA is used for both increment and decrement steps.

3.1.3 Random staircase stimulus selection

The random staircase procedure starts with a pre-defined set of a number of equidistant stimulus amplitudes (NoA) of NoA=5 or NoA=10, with the lower and upper values set to 0 and 0.3 mA respectively (thus, amplitudes are separated with 0.075 mA and 0.033 mA, respectively). New amplitudes are randomly selected from this set. All amplitudes in the set are increased with a fixed step-size after a not-perceived stimulus and decreased after a perceived stimulus with the same step-size. The step-size is set to either 0.05 or 0.1 mA.

3.1.4 Minimum entropy stimulus selection

The minimum entropy procedure used in this study is based on the procedure described by Kontsevich and Tyler (1999). New stimulus amplitudes are chosen such that they will minimize the expected entropy in the posterior probability distribution. Similar to the work by Kontsevich and Tyler (1999), only a Bayesian estimation method is used to obtain threshold estimates from the stimulus response pairs.

3.1.5 Threshold estimation methods

Two estimation methods were compared in simulations: logistic regression (Hosmer and Lemeshow 2000) and Bayesian estimation (Treutwein 1995; King-Smith and Rose 1997; Kingdom and Prins 2009).

For the implementation of the Bayesian estimation method, a logistic function (Eq. 1) was chosen for conditional probabilities of responses to stimulus amplitudes given thresholds α and slopes β. Moreover, different ranges and resolution of the threshold α, slope β and stimulus amplitude x in the probability distribution (PD) were used in simulations (Table 1). PD #1 and PD #2 are chosen such that the true slope β is at the edge of the slope range. PD #3 is chosen such that the true slope β is in the middle of the slope range. A uniform *a priori* probability distribution was assumed at the start of each simulation.

Table 1. Settings used for the probability distribution (PD): the range of the threshold α, slope β and stimulus amplitude x were varied.

PD setting (#)	Threshold α range [mA]			Slope β range [mA]			Amplitude x range [mA]		
	Min	Step	Max	Min	Step	Max	Min	Step	Max
1	0	0.01	1	0	0.05	1	0	0.01	1
2	0	0.01	2	0	0.05	1	0	0.01	2
3	0	0.015	1	0	0.0025	0.1	0	0.015	1

3.1.6 Simulations

Adaptive stimulus selection procedures and estimation methods were compared in two situations: 1) during a stationary threshold α (Fig. 1A-C) and 2) during a nonstationary threshold α (Fig. 1D-F). The stationary threshold situation estimates thresholds based on 15, 20, 25 and 30 stimulus-response pairs. Each simulation was repeated 5.000 times. The nonstationary threshold situation simulates a ten minute experiment where the threshold α changes from 0.3 to 0.6 mA after five minutes. The momentary threshold is estimated by using the preceding 25 stimulus-response pairs. Each simulation was repeated 1.000 times.

In our human subject experiments, subjects need to hold a button until detection of a stimulus (see section 4.1.5). Due to this, the time between two stimuli is shorter when the first is not perceived than when the first is perceived. When thresholds are stationary, this timing effect is not of interest, but it is for nonstationary thresholds. Therefore, the nonstationary threshold situation includes inter-stimulus intervals of 1.5 s and 3.5 s after a not-perceived and perceived stimulus respectively.

3.1.7 Analysis

Bias and precision of threshold estimates were determined in all simulations (Treutwein 1995). Bias was defined as the mean difference between the true threshold and the estimated threshold and precision was defined as the reciprocal of the variance of threshold estimates. To get equally spaced estimates, nonstationary results are linearly interpolated. A rate of 1 Hz is used for interpolation to prevent under sampling. All simulation models and analyses were performed with MATLAB 7.14.

3.2 Results

3.2.1 Stationary threshold

Bias was found to be lower than 1% of the true threshold for all but one simulation. A bias larger than 1% was found when combining the simple staircase procedure (step-size=0.1 mA) with Bayesian estimation (PD #1). Precision for all stimulus selection procedures and estimation methods are presented in Fig. 2. Precision increased when more stimulus-response pairs were included in the estimation for all selection procedures and estimation methods.

Estimation precision was found highest with the simple staircase procedure in almost all cases. The precision of the minimum entropy is higher than the random staircase procedure when the true slope β is in the middle of the PD (i.e. PD #3). Otherwise, precision of the random staircase procedure is higher than the precision of the minimum entropy procedure.

The simple staircase procedure showed increasing precision with smaller step-sizes (Fig. 2A). The precision of the random staircase procedure is similar for different settings (i.e. step-size, NoA and estimation method) (Fig. 2B). Logistic regression as estimation method resulted in higher precision than Bayesian estimation in both the simple staircase and random staircase procedure. The minimum entropy procedure is more precise when the true slope β is in the middle of the PD (i.e. PD #3) than when the true slope β is at the edge of the PD (i.e. PD #1 and PD #2. Fig. 2C).

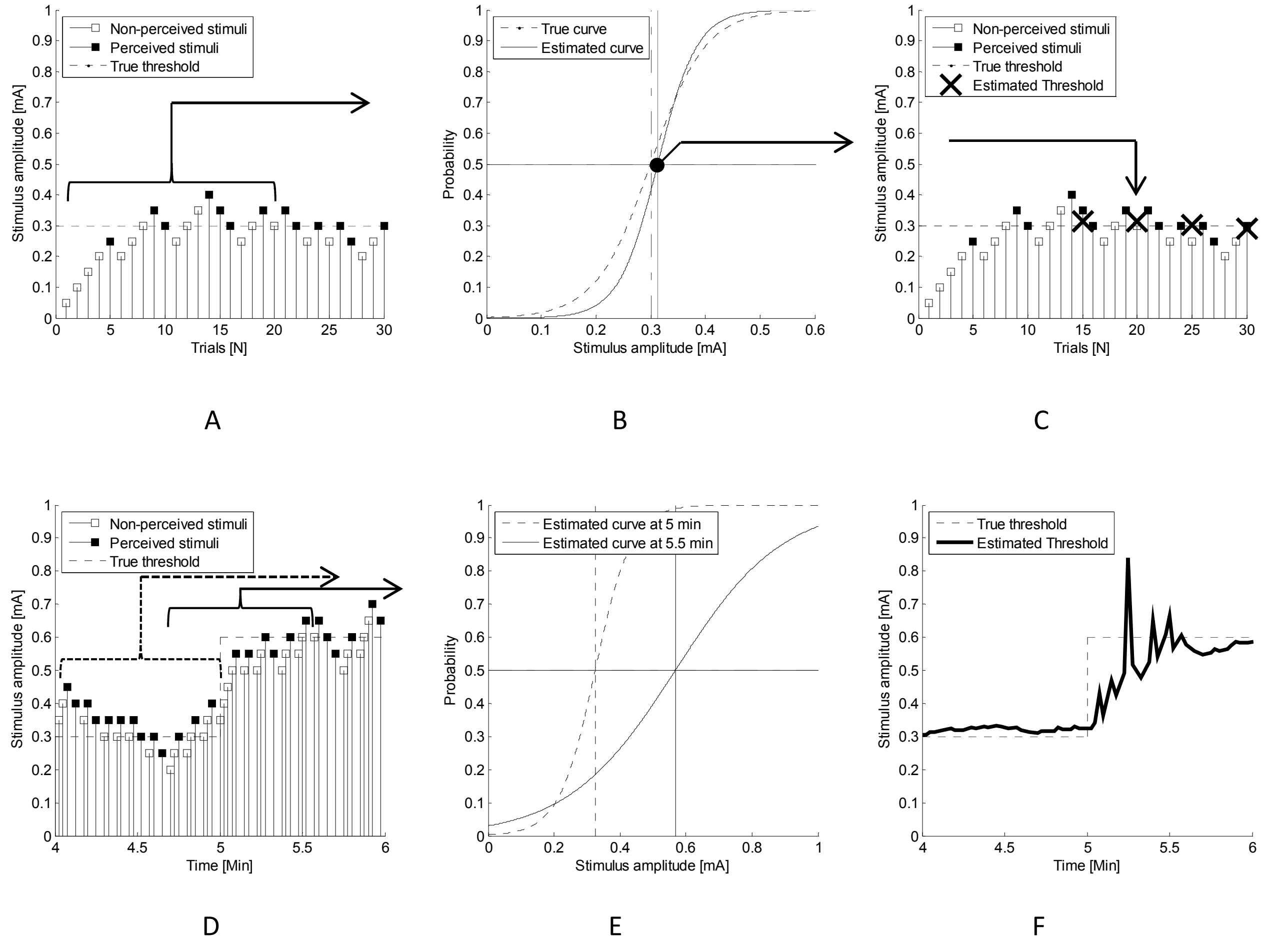


Fig. 1. Simulation procedures: (A-C) stationary threshold simulations: a window including N stimulus-response pairs (A) is used in estimation methods to obtain the threshold estimate (B). (C) A single simulated experiment. (D-F) nonstationary threshold simulations: a moving window including 25 stimulus-response pairs (D) is used to estimate (E) and track (F) thresholds over time.

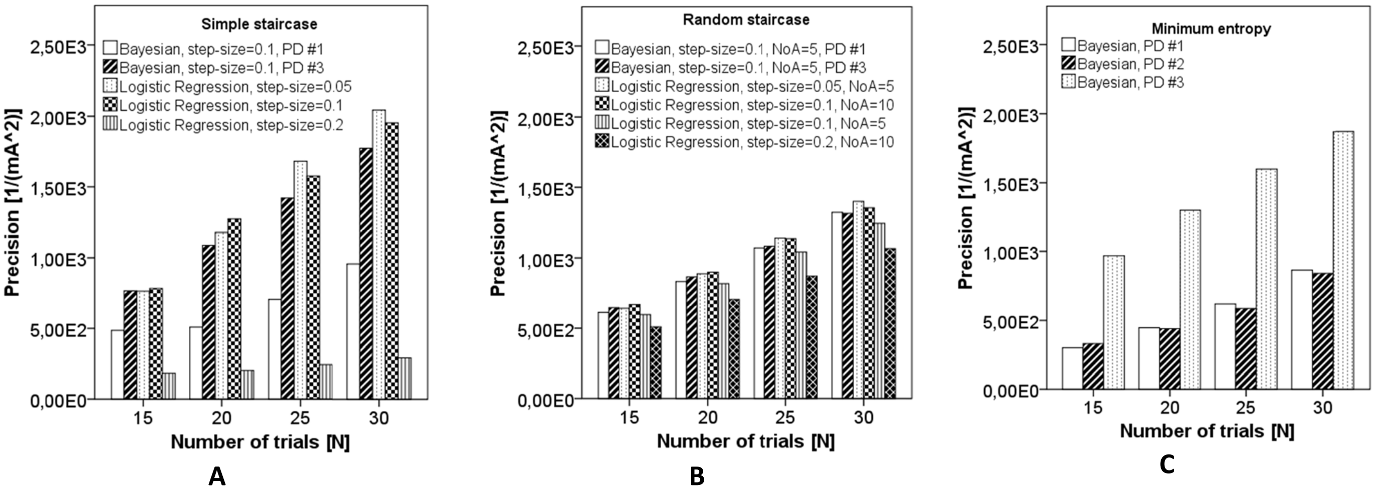


Fig. 2. Precision of simulated adaptive procedures with different settings. (A) Precision of the simple staircase procedure: the step-size, estimation method and PD settings (Table 1) are varied. (B) Precision of the random staircase procedure: the step-size, number of amplitudes (NoA), estimation method and PD number are varied. (C) Precision of the minimum entropy procedure: the PD settings are varied.

3.2.2 Nonstationary threshold

Based on the results of the stationary threshold simulations, we chose to simulate nonstationary thresholds with two settings of the simple staircase procedure: step-size=0.05 and 0.1 mA. Three settings were chosen for the random staircase procedure: step-size=0.05 mA and NoA=5, step-size=0.1 mA and NoA=10, and step-size=0.1 mA and NoA=5. PD #1 and PD #3 were simulated for the minimum entropy procedure.

The bias and precision over time around the change in threshold (i.e. between 4.5 and 7 minutes) is shown in Fig. 3. The bias is displayed in the upper graphs, precision in the lower graphs. The bias is negligible in all simulations before the change in threshold occurs. All procedures and methods need similar time to converge to negligible bias after the change in threshold. Moreover, all simulations showed a decrease in precision after a change in threshold occurred. The time necessary to return to a stable precision with the simple staircase and random staircase procedure is similar. The minimum entropy procedure needed more time to converge to a stable precision.

Before the change in threshold, the simple staircase procedure showed higher precision in threshold estimates than the other procedures. Right after the change in threshold, the precision of the simple staircase and random staircase procedure are similar. However, with a smaller step-size, the precision of the simple staircase procedure is drastically reduced. Right after the change in threshold, the precision of the minimum entropy procedure is lower than the other procedures. Before the change in threshold, precision of the minimum entropy is only higher than precision of the random staircase procedure when the true slope β is in the middle of the PD.

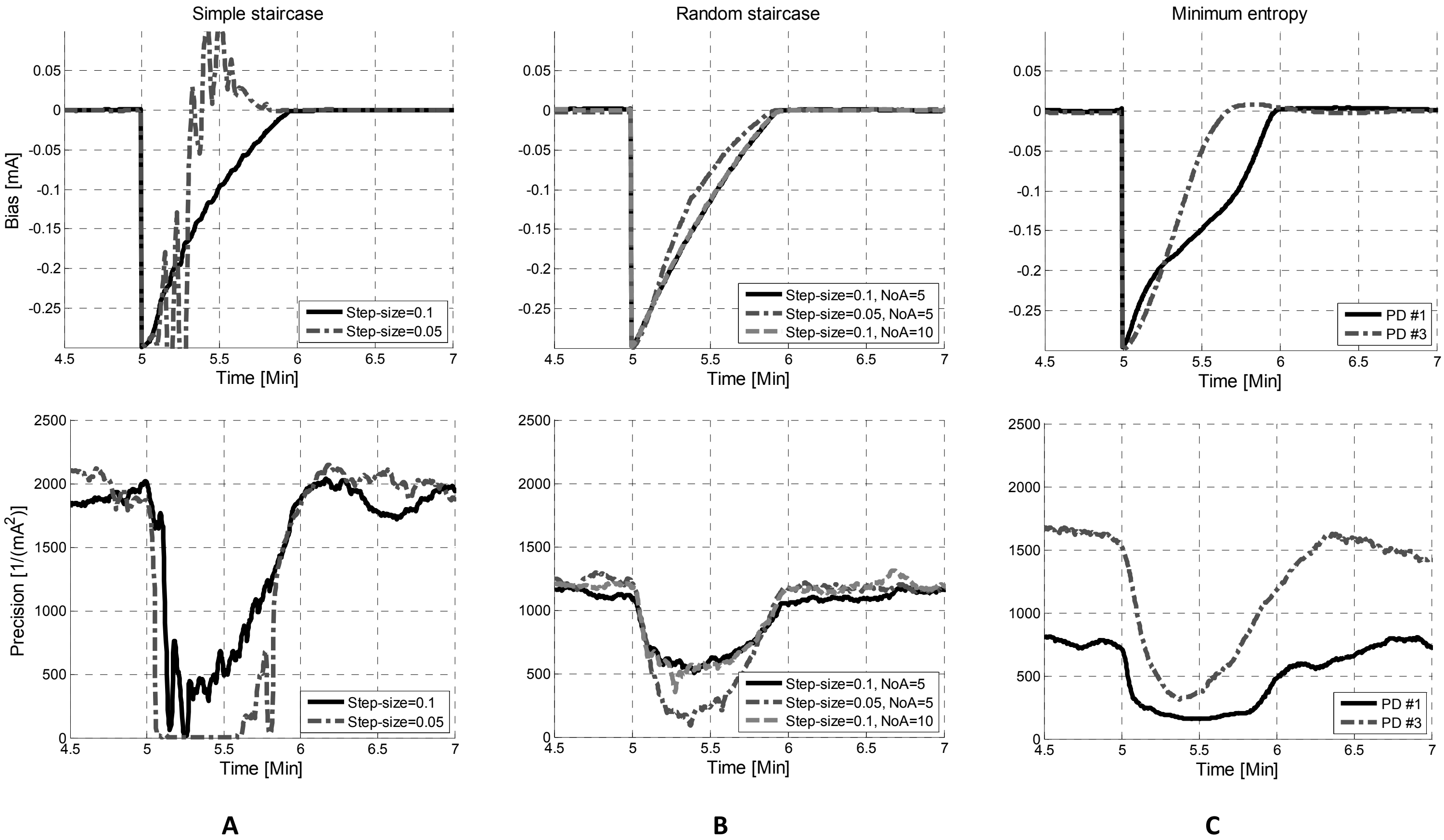


Fig. 3. Bias and precision over time (between 4.5 and 7 minutes) of simulated adaptive procedures with different settings. (A) Bias and precision of the simple staircase procedure, (B) bias and precision of the random staircase procedure and (C) bias and precision of the minimum entropy procedure.

4 HUMAN SUBJECT STUDY

The simple staircase procedure and random staircase procedure, both in combination with logistic regression, were compared in a human subject study. Based on the simulation results, we decided to exclude the minimum entropy procedure and the Bayesian estimation method.

4.1 Methods

4.1.1 Subjects

The two stimulus selection procedures were compared in a psychophysical experiment including 31 pain free human subjects (20 men and 11 women, three left-handed) aged 24.38±2.86 (mean±SD), with a range of 19—32 years. The local Ethics Committee approved all experiment procedures. All subjects provided written informed consent and were rewarded a gift voucher after participation of the experiment.

4.1.2 Test-stimuli

Single cathodic square wave electrical pulses with a pulse-width of 525 µs were applied as test-stimuli using a compound electrode attached to the subjects' left forearm (Steenbergen, Buitenweg et al. 2012). The compound electrode consists of an array of five needles and an array of four flat electrodes. The needles served as cathode and the flat electrodes as anode. Inter-stimulus interval times were randomly varied between 600-1000 ms. Electrocutaneous stimulation using a needle electrode is used as it is shown to selectively stimulate nociceptive related Aδ-fibers using low current amplitudes (Mouraux, Iannetti et al. 2010; Inui and Kakigi 2012).

4.1.3 Conditioning stimulus

A three minute cold pressor task was used as a nociceptive conditioning stimulus (Talbot, Duncan et al. 1987; Mitchell, MacDonald et al. 2004; Pud, Granovsky et al. 2009). Subjects were asked to immerse their right hand up to the wrist into a polystyrene container filled with water and crushed ice (water temperature is between 0-3 degrees Celsius). Subjects were allowed to remove their hand from the water when pain was no longer tolerable. However, they were instructed to continue with the protocol.

4.1.4 Protocol

The simple staircase procedure is compared with the random staircase procedure, both using logistic regression to obtain thresholds. A computer program specially programmed for our purposes (LabVIEW 2011, SP1) controlled all stimulation procedures as well as registration of stimuli in mA, time in ms and responses to stimuli. Momentary thresholds were estimated after every stimulus, including the preceding 25 stimulus-response pairs.

A step-size of 0.1 mA was used for both ascending and descending stairs in the simple staircase as well as for the random staircase method. The random staircase procedure had an initial NoA of five amplitudes, equidistantly separated between 0 and 0.3 mA.

4.1.5 Procedure

The experiment procedure was subdivided in two tests: a static and a dynamic test. The static test took ten minutes whereas the dynamic test took 23 minutes and included a cold pressor task between the fifth and eighth minute (Fig. 4). Subjects were instructed to indicate perceived stimuli by means of releasing a response button and to press the button again after about half a second to a second. Stimuli were chosen according to either the simple staircase procedure or the random staircase procedure. Both procedures were randomly, but balanced, alternated after a perceived stimulus (Fig. 5). Subjects were familiarized with the test-stimuli before the start of both tests by applying several test-stimuli of various amplitudes.

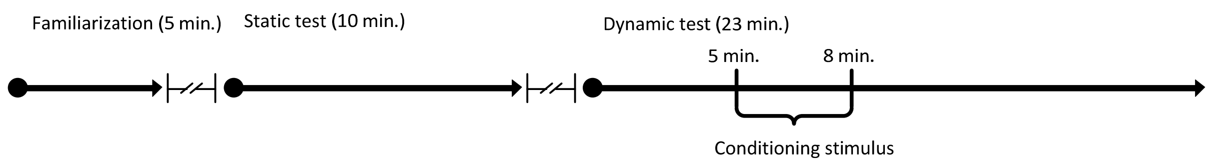
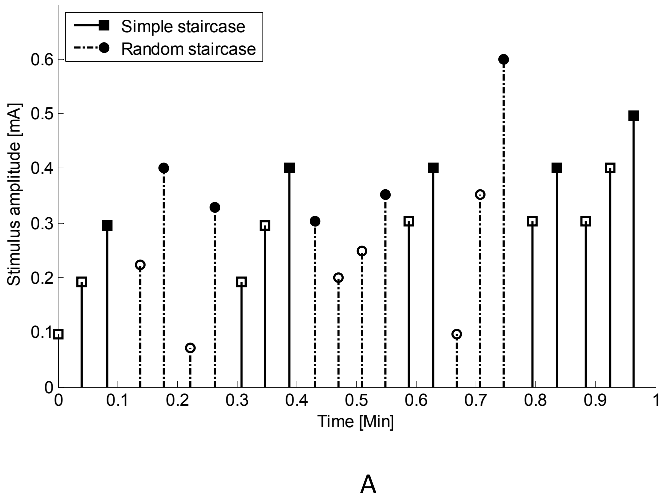


Fig. 4. Schematic representation of the experiment timeline.



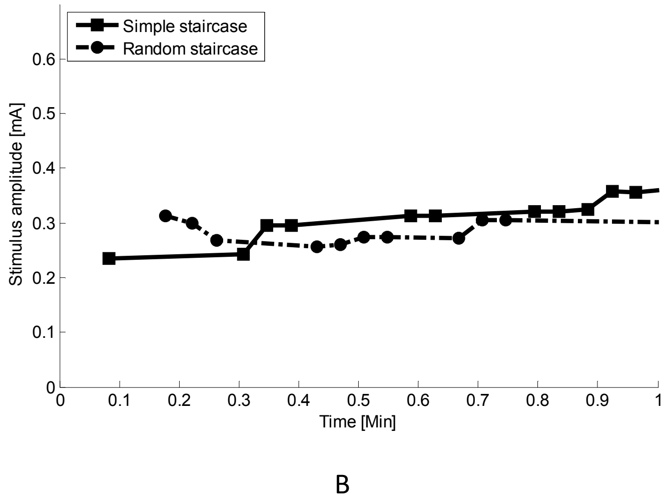


Fig. 5. Stimulus application in the human subject study. Stimuli are randomly, but balanced, selected by either the simple staircase procedure or the random staircase procedure. (A) Open markers indicate non-perceived stimuli, filled markers indicate perceived stimuli.

4.1.6 Data analyses

All data preparation was performed in MATLAB. Linear mixed model analyses (LMM) were performed in SPSS Statistics 20.0. Default settings are used for all other options in all LMMs unless stated otherwise. All other analyses were performed in MATLAB.

Static test. Estimated thresholds during the first two minutes were excluded from the dataset of all subjects to exclude possible procedural start-up effects. Thresholds were analyzed using LMM analysis. The model estimated the fixed effects of time, method and intercepts, fixed interaction effects between time and procedure as well as random effects of time (covariance type: AR(1): Heterogeneous) on threshold estimates. The estimation method was set to Maximum Likelihood in the LMM as it is said to be a better estimator of fixed effects (Twisk 2006).

Residuals of LMM estimates were saved and exported to MATLAB. We assume that less residuals indicate a lower variance of threshold estimates and thus a higher precision. An F-test was used to determine whether there is a difference between residuals per procedure.

Dynamic test Estimated thresholds during the first two minutes were excluded from the dataset of all subjects to exclude possible procedural start-up effects. Thresholds were analyzed using LMM analysis. The model estimated the fixed effects of time, method, conditioning stimulus and intercept, the fixed interaction effect between time and procedure as well as the random effects of time and condition stimulus (covariance type: AR(1): Heterogeneous). The estimation method was set to Maximum Likelihood in the LMM as it is said to be a better estimator of fixed effects (Twisk 2006). As subjects were allowed to remove their hand from the water whenever they wanted, the conditioning stimulus time varied. For this reason, we defined conditioning stimulus to have three time intervals for each subject: before, during and after the cold pressor task.

4.2 Results

A total of 31 subjects participated the experiment. One subject stopped the procedure after becoming unwell during the dynamic test, leaving a total subject population of N=30.

4.2.1 Static test

The mean and standard deviations of threshold estimates over time of all subjects are presented in Fig. 6A. As can be seen from the graph, the threshold increased gradually over time. With the simple staircase procedure in a few subjects, higher thresholds were estimated between 4.5 and 7 minutes than with the random staircase procedure.

The results of the LMM are presented in Table 2. Thresholds significantly increased over *Time*. No significant effect was found for *Procedure* as well as for the interaction effect between *Procedure* and *Time*. Estimated regression parameters are presented in Table 3. In addition, a significant difference was found between the LMM output residuals ($F_{(2522,2507)}$=8.44, p<0.001). Residuals of the simple staircase procedure were found to be higher than the random staircase procedure.

Table 2. Linear mixed model fixed effects on tracked thresholds results for the static test

Factor	df[a]	F	p
Procedure	1/4941.5	0.492	0.483
Time	1/30.266	33.7	<0.001
Procedure × Time	1/4941.5	2.19	0.139

[a]Numerator/denominator degrees of freedom

Table 3. Parameter estimates of fixed effects of the linear mixed model of threshold estimates as a function of time.

Parameter	Estimate	95% confidence interval
Intercept	0.188 [mA]	[0. 127 0.247]
Time	0.0307 [mA/min]	[0.0191 0.424]

4.2.2 Dynamic test

Twenty-four subjects immersed their hand for three minutes. The other six immersed their hand for 88±34 seconds (mean±SD). The mean and standard deviations of threshold estimates over time of all subjects are presented in Fig. 6B. As can be seen from the graphs, thresholds of both procedures are similar, also during the cold pressor task (between the fifth and eighth minutes). As in the static test, a gradual increase of threshold over time is visible.

It can be seen from Fig. 6B that thresholds increase right after immersion of the hand into the water. Moreover, after removal of the hand, the estimated threshold does not immediately decrease, but remain increased for about one to two minutes.

The results of the LMM are presented in Table 4. Significant effects were found for *Conditioning stimulus*, *Time* as well as the interaction effect between *Procedure* and *Time*, and interaction effect between *Conditioning stimulus* and *Time*. No significant effect was found for *Procedure*. Estimated regression parameters are presented in Table 5.

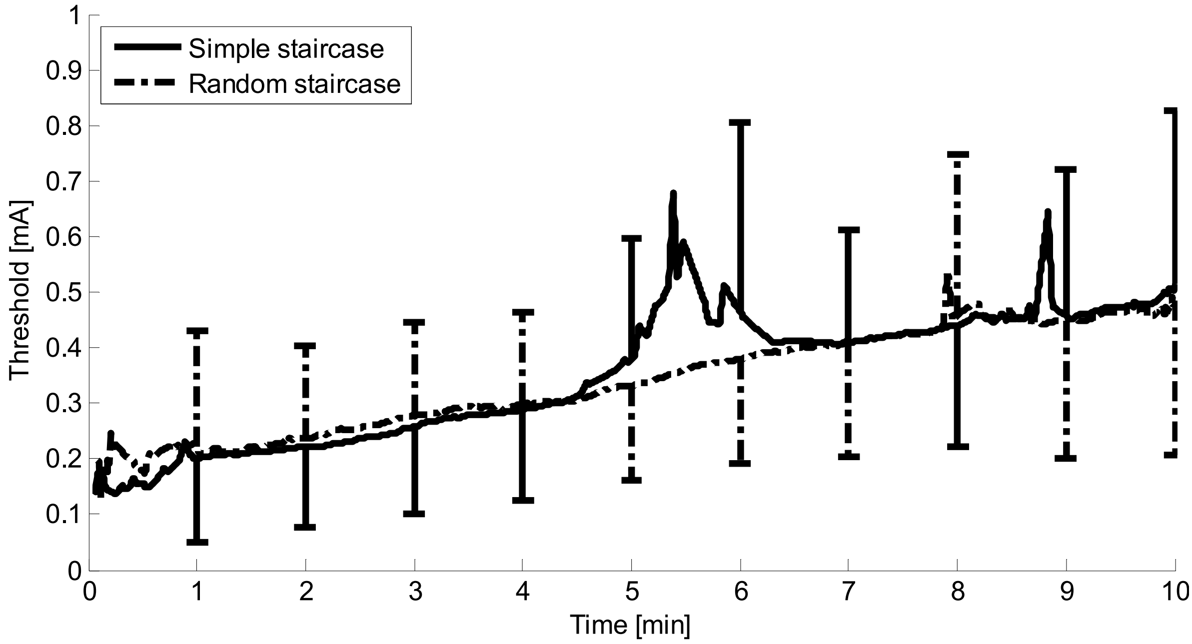
Table 4. Linear mixed model fixed effects results for the dynamic test

Factor	df[a]	F	p
Procedure	1/13067.2	1.43	0.232
Conditioning stimulus	2/30.3	41.5	<0.001
Time	1/18.004	62.0	<0.001
Procedure × Time	1/13054.3	4.35	0.037
Conditioning stimulus × Time	2/30.62	20.21	<0.001

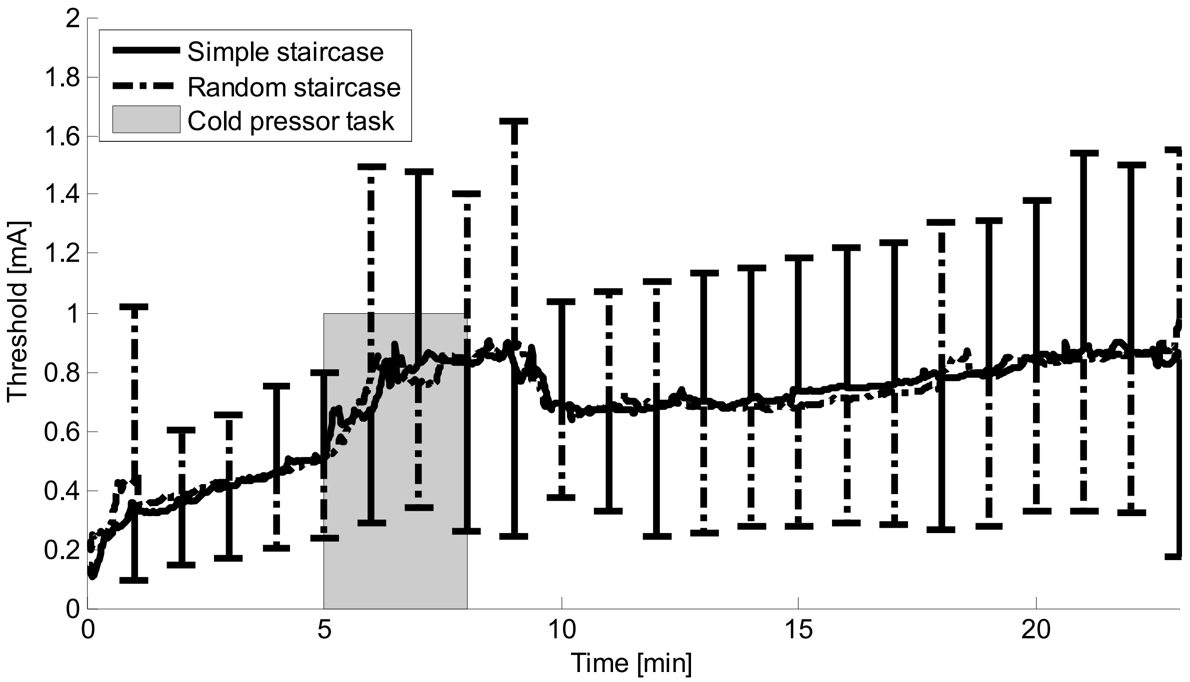
[a]Numerator/denominator degrees of freedom

Table 5. Parameter estimates of fixed effects of the linear mixed model of threshold estimates as a function of time and conditioning stimulus (before, during and after conditioning stimulus).

Parameter	Estimate	95% confidence interval
Intercept	0.628 [mA]	[0.512 0.764]
Conditioning stimulus		
Before	-0.355 [mA]	[-0.454 -0.256]
During	-0.418 [mA]	[-0.604 -0.232]
Time	0.00797 [mA/min]	[-0.0004 0.0163]
Procedure × Time		
Simple staircase	0.00111 [mA/min]	[0 0.00216]
Conditioning stimulus × Time		
Before	0.038 [mA/min]	[0.0231 0.0529]
During	0.0678 [mA/min]	[0.0155 0.120]



A



B

Fig. 6. Human experiment results: mean and standard deviations of threshold estimates over time. (A) Estimated thresholds during the static test. (B) Estimated thresholds during the dynamic test. A conditioning stimulus is applied between the fifth and eighth minutes.

5 DISCUSSION AND CONCLUSIONS

We compared the performance of stimulus selection procedures and threshold estimation methods for the use in threshold tracking experiments. Bias and precision of threshold estimates (using either logistic regression or Bayesian estimations) of the simple staircase, random staircase and minimum entropy procedures were compared by means of Monte Carlo simulations. Consequently human subject study was carried out to compare the simple staircase and random staircase procedures.

5.1 Monte Carlo simulations

We performed two types of simulations: the first simulation a stationary threshold, the second a nonstationary threshold. Various procedure specific settings (e.g. step-size) were used in the simulations.

When the threshold is stationary, we observed a negligible bias in threshold estimations in most cases. Non-negligible biases were observed when the threshold changed (Fig. 3). However, the convergence time to negligible bias was similar in all simulations. Obviously, the convergence time of the bias is dependent on the size of the moving time window. Therefore, we expect that smaller windows will result in faster convergence and larger windows in slower convergence. However, as can been seen from Fig. 2, the cost of smaller windows is a loss in estimation precision.

When comparing the estimation precision between the stimulus selection procedures, we observed more precise estimates with the simple staircase procedure than with the other procedures. During stationary thresholds we found that estimation precision of both the simple staircase procedure and minimum entropy is more sensitive to various procedure specific settings than the random staircase procedure. In addition, during nonstationary thresholds, the minimum entropy procedure showed higher sensitivity to various settings than both the simple staircase procedure and random staircase procedure. However, small step-sizes in the simple staircase procedure might result in a high loss in estimation precision. From this, we deduce that estimation precision can become very low when choosing inappropriate procedure settings prior to a human subject experiment. Therefore, we recommend to make a tradeoff between potential precision and low sensitivity to various procedure settings prior to experiments.

Regarding the estimation method, we observed that, in similar cases, logistic regression results in more precise estimates of stationary thresholds than Bayesian estimation (Fig. 2). Especially when the true slope β is placed near the edge of the PD (i.e. PD #1 and PD #2, (Table 1), precision is highly reduced for both the simple staircase procedure and the minimum entropy procedure, but to a lesser extent for the random staircase procedure. This implies that the Bayesian estimation precision depends more on *a priori* choices than logistic regression and can be problematic when little information is available about the psychometric function parameters prior to a human subject experiment. Increasing the size of the PD might be a solution to this issue, but results in increased computational complexity (Kontsevich and Tyler 1999; Kujala and Lukka 2006; Kingdom and Prins 2009). Therefore, logistic regression as estimation method might be a more appropriate solution to this issue, not only because of higher precision, but also because it needs fewer *a priori* assumptions.

5.2 Human subject study

We compared the simple staircase procedure with the random staircase procedure in a human subject study on thresholds. Both procedures used logistic regression as estimation method. Two tests were carried out, the first not containing a conditioning stimulus, the second including a three minute cold pressor task as conditioning stimulus. Both stimulus selection procedures were randomly, but balanced, alternated after every perceived stimulus.

In human subject experiments, it is impossible to find the true threshold and therefore a bias could not be determined. Moreover, calculation of precision is difficult due to between subject variance effects and possible habituation effects. However, the used LMM in this study took the random effects between subjects and effects of time into account. Therefore, if threshold estimations within subjects are similar, the output residuals of the fitted LMM are an indication of estimation precision. Larger residuals imply a lower estimation precision and lower residuals imply a higher estimation precision.

We did not find differences in estimated thresholds between the simple staircase procedure and random staircase procedure. Moreover, no interaction effect between time and stimulus selection procedure was found implying that estimated thresholds were similar in both procedures. We observed that the fitted LMM resulted in lower residuals for the random staircase procedure than for the simple staircase procedure. Therefore, we observed higher precision with the random staircase procedure than the simple staircase procedure.

This observation is contrary to the finding of the Monte Carlo simulations in which estimations were more precise with the simple staircase procedure than with the random staircase procedure in most cases. However, we also observed in simulations that precision of the simple staircase procedure is more sensitive to various procedure settings than the random staircase procedure. It might be that the used procedure settings in the human subject experiments were not optimally chosen and thus resulting in lower estimation precision. Moreover, the temporary higher estimated threshold with the simple staircase procedure between the fourth and seventh minutes in the static test could have caused higher residuals than with the random staircase procedure. However, this temporary increase are due to a few subjects and was not found with the random staircase procedure. Therefore, we account the higher residuals to possible simple staircase procedure effects.

We found habituation effects over time in both the static experiment and dynamic experiment. Thresholds increased further during the conditioning stimulus. As can be seen from Fig. 6B, thresholds increase immediately after the start of the conditioning stimulus. It takes about one to two minutes before reaching the highest threshold. Van Wijk and Veldhuijzen (2010) reported that most studies found a prolonged effect of the conditioning stimulus to be lower than five minutes and a few up to eight minutes. We observed that after removal of the conditioning stimulus, the threshold started to decrease after about a minute and conditioning stimulus effects were still visible until about two minutes after removal. From the simulation results, we observed that threshold estimates immediately responded to a change in true threshold and thus showed no increasing delays. Therefore, the prolonged effect after removal is not due to the stimulus selection procedures, but due to the conditioning stimulus.

5.3 Conclusion

The Monte Carlo simulations and human subject study demonstrated the possibility of tracking changes in threshold over time. We found that logistic regression results in more precise threshold estimates than Bayesian estimation. The simple staircase procedure showed the highest threshold estimation precision. Lowest sensitivity to procedure settings was observed in the random staircase procedure. The minimum entropy procedure showed lower precision than the simple staircase procedure and is more sensitive to procedure settings than the other procedures. The random staircase procedure was found to have higher estimation precision than the simple staircase procedure in the human subject experiments.

Therefore, we recommend the use of the random staircase procedure, in combination with logistic regression for nonstationary nociceptive threshold tracking experiments.

6 ACKNOWLEDGEMENT

This research is supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organisation for Scientific Research (NWO) and partly funded by the Ministry of Economic Affairs, Agriculture and Innovation. The authors would like to thank Victor Sluiter and Frodo Muijzer for their technical contribution.

7 REFERENCES

Cornsweet, T. N. (1962). "The Staircase-Method in Psychophysics." The American Journal of Psychology **75**(3): 485-491.

Ehrenstein, W. H. and A. Ehrenstein (1999). "Psychophysical methods." Modern Techniques in Neuroscience Research: 1211-1241.

Gescheider, G. A. (1985). Psychophysics: method, theory, and application, L. Erlbaum Associates.

Hosmer, D. W. and S. Lemeshow (2000). Applied Logistic Regression, Wiley.

Inui, K. and R. Kakigi (2012). "Pain perception in humans: use of intraepidermal electrical stimulation." J Neurol Neurosurg Psychiatry **83**(5): 551-556.

King-Smith, P. E. and D. Rose (1997). "Principles of an adaptive method for measuring the slope of the psychometric function." Vision Res **37**(12): 1595-1604.

Kingdom, F. A. A. and N. Prins (2009). Psychophysics: A Practical Introduction, Elsevier Science & Technology.

Klein, S. A. (2001). "Measuring, estimating, and understanding the psychometric function: a commentary." Percept Psychophys **63**(8): 1421-1455.

Kontsevich, L. L. and C. W. Tyler (1999). "Bayesian adaptive estimation of psychometric slope and threshold." Vision Research **39**(16): 2729-2737.

Kujala, J. V. and T. J. Lukka (2006). "Bayesian adaptive estimation: The next dimension." Journal of Mathematical Psychology **50**(4): 369-389.

Leek, M. R. (2001). "Adaptive procedures in psychophysical research." Percept Psychophys **63**(8): 1279-1292.

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics." Journal of the Acoustical Society of America **49**(2).

Mitchell, L. A., R. A. MacDonald, et al. (2004). "Temperature and the cold pressor test." J Pain **5**(4): 233-237.

Mouraux, A., G. D. Iannetti, et al. (2010). "Low intensity intra-epidermal electrical stimulation can activate Aδ-nociceptors selectively." Pain **150**(1): 199-207.

Pud, D., Y. Granovsky, et al. (2009). "The methodology of experimentally induced diffuse noxious inhibitory control (DNIC)-like effect in humans." Pain **144**(1-2): 16-19.

Simpson, W. A. (1988). "The method of constant stimuli is efficient." Percept Psychophys **44**(5): 433-436.

Steenbergen, P., J. R. Buitenweg, et al. (2012). "A system for inducing concurrent tactile and nociceptive sensations at the same site using electrocutaneous stimulation." Behav Res Methods.

Talbot, J. D., G. H. Duncan, et al. (1987). "Diffuse noxious inhibitory controls (DNICs): psychophysical evidence in man for intersegmental suppression of noxious heat perception by cold pressor pain." Pain **30**(2): 221-232.

Treutwein, B. (1995). "Adaptive psychophysical procedures." Vision Res **35**(17): 2503-2522.

Treutwein, B. and H. Strasburger (1999). "Fitting the psychometric function." Percept Psychophys **61**(1): 87-106.

Twisk, J. W. R. (2006). Applied Multilevel Analysis: A Practical Guide for Medical Researchers, Cambridge University Press.

Van Wijk, G. and D. S. Veldhuijzen (2010). "Perspective on diffuse noxious inhibitory controls as a model of endogenous pain modulation in clinical pain syndromes." J Pain **11**(5): 408-419.

von Dincklage, F., M. Hackbarth, et al. (2009). "Introduction of a continual RIII reflex threshold tracking algorithm." Brain Res **1260**: 24-29.

Watson, A. B. and A. Fitzhugh (1990). "The method of constant stimuli is inefficient." Percept Psychophys **47**(1): 87-91.

Watson, A. B. and D. G. Pelli (1983). "QUEST: a Bayesian adaptive psychometric method." Percept Psychophys **33**(2): 113-120.

Wilder-Smith, O. H., T. Schreyer, et al. (2010). "Patients with chronic pain after abdominal surgery show less preoperative endogenous pain inhibition and more postoperative hyperalgesia: a pilot study." J Pain Palliat Care Pharmacother **24**(2): 119-128.

Yarnitsky, D., Y. Crispel, et al. (2008). "Prediction of chronic post-operative pain: pre-operative DNIC testing identifies patients at risk." Pain **138**(1): 22-28.

Yarnitsky, D., M. Granot, et al. (2012). "Conditioned pain modulation predicts duloxetine efficacy in painful diabetic neuropathy." Pain **153**(6): 1193-1198.