



Absolute-detection results with a Bayesian adaptive tracking procedure using simulated and human subjects.

Georg Stiefenhofer, William S. Woods, Brent Edwards, Starkey Hearing Research Center, 2150 Shattuck Ave., Berkeley, CA
 georg.stiefenhofer@stud.fh-oldenburg.de, bill_woods@starkey.com, brent_edwards@starkey.com



Motivation

Adaptive tracking procedures are widely used in psychoacoustics. The common, fixed step-size, up-down procedures are: known to be biased and yield convergence dependent upon the step-size used [5]; provide an estimate of only one parameter of the psychometric function; and in general make nonoptimal use of information (*a priori* and *a posteriori*) concerning a given subject's psychometric function. Recently introduced Bayesian adaptive procedures [1] make optimal use of all information in both the estimation of parameters of the psychometric function, as well as in stimulus placement via a minimum-entropy rule. No direct comparison of the up-down and Bayes procedures has been carried out, however.

Our main goal was to compare the Bayesian and up-down procedures using simulations, and examine the Bayesian procedure's performance with real subjects.

Algorithm

The algorithm introduced by Kontsevich and Tyler [1] uses Bayesian learning to modify an a priori probability distribution of parameters related to the sensitivity being measured. Before any measurements are made the prior distribution is built from existing information. The Bayesian representation is used to determine the stimulus presentation levels that most quickly minimize the entropy (or "uncertainty") as the procedure advances.

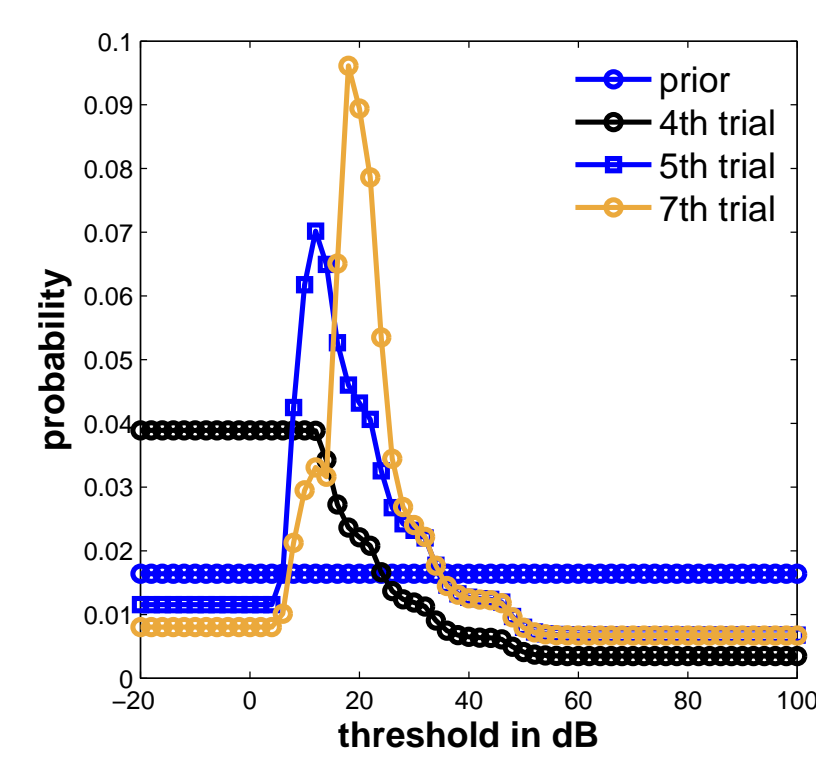


Figure 1. Initial distribution of the uniform prior and posterior distribution after the 4th, 5th and 7th trial. Until the 5th trial the routine lowers the probability at the high threshold values. It is learning that the true threshold is most likely not at these values. After the 7th trial the routine has gathered enough data to be quite sure that the actual threshold is somewhere around 20dB.

The sensitivity of the subject is determined from the final ("posterior") probability distribution of the parameters of sensitivity:

$$p_t(\theta|r_x) = \frac{p_t(\theta)p(r_x|\theta)}{p_t(r_x)}$$

where $p_t(\theta)$ is the prior distribution and $p_t(r_x)$ is a normalizing term.

The procedure creates a hypothesis space of psychometric functions, where $p(r_x|\theta)$ is the probability of a correct ($r_x = 1$) or incorrect ($r_x = 0$) response at an intensity x given the parameters of sensitivity θ . An initial prior distribution $p_0(\theta)$ from existing information is built and $p(r_x|\theta)$ is precomputed.

Next a loop for all trials $t = 0 \dots T - 1$ is run:

1. Compute for all x and $r_x \in \{0, 1\}$

$$p_t(r_x) = \sum_{\theta} p_t(\theta)p(r_x|\theta)$$

2. Compute the posterior distribution for all θ , x and $r_x \in \{0, 1\}$.

3. For all x and $r_x \in \{0, 1\}$ compute the posterior entropy:

$$H_t(\theta|r_x) = - \sum_{\theta} p_t(\theta|r_x) \log p_t(\theta|r_x).$$

4. For all x compute the expected posterior entropy:

$$H_t(\theta|R_x) = - \sum_{r_x \in \{0,1\}} p_t(r_x) H_t(\theta|r_x).$$

5. A trial is run at an intensity that minimizes $H_t(\theta|R_x)$.

6. The result is used to update the prior information for the next trial:

$$p_{t+1}(\theta) = p_t(\theta|r_x^t).$$

Algorithm loops and trials presented until stopping criteria (e.g., entropy threshold or max number of trials) are met.

Simulation Methods

1000 simulations of 150-trial, typical Monte Carlo runs (cf. [6]) were simulated with:

1. A common 3-down 1-up staircase procedure with start step size of 10dB changing to 2dB after the 4th reversal. Threshold by definition is the 79% correct point and was obtained by averaging an even numbers of reversals from the 5th on.

2. The Bayesian procedure.

Psychometric functions were sampled every 1dB in stimulus intensity (to be consistent with detection experiment - see below) and intensity ranged from -20dB to 100dB. Function parameters θ were horizontal shift θ_{subj} , slope, and lapsing rate λ_{subj} . The shift parameter ranged from -20dB to 100dB with an increment of 2dB. The slope parameter was fixed with a value of 4 or ranged from 2 to 10.

The psychometric functions were implemented based on a power law approximation to the d' function with $d'(x) = (\frac{x}{a})^b$, where x is stimulus intensity, a shift parameter and b slope. Psychometric functions were cumulative Gaussians ranging from guessing probability to $1 - \lambda$:

$$P_c(d') = \lambda + (1 - 2\lambda) \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d'} e^{-\frac{x^2}{2\sigma^2}} dx.$$

thresholds θ_{subj}	lapsing rates λ_{subj}	slope space	prior distr.
0 dB	0.01%	fixed	gaussian
40 dB	1%	free	uniform
70 dB	10%	free	uniform

Table 1. Different conditions and parameters for the subject's function simulated.

To allow the staircase procedure to make use of prior information initial stimulus level was set appropriately.

$$L_{init} = \sum_{\Theta} L(79\%|\Theta) \cdot P(\Theta)$$

where $P(\Theta)$ is the prior and $L(79\%|\Theta)$ are all levels of the assumed hypotheses Θ that lead to 79% correct.

Simulation Results

1. Bayesian converges faster to true threshold than the staircase procedure (figure 2).

2. Staircase procedure provides estimates close to Bayesian especially if starting level is close to true threshold. This is always the case for Gaussian priors and coincidentally for some conditions with uniform priors (figure 2 upper right panel).

3. Bayesian estimates reach asymptotic values with a deviation of ± 1 dB from trial 35 on, whereas staircase estimates need about 10 trials more and are further more biased.

4. Bayesian is more biased than staircase for very early trials. This is due to leaving out the first 4 reversals of the up-down procedure, hence results for staircase procedure are available from trial 15-20 on.

5. A priori information improves results, especially for trials < 50 and for a high lapsing rate.

6. Impact of a free slope parameter on Bayesian thresholds is almost not pronounced (figure 4).

7. Slope estimates are effected from trial 50 on and are more biased (overestimated) than threshold estimates (figure 5).

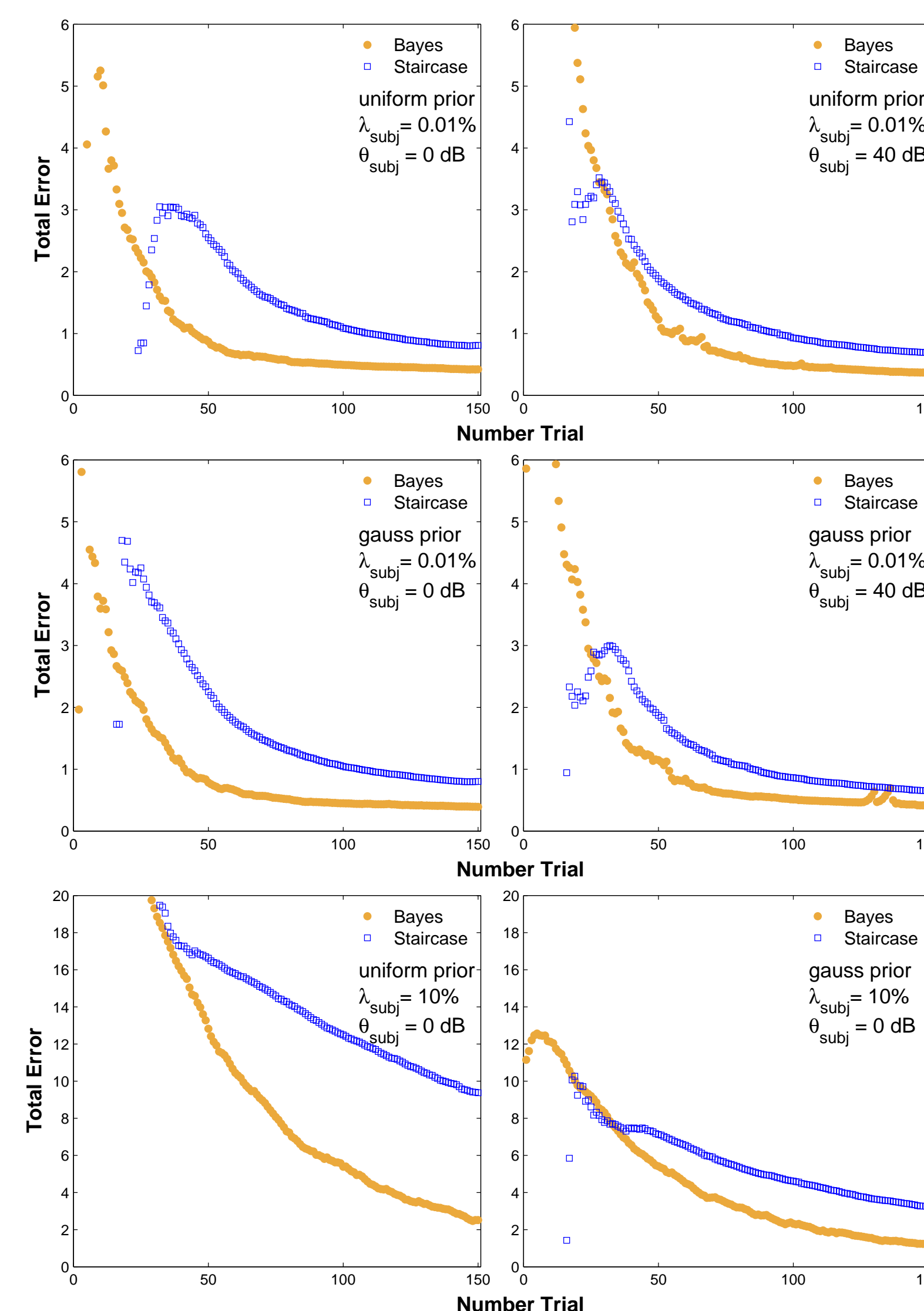


Figure 2. Total error for 79% threshold estimation in different simulation conditions. $TotalError(i) = \sqrt{\frac{1}{N} \sum_{n=1}^N (\hat{\theta}_{79\%}(i) - \theta_{79\%})^2}$, where $\hat{\theta}_{79\%}(i)$ is the estimated threshold in dB at trial i , $\theta_{79\%}$ true threshold and N number of simulations.

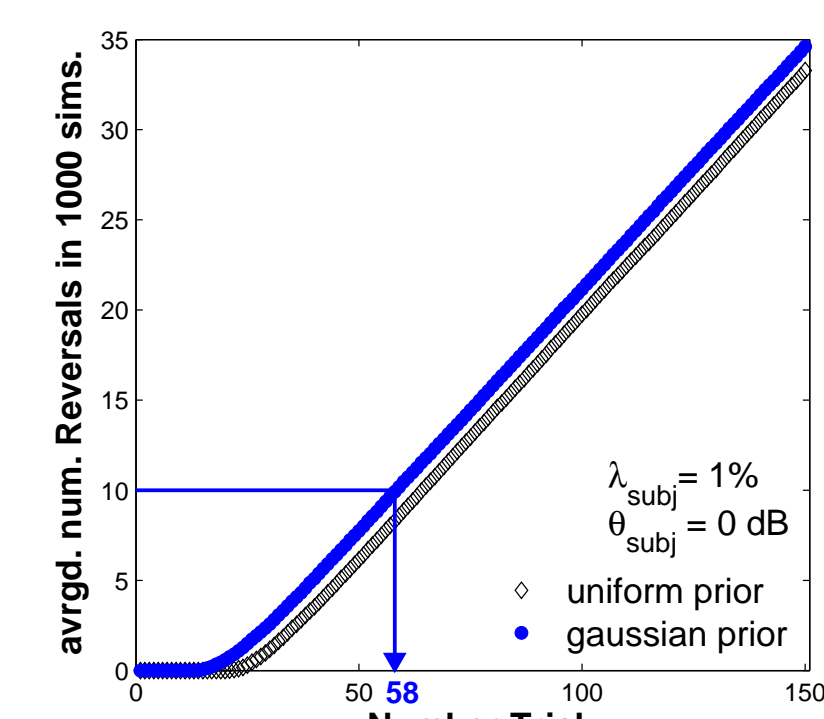


Figure 4. Number of reversals in 1000 1-up-3-down 2-AFC simulations. The first 4 reversals were not counted. With Gaussian priors the initial stimulus level is closer to true threshold than with uniform, hence the procedure gives more reversals. With a Bayesian approach stimulus selection rule does not change over time. Arrow indicate trials needed for 10(+4) reversals.

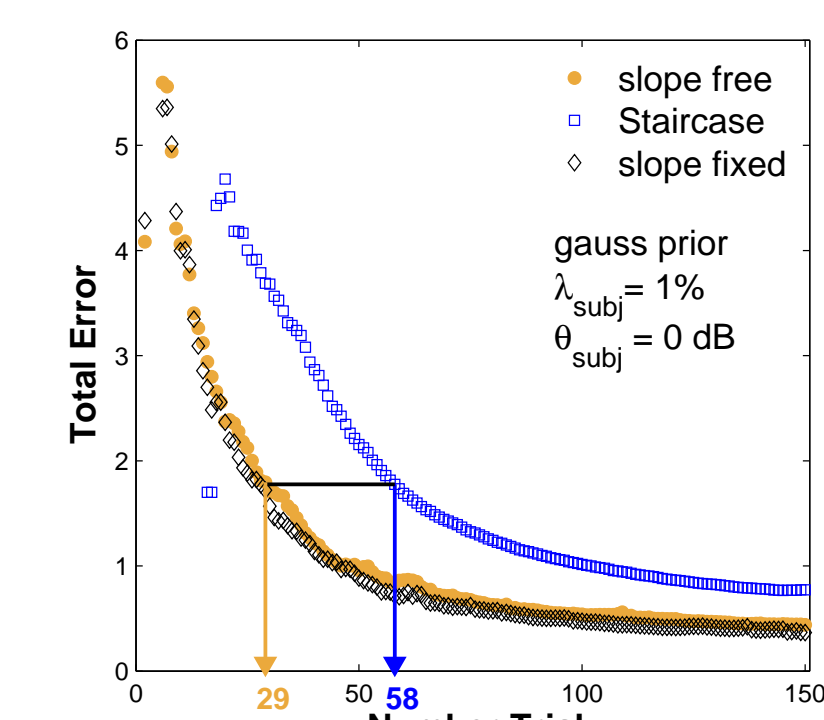


Figure 3. No pronounced difference in total error between Bayesian procedure with fixed and free slope parameter in the hypothesis space. Bayesian procedure has better threshold estimates and in addition provides an estimate of the slope parameter. Lines indicate total error for 10(+4) reversals with staircase and trials needed for the same error with the Bayesian procedure.

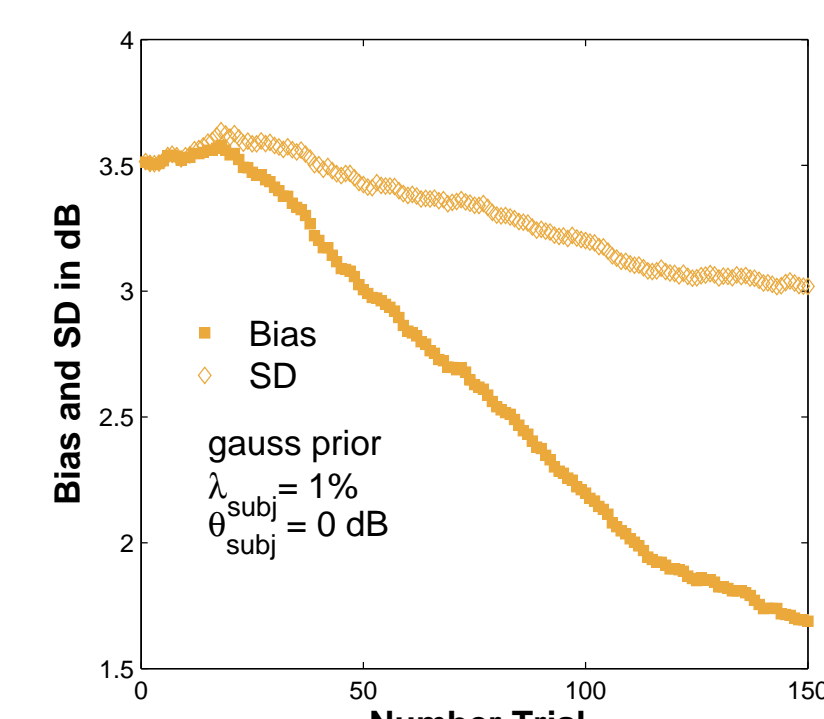


Figure 5. Bias and SD of slope estimation, where $bias(b)_t = \frac{\sum_{i=1}^N (\log \hat{b}_i - \log b)}{N} \cdot 20dB$ and $SD(b)_t = \sqrt{\frac{\sum_{i=1}^N (\log \hat{b}_i - \log b)^2}{N-1}} \cdot 20dB$ and \hat{b}_t is the estimated slope at trial t and b is true slope.

Experimental Methods

1. Standard audiograms (2dB and 5dB).
2. Six blocks with 30 trials each of the 2-AFC Bayesian procedure.

3. Three blocks with 30 trials each of a 2-AFC fixed level procedure. The levels were the averaged Bayesian thresholds and were adjusted ± 1 -2dB depending on subjects' performance.

Psychometric functions were fitted to the fixed level data minimizing the squared error:

$$Error^2 = \frac{(P_{correct}^{fit}(x) - P_{correct}^{actual}(x))^2}{\sigma_{P_{correct}}^2},$$

where $P_{correct}(x)$ is the probability correct at a given stimulus level for the *fitted* psychometric function and the *actual* probability correct in the fixed level procedure. $\sigma_{P_{correct}}^2$ is the binomial variance for the *fitted* probability correct.

Measurements were conducted monaural (right ear) on 14 paid subjects with different hearing loss. Frequencies measured were 4kHz, 2kHz, 1kHz and 500Hz in that order. The measuring equipment were calibrated to the audiometer.

A priori information for the shift parameter was obtained from a database [7] which provided mean and standard deviation of hearing loss. The first frequency (4kHz) was determined with uniform prior distribution, but with 10 more trials. Second and subsequent priors were built by finding the threshold value in the database which was closest to the actual threshold and using mean and S.D. of the following frequency (table 2). Prior distribution for the slope parameter was always uniform.

4kHz	6	9	13	17	28	39	54	60	70	80	90
2kHz	6	9	11	11	18	28	41	50	60	70	80
1kHz	5	8	9	11	16	23	35	45	55	65	75
500Hz	9	12	12	9	13	20	30	40	50	60	70

Table 2. Prior database mean hearing loss for women, e.g. having determined a threshold of 72dBHL at 4kHz, 70dBHL is the closest value, hence mean of the prior distribution for 2kHz would be 60 dBHL. Same procedure for S.D. All values in dBHL.

Experimental Results

1. Bayes thresholds are up to 10dB lower than 2dB-step audiogram thresholds for all subjects and frequencies, except for 2 subjects at 500Hz. Averaged across subjects Bayes thresholds are 4 to 7dB lower than 2dB-step audiogram thresholds.

2. Standard deviation for Bayes estimates averaged across subjects is ≤ 2 dB for all frequencies.

3. Slope estimates averaged across subjects is about 6.

4. Correlation between thresholds on the psychometric function fitted to the fixed level data and Bayesian estimates is excellent (figure 7).

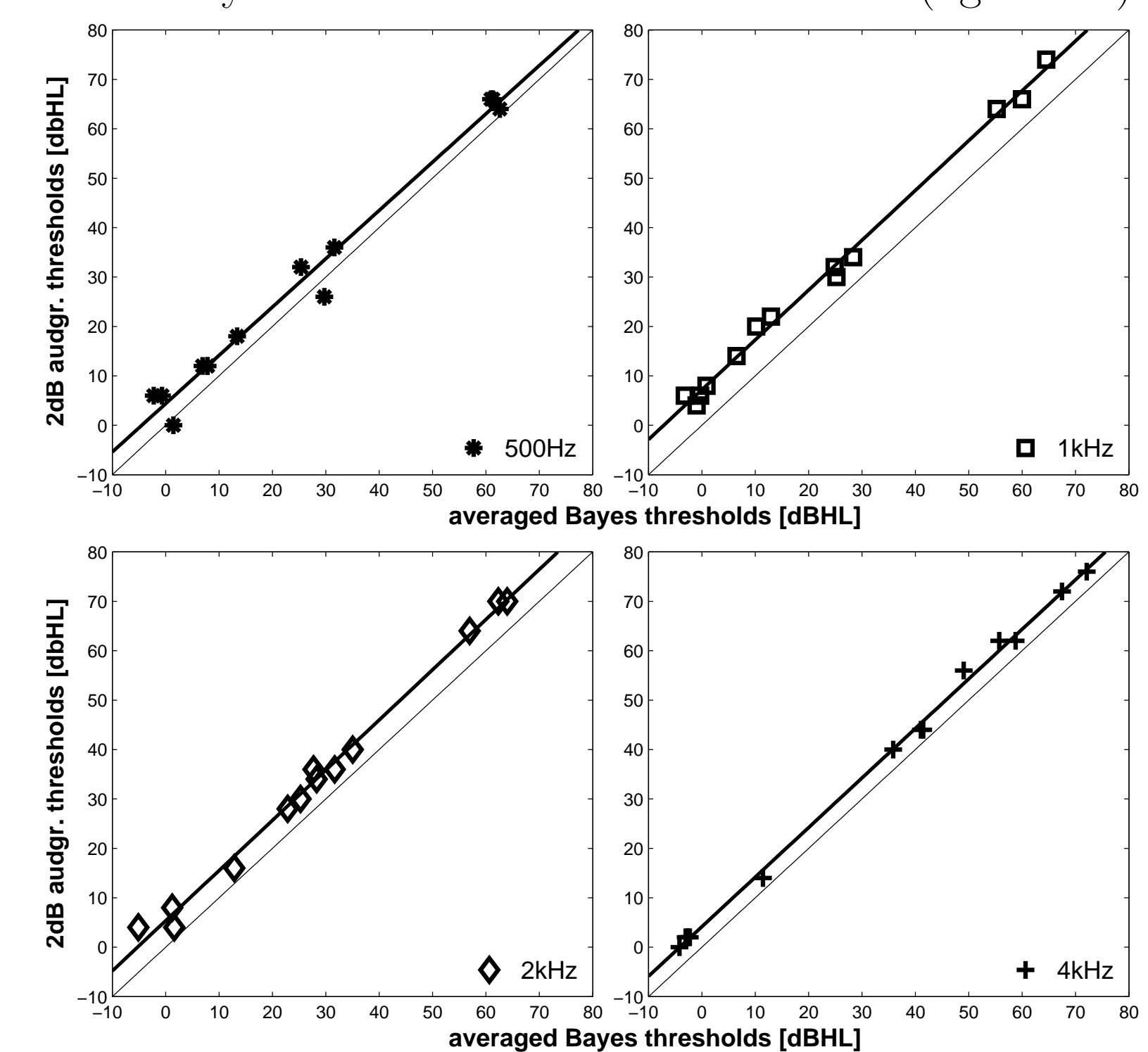


Figure 6. Averaged Bayes and 2dB audiogram data for all subjects. Solid lines are linear best-fits with slopes of 1 and offsets of 4.2, 5.4, 7.2 and 4.4dB for 4kHz, 2kHz, 1kHz and 500Hz. Thin lines indicate equality of Bayes and audiogram thresholds.

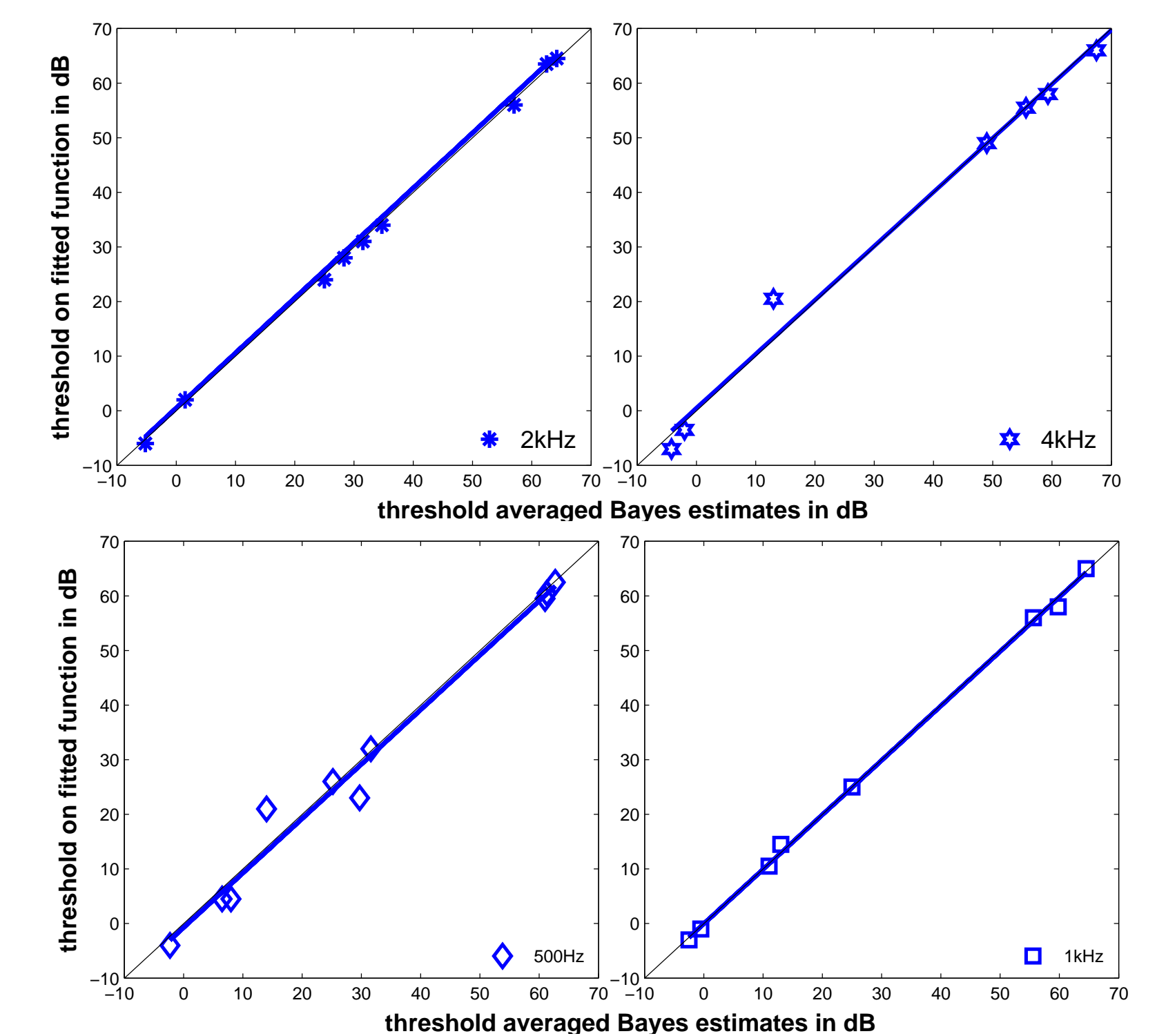


Figure 7. Bayesian data averaged across all runs and fitted psychometric function threshold for all subjects. Lines are best-fit linear regression to the data and have slopes of 1 and offsets < 0.75 dB.

Discussion

1. More sophisticated derived a priori information would further improve Bayesian performance, e.g. several databases for different types hearing losses especially if more than 4 frequencies are to be measured.

2. The literature gives different values for the slope of psychometric functions for absolute threshold detection. Hanna et al [2] found slopes ranging from 3 to 5 whereas Arehart et al [3] found slopes ranging from 0.5 to 2.5. Watson et al [4] claimed that slopes of psychometric functions are steep at higher frequencies and require between 3 and 8dB increase in signal to raise performance from 60% to 95%. This actually means a slope of up to 6. Contradictory to that the slope parameter in simulations is often assumed to have values between 0.5 and 2. However, our slope estimates are about 6 on average for all frequencies. Simulations showed that slope is overestimated by about 3dB around trial 30 which means by a factor of 1.4. Using this correction slope for absolute detection would be about 4 to 5.

3. The entropy of the posterior distribution could also be used to detect instances when the current patient's performance is inconsistent. Additionally more sophisticated stopping criteria could be designed by analyzing the progress of the entropy trial by trial.

Conclusions

Results from simulations and experiments show that a Bayes procedure coupled with a minimum-entropy stimulus placement rule can lead to very reliable parameter estimates with fewer trials than a standard staircase procedure, even while estimating more parameters. Appropriate use of a "lapsing" parameter allows the Bayes procedure to react to subject lapsing in much the same way as the staircase procedure. Up-down procedures increase stimulus level until a subject stops lapsing, Bayes procedure rather increases level if lapsing is less likely than a threshold at a higher intensity. Future work will consider the use of particle filtering to eliminate dependence on discretized hypothesis space.

References

- [1] L.L. Kontsevich and C.W. Tyler, Vision Research **39**, 2729-2737 (1999)
- [2] T.E. Hanna, S.M. von Gierke, D.M. Green, J. Acoust. Soc. Am. **80**(5), November 1986
- [3] K.H. Arehart, E.M. Burns, R.S. Schlauch, Journal of Speech and Hearing, Vo. **33**, 433-439, 1990
- [4] C.S. Watson, J.R. Franks, D.C. Hood, J. Acoust. Soc. Am. , 1972
- [5] B. W. Edwards and G. H. Wakefield, J. Acoust. Soc. Am. **83**, S17 1988
- [6] R.S. Schlauch, R.M. Rose, J. Acoust. Soc. Am. **88**(2), (1990)
- [7] National Health and Nutrition Examination Survey (NHANES) <http://www.edc.gov/nchs/nhanes.htm>

Table 1:

Table 2:

Fig. 1.—

Fig. 2.—

Fig. 3.—

Fig. 4.—

Fig. 5.—

Fig. 6.—

Fig. 7.—