Comparative simulations of adaptive psychometric procedures

Stefanie Otto, Stefan Weinzierl

Technische Universität Berlin, Fachbereich Audiokommunikation, Germany

Email: stefanie.otto@mailbox.tu-berlin.de, stefan.weinzierl@tu-berlin.de

Introduction

Various psychometric procedures have been proposed to measure perceptual thresholds as points on a subject's psychometric function. Adaptive procedures are generally preferred due to their precision and efficiency. However, different approaches exist with respect to the stimulus adaptation rule and the final estimation of the resulting threshold. In the presented study, the performance of maximum likelihood methods (Best PEST and ZEST) is compared to traditional adaptive methods (Staircase und PEST). All methods are examined with different paradigms of stimulus presentation (Yes/No, 2AFC, 3AFC). The procedures are evaluated in computer simulations with regard to their performance over a large sample size in the case of known truth. Bias, accuracy and efficiency of each method are analysed. The obtained results provide criteria for selecting an appropriate method for a particular application.

Methods

The following adaptive methods, which are are widely used in psychophysic tests, were selected for evaluation:

- Transformed Up-Down Staircase (Levitt, 1971)

- PEST (Taylor and Creelman, 1967)

- Best PEST (Pentland, 1980)

- ZEST (King-Smith et al., 1994)

The chosen methods differ in various aspects: The use of a priori information about the assumed threshold, the stimulus selection rule, and the calculation of the final threshold estimation.

The **Transformed Up-Down Staircase method and PEST** do not assume specific information about the psychometric function. The **Transformed Up-Down** method proposed by Levitt [5] is more flexible than the simple staircase, because it is possible to converge to different target probabilities. The final threshold estimate is obtained by averaging the turnaround points leaving out the first one.

PEST (Parameter Estimation by Sequential Testing) is adaptive in step size and based on blocks of trials. The Wald test, a simplified version of sequential probability ratio test, is used to decide whether to change the stimulus level or not. Then a set of heuristic rules is employed to chose the next step size. On every reversal of the response category it is halvened. After one run the last stimulus level is taken as the most probable threshold [3].

Best PEST and ZEST are parametric methods that use all available information to place the next stimulus presentation as close as possible to the true, but unknown, threshold and to maintain an accurate estimate with the least possible number of trials. This means that the experimenter has to decide in advance about the general form of the psychometric function, e.g. the cumulative normal, the Weibull or the logistic distribution. As the threshold value was the only free parameter to be estimated, the slope parameter was set to a fixed value. Thus the different possibilities for the psychometric functions are reduced to translations of one template parallel to the x-axis.

Best PEST computes implicit trials before starting the measurement. For the Yes-No paradigm the first implicit trial was presented at the lowest possible value and the answer set as incorrect. The second implicit trial is placed at the highest stimulus level and the answer set correct. In this way prior information about the location of the threshold is included in the subsequent calculations. Maximum likelihood estimation is employed to select the next stimulus level. The maximum (mode) of the likelihood function is supposed to be the best current estimate. The last tested value is the final threshold estimate [2].

ZEST (Zippy Estimation by Sequential Testing) starts with a prior probability density function (p.d.f.) that has to be provided before testing. For non-informative prior the rectangular or uniform distribution can be used. A Bayesian approach is employed to estimate the most probable threshold value. ZEST uses the mean of the posterior probability density function to chose the next stimulus level. Bayesian estimation with the mode (maximum) of the posterior pdf as estimator and a rectangular prior distribution would be equivalent to maximum likelihood estimation [1],[6].

Previous studies

A couple of comparative simulations have been done, mainly by authors suggesting a new method or an improvement of existing procedures. Pentland [2] compared the simple staircase method, PEST, improved PEST and Best PEST in computer simulations but employs only the yes-no paradigm. The results show that PEST versions and Best PEST differ only little regarding the accuracy. Wheras Pentland claimed that Best PEST provides unbiased and accurate treshold estimates. However, other authors have shown that Best PEST produces biased results when combined with AFC paradigms [6],[8].

Madigan and Williams [7] compare the accuracy of PEST, Best PEST and QUEST in a Yes-No and a 2AFC task. The results show that PEST is less efficient. Best PEST and

QUEST show similar accuracy. All tested methods proved more accurate estimates when combined with the Yes-No paradigm.

King-Smith et al. [6] were able to show by simulations that the mean of the posterior p.d.f. yielded more reliable and less biased estimates than the mode.

The present study wants to combine all this previous work in simulations under the same conditions. All four types of methods are combined with the Yes-No, 2AFC and 3AFC paradigm producing 12 different settings. Special interest regards the performance of the parametric methods in a 3AFC task, because this is the most frequently used combination. The performances are evaluated not only regarding the accuracy but also the bias.

Simulations

The stimulus set available was defined in logit units ranging from -5 to +5 logit units in steps of 0.25. The real psychometric function was chosen to be the logistic function with a mean threshold at 0 and a slope of 1 logit unit. The subject's true threshold was randomly distributed within -3 and +3 logit units. The real lapsing rate (the rate of false negative errors above threshold) was set to 2%.

Previous simulations have been computed to choose several parameters for each method. The staircase method was implemented with an initial step size of 2 logit units which is reduced to 1 logit unit after the first turnaround. For the PEST method previous tests showed that an initial step size of 4 logit units allows for a sufficient number of reversals during one run. The test stops when the step size arrives at 0.25 logit units, which is the minimal resolution of the stimulus set. For Best PEST and ZEST the parameters of the assumed psychometric function (also chosen to be of the logistic form) were set to correspond to the real psychometric function. The prior p.d.f. for ZEST is a flat Gaussian distribution, which is similar to the Likelihood function in Best PEST after two implicit trials. One run was terminated after a fixed number of simulated judgements, ranging from 10 to 50 trials. One thousand subjects were simulated for each condition investigated.

Results

Yes-No paradigm

The bias of the final results is simply the mean of the threshold estimation error. It is shown in Figure 1 for all four methods combined with the Yes-No task. Only PEST shows positive bias. All other three methods produce unbiased estimates.

Figure 2 shows the variance of threshold estimation errors for the methods combined with the Yes-No paradigm. Almost all methods converge very fast and provide accurate results. Only PEST needs more than 50 trials to converge at the 50% correct point.
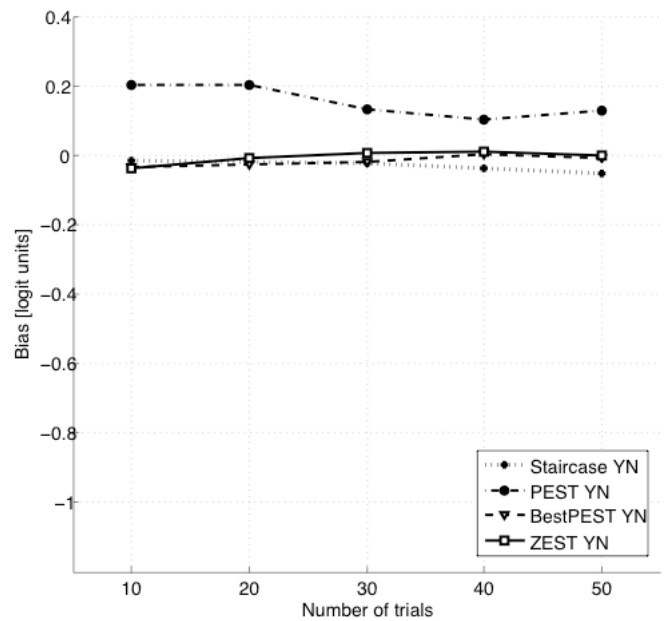

Figure 1: Bias of threshold estimates for all four methods combined with Yes-No paradigm and different number of trials.
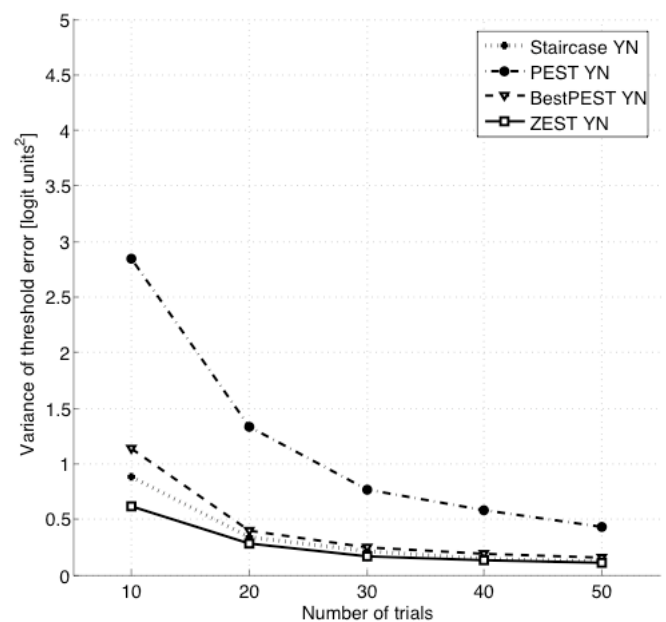

Figure 2: Variance of threshold error for all four methods and different number of trials in a Yes-No task.

2AFC paradigm

The bias of the final results is shown in Figure 3 for all four methods combined with the 2AFC paradigm. The staircase method shows constant positive bias. This is due to the selected 3Down-1Up rule, because in a 2AFC task the procedure converges at 61.2% correct. Best PEST and PEST show negative bias, which is remarkably high if only a small number of trials is taken. Only ZEST provides constantly bias-free results.

Figure 4 shows the variance of threshold estimation errors for the methods combined with the 3AFC paradigm. ZEST provides best results, already after few trials. PEST and Best

PEST show large variance especially when only few trials are computed. These methods tend to require more measurements to provide accurate threshold estimates.
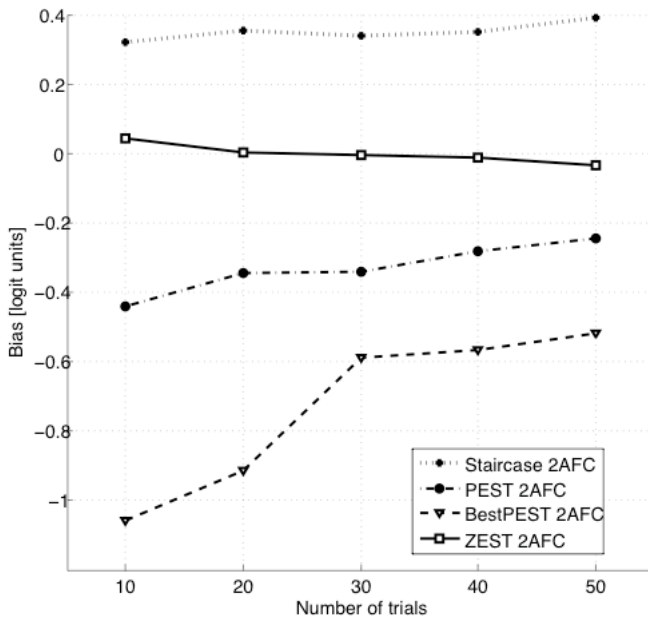

Figure 3: Bias of threshold estimates for all four methods combined with 2AFC paradigm and different number of trials.
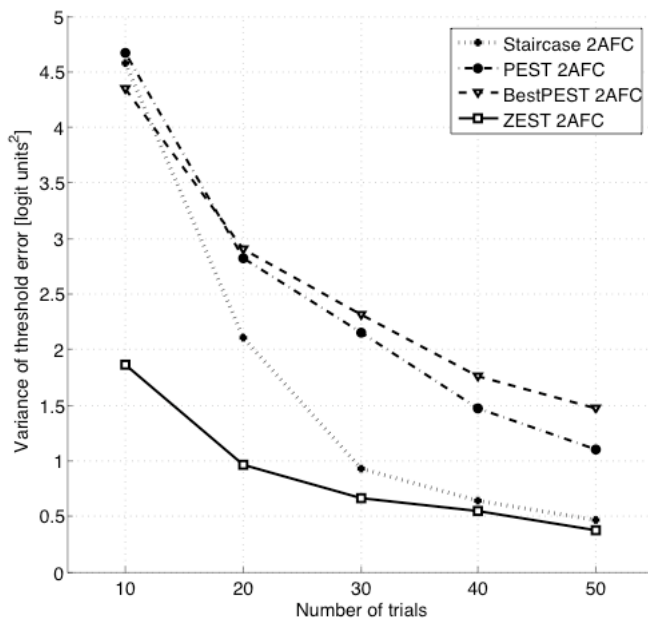

Figure 4: Variance of threshold error for all four methods combined with 2AFC and different number of trials.

3AFC paradigm

The bias in a 3AFC task is shown in Figure 5 for all four methods. The staircase method shows little, but constant positive bias, due to the selected 2Down-1Up-rule. In a 3AFC task this procedure converges at 57.8% correct instead of 50%. Best PEST and PEST tend to have negative bias,

which reduces with the increasing number of trials. Only ZEST shows constantly bias-free results.

Figure 6 shows the variance of threshold estimation errors for 3AFC task combined with all four methods. As with 2AFC, PEST and Best PEST need more measurements to provide accurate threshold estimates, while ZEST provides the most accurate results.
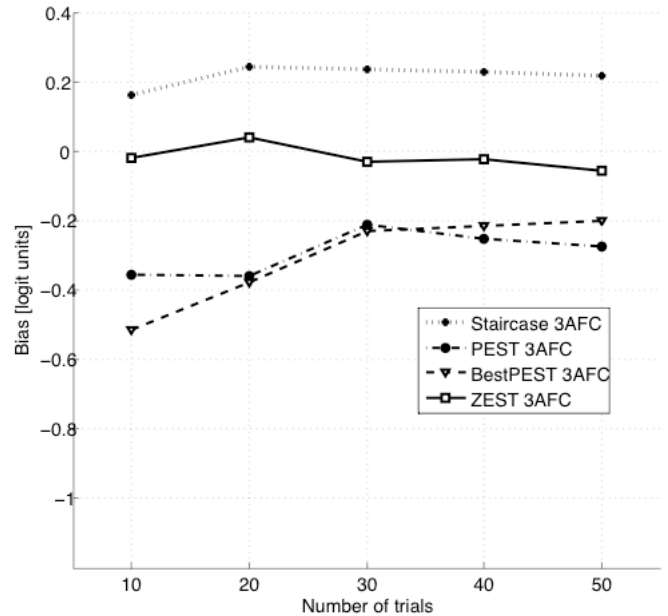

Figure 5: Bias of threshold estimates for all four methods combined with 3AFC and different number of trials.
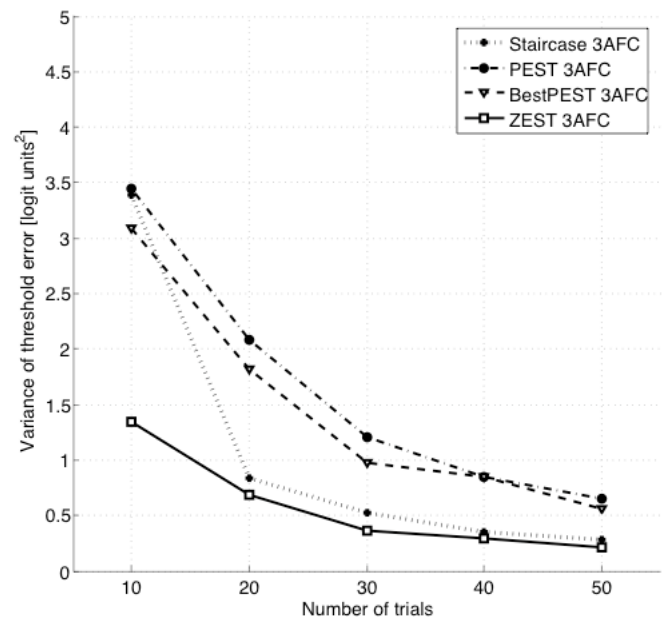

Figure 6: Variance of threshold error for all four methods combined with 3AFC and different number of trials.

Discussion

The Staircase method provides reliable results, when a certain amount of trails has been computed. In combination with an AFC task, it returns a predictable and thus correctable bias due to the different convergence point. The

PEST method shows biased results and requires a high number of measurements to provide accurate estimates. Best PEST did not prove to be superior to the other methods (as claimed by Pentland [2]). If used in combination with AFC paradigms, it provides biased and inaccurate results. ZEST proved to be superior in all tested conditions. Hence, our simulations confirmed earlier studies [6] showing that especially for asymmetrical p.d.f.s the mean of the posterior p.d.f. (as used in ZEST) is a more reliable estimator for placing the next stimulus than the mode, as used in the original Best PEST procedure. According to our simulations, a combination of 3AFC and ZEST is particularly recommended to yield fast, unbiased and accurate threshold estimates.

References

[1] Treutwein, B.: Adaptive Psychophysical Procedures. Vision Research **35** (1995), 2503-2522

[2] Pentland, A.: Maximum Likelihood Estimation: the best PEST. Percept. Psychophysics **28** (1980), 377-379

[3] Taylor, M.M. and Creelman, C.D.: PEST: Efficient Estimates on probability functions. J. Acoust. Soc. Am. **41** (1967), 782-787

[4] Watson, A.B. and Pelli, D.P.: QUEST: A Bayesian adaptive psychometric method. Percept. Psychophys. **33** (1983), 113-120

[5] Levitt, H.: Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. **49** (1971), 467-477

[6] King-Smith et al.: Efficient and Unbiased Modifications of the QUEST Threshold Method: Theory, Simulations, Experimental Evaluation and Practical Implementation. Vision Res. **34** (1994), 885-912

[7] Madigan, R. and Williams, D.: Maximum-likelihood psychometric procedures in 2-alternative forced-choice: Evaluation and recommendations. Percept. Psychophys. **42** (1987), 240-249

[8] Emerson, P.L.: Observations on maximum-likelihood and Bayesian methods of forced-choice sequential threshold estimation. Percept. Psychophys. 39 (1986a), 151-153