Atten Percept Psychophys
DOI 10.3758/s13414-013-0438-9

Comparing adaptive procedures for estimating the psychometric function for an auditory gap detection task

Yi Shen

Abstract A subject's sensitivity to a stimulus variation can be studied by estimating the psychometric function. Generally speaking, three parameters of the psychometric function are of interest: the performance threshold, the slope of the function, and the rate at which attention lapses occur. In the present study, three psychophysical procedures were used to estimate the three-parameter psychometric function for an auditory gap detection task. These were an up–down staircase (up–down) procedure, an entropy-based Bayesian (entropy) procedure, and an updated maximum-likelihood (UML) procedure. Data collected from four young, normal-hearing listeners showed that while all three procedures provided similar estimates of the threshold parameter, the up–down procedure performed slightly better in estimating the slope and lapse rate for 200 trials of data collection. When the lapse rate was increased by mixing in random responses for the three adaptive procedures, the larger lapse rate was especially detrimental to the efficiency of the up–down procedure, and the UML procedure provided better estimates of the threshold and slope than did the other two procedures.

Keywords Psychometrics · Testing · Psychoacoustics

Many psychophysical experiments measure people's ability to detect a change in one or more aspects of a physical stimulus. As the magnitude of the change, or the signal strength, increases, the probability of detecting the change typically increases as well. Psychometric functions describe the dependence of detectability on signal strength, making psychometric functions an important tool for the study and modeling of perceptual sensitivity. Typically, a psychometric function takes the form of a sigmoidal function, which can be described using three parameters: (1) the detection threshold—that is, the signal strength at the center of the psychometric function's dynamic range; (2) the slope of the psychometric function; and (3) the lapse rate—that is, the distance between 100 % correct and the function's upper asymptote. The lapse rate is named so because attentional lapses provide an explanation for the observation that the measured psychometric function may not reach 100 % correct responses.

A classic procedure for estimating the psychometric function utilizes the method of constant stimuli. In this procedure, the percent correct is estimated at each of several preselected signal strengths. The psychometric function is obtained by fitting the resulting percent-correct data to a prespecified function. Although it is a widely accepted procedure, the method of constant stimuli can lead to very time-consuming data collection. Consequently, adaptive tracking procedures have been proposed with the goal of more efficient estimation of either threshold, or more generally, psychometric functions. Most of these procedures use information gained from previous trials to determine the stimulus placement on the following trials and can roughly be classified into three categories: (1) up–down staircase adaptive tracking procedures (e.g., Levitt, 1971), (2) min-variance Bayesian procedures (e.g., Green, 1990), and (3) min-entropy Bayesian procedures (e.g., Kontsevich & Tyler, 1999).

Nonparametric staircase procedures are typically used to estimate the signal strength at a predetermined percentage correct (i.e., threshold). Typically, signal strength is reduced following a correct response or increased following an incorrect response. To ensure that the signal strength converges rapidly toward the target percentage correct, many procedures adaptively adjust the step size by which the signal strength is manipulated following each trial (e.g.,

Y. Shen (✉)
Department of Cognitive Sciences, University of California,
Irvine, 3151 Social Sciences Plaza, Irvine, CA 92697-5100, USA
e-mail: shen.yi@uci.edu

Published online: 16 February 2013

PEST procedure; Kaernbach, 1991; Taylor & Creelman, 1967), while others use a fixed step size and modulate the probability of increasing versus decreasing the signal strength after each trial (e.g., Derman, 1957; Durham & Flournoy, 1995). Levitt (1971) described a transformed up–down procedure, in which different rules for the increments and decrements of signal strength yield different target percent-correct convergences. For instance, given a fixed step size, if an increment of signal strength takes place following one incorrect response and a decrement occurs after one correct response (one-down, one-up track), the procedure would place the signal strength at the 50 %-correct point on the psychometric function. On the other hand, if two consecutive correct responses are required for a decrement but a single incorrect response leads to an increment (two-down, one-up track), the signal strength would be placed at the 70.7 % point.

Although staircase procedures were originally proposed to estimate one point on the psychometric function (i.e., the threshold), several studies have been conducted to explore their usefulness in estimating the slope of the psychometric function using maximum-likelihood algorithms. Leek, Hanna, and Marshall (1992) investigated the reliability and accuracy of using transformed up–down procedures for slope estimation and found that accurate psychometric function slope estimates can be achieved using these procedures; however, when the number of experimental trials is small (e.g., less than 100 trials), the slope estimates could be biased, particularly when the slope of the true psychometric function is relatively shallow. In a simulation study, Kaernbach (2001) studied the origin of this type of bias in the slope estimates and found that the staircase procedures that estimate thresholds at a certain percentage correct on the psychometric function would introduce sequential dependency between adjacent trials. However, the maximum-likelihood estimation of the psychometric function requires independent data across trials. As a result, the estimated slopes using maximum-likelihood algorithms are usually steeper than the true psychometric function slopes. The author suggested that to prevent biases in the slope estimates, experimenters could use either interleaved staircase tracks targeting two different points on the psychometric function or Bayesian adaptive procedures that provide interim estimation of the psychometric function slope during the run. Unlike for the slope parameter, no study has systematically addressed the question of whether the lapse rate can be reliably estimated using transformed up–down procedures.

In addition to staircase procedures, Bayesian adaptive procedures have been proposed (e.g., QUEST, Watson & Pelli, 1983; ZEST, King-Smith & Rose, 1997), and, with increasingly fast computers, are gaining popularity. These procedures update the posterior distributions of the psychometric function parameters on a trial-by-trial basis according to Bayes's rule. Closely related to these Bayesian procedures are maximum-likelihood procedures (Green, 1990, 1993) based on the same type of analysis, although rather than posteriors, they use the likelihood functions for parameter estimation.[1] On a given trial, a current best estimate of the psychometric function is obtained on the basis of either posterior distributions or likelihood functions. The signal strength of the following trial is then chosen, according to a predefined sampling strategy, at one of a few specific points on the best-fitting psychometric function. These special locations for stimulus placement, the so-called "sweet points," are derived so as to minimize the expected variances in the parameter. It has been found that for psychometric functions taking the form of a logistic function, there exists one sweet point optimized for the estimation of the threshold parameter (Green, 1990), and two sweet points for the slope parameter (Brand & Kollmeier, 2002; King-Smith & Rose, 1997). In a recent study, Shen and Richards (2012) also showed that the best sampling strategy for estimating the lapse rate is to present stimuli at the upper limit of signal strength. Therefore, several sweet points might coexist, and rules must be implemented to select the appropriate sweet point on each trial. In their ZEST procedure, King-Smith and Rose used an alternating sweet-point selection rule, in which the stimulus was placed alternatingly at one of the two sweet points for the slope parameter.

To enable the inclusion of more than two sweet points for concurrent estimation of multiple parameters of the psychometric function, Shen and Richards (2012) described a sweet-point selection rule based on the transformed up–down procedure, in which the signal strength was shifted down to the next-highest sweet point after *n* consecutive correct responses, and shifted up to the next-lowest sweet point after a single incorrect response. Note that the signal strength might change even though the same sweet point was visited on two different trials, because sweet-point estimates are changed as the estimate of the psychometric function is updated trial by trial. This *n*-down, one-up sweet-point selection rule has the potential advantage of making the experiment easy to follow for naïve listeners, allowing the subjects to maintain performance at a certain percentage correct. Using this sweet-point selection rule, Shen and Richards investigated the efficiency of an updated maximum-likelihood procedure (UML) that utilized four

[1] Note that the maximum-likelihood procedure (e.g., Green, 1990) and the maximum-likelihood algorithm for estimating the psychometric function are two different concepts. The maximum-likelihood algorithm is a computational method to fit the psychometric function to collected data. On the other hand, the maximum-likelihood procedure is an adaptive psychophysical procedure by which the psychometric function parameters are estimated following each experimental trial using the maximum-likelihood algorithm.

sweet points in a simulation study. Their results suggested that extending the stimulus placement to four sweet points improved the estimation of the psychometric function parameters, especially for the slope and lapse rate.

The procedures described above have used sampling strategies that minimize the variances of the parameter estimates. Criteria other than variance minimization have also been used in Bayesian adaptive procedures. Kontsevich and Tyler (1999) described a procedure that determines signal strength by performing a one-step-ahead search that minimized the expected entropy function. For example, after a trial, the parameter posterior distributions might be concentrated into narrow regions, and thus exhibiting low entropies, or widely spread across the parameter space, exhibiting high entropies. Depending on the signal strength on the following trial, the total entropies might be expected to increase or decrease. According to Kontsevich and Tyler, the optimal place to sample is at the signal strength that minimizes overall expected entropy, thereby maximizing the expected information gain. Both computer simulations and psychophysical experiments suggested that the entropy-based Bayesian procedure with a two-alternative forced choice task yielded accurate threshold estimates (within 2 dB) with as few as 30 trials, while a good estimate of the psychometric function slope takes, on average, 300 trials.

In the present study, three adaptive procedures were compared in auditory gap detection experiments: the up–down staircase procedure (Levitt, 1971), the entropy-based Bayesian procedure (Kontsevich & Tyler, 1999), and the UML procedure (Shen & Richards, 2012). In Experiment 1, the three procedures were evaluated in terms of the variability of the parameter estimates, test–retest repeatability and rates of convergence. Experiment 2 evaluated the performance of these procedures when frequent lapses of attention occurred.

Experiment 1: Estimating the psychometric function using three adaptive procedures

Method

Subjects Four normal-hearing listeners (S1–S4) participated in the present experiment. All of the listeners were 18 to 35 years of age and had audiometric thresholds equal to or better than 15 dB HL between 250 and 8,000 Hz in both ears. The left ears of the listeners were tested in the experiment. The subjects practiced the gap detection task for at least 2 h before the data collection began, after which the experiment was conducted in 2-h sessions. For each listener, no more than one session was run on a single day, and listeners were paid for their participation.

Stimuli The ability to detect a silent gap in an otherwise continuous sound is a measure of the auditory system's sensitivity to intensity fluctuations over time (e.g., Fitzgibbons & Wightman, 1982; Penner, 1975; Plomp, 1964; Shailer & Moore, 1983). In the present study, the detection of a silent gap in a broadband noise carrier was measured for four young, normal-hearing listeners. Four sound intervals were presented on each trial, separated by 500-ms interstimulus intervals. Each interval contained broadband noise presented at 70 dB SPL. The duration of the noise was 500 ms, including 5-ms cosine-squared onset/offset ramps. In either the second or the third interval, a brief silent gap was introduced to the temporal center of the noise. The duration of the gap was defined from the half-amplitude point of its cosine-squared onset to that of its offset. The listeners were instructed to select the interval that contained the gap, with the understanding that the gap would only occur in one of the middle two intervals.

All stimuli were generated digitally at a sampling frequency of 44,100 Hz and were presented to the left ear of each listener via a 24-bit sound card (Envy23 PCI controller, VIA Technologies, Inc., Taipei, Taiwan) installed on the experimental computer, a programmable attenuator (PA4, Tucker-Davis Technologies, Inc., Alachua, FL), a headphone buffer (HB6, Tucker-Davis Technologies), and a headphone (HD410 SL, Sennheiser, Old Lyme, CT). Each stimulus presentation was followed by visual feedback indicating the correct response. The experiment was conducted in a double-walled, sound-attenuating booth.

Procedure For the gap detection task, the psychometric function was assumed to take the form of a logistic function:

$$p = \gamma + (1 - \gamma - \lambda) \Big/ \left(1 + e^{-\beta(x-\alpha)}\right), \tag{1}$$

where p indicates the proportion correct; x, the gap duration in decibel units [$x = 20 \log(\text{gap duration})/(1 \times 10^{-3})$]; α, β, and λ are the threshold, slope, and lapse rate of the psychometric function; and $\gamma = .5$ is the chance performance level for the two-alternative forced choice paradigm.

Three procedures were used for data collection: (1) the up–down staircase procedure, (2) the entropy-based Bayesian procedure, and (3) the UML procedure. For the staircase procedure, 200 trials were run, which consisted of four adaptive tracks of 50 trials. On the first trial, the gap duration was 35 dB (56.2 ms, in the physical scale), which was reduced after two consecutive correct responses and increased after a single incorrect response. The initial step size of 8 dB was reduced to 5 dB after the first two reversals, and was reduced further to 2 dB after the first four reversals.

For the entropy procedure, the parameter space was a grid of α, β, and λ values. The α parameter took 18 values, ranging from −3 to 31 dB (0.7 to 35.5 ms) with 2-dB spacing. The β parameter took 11 log-spaced values, ranging from 0.1 to 10. The λ parameter took five values, linearly spaced between 0 and 0.2. Flat, uninformative priors were used for the three parameters. The signal strength—that is, the gap duration—had 21 potential values, linearly spaced between −9 and 35 dB (logarithmically spaced between 0.35 and 56.2 ms). Each adaptive track consisted of 200 trials, which were divided into four blocks of 50 trials. Following the procedure described by Kontsevich and Tyler (1999), before each trial, the posterior parameter distributions were calculated for each potential gap duration and each potential response (correct or incorrect). The entropies of these parameter distributions were calculated, and the expected total entropy was then derived for each potential gap duration. The gap duration that led to the minimum expected entropy was used in the following stimulus presentation. After obtaining the listener's response, the posterior parameter distributions were updated, and the procedure was repeated to select the gap duration for the next trial.

The parameter space for the UML procedure was the same as the one used in the entropy procedure. Each adaptive track consisted of 200 trials, which were divided into four blocks of 50 trials. The initial gap duration was 35 dB (56.2 ms). Following each trial, the posterior parameter distributions were calculated on the basis of the listener's response, which updated the best-fitting psychometric function. Then, the signal strength was placed at one of the four sweet points according to a two-down, one-up sweet-point selection rule (Shen & Richards, 2012). From short to long gap durations on the psychometric function, the four sweet points were the lower β sweet point, the α sweet point, the upper β sweet point, and the λ sweet point. The sweet points for the α and β parameters were reestimated on a trial-by-trial basis, while the sweet point for the λ parameter was fixed at 35 dB.[2] The gap duration was shifted to the adjacent lower sweet point after two consecutive correct responses, and was shifted to the adjacent higher sweet point after a single incorrect response. When the gap duration was already at the lowest sweet point (i.e., the lower β sweet point), the gap duration remained the same even if two correct responses were collected. Similarly, when the gap duration was at the highest sweet point (i.e., the λ sweet point), the gap duration stayed at that sweet point, even after an incorrect response was collected.

For each listener, gap detection data were collected using the three procedures in random order.[3] This included four adaptive tracks for the up–down procedure, and one track for each of the entropy and UML procedures.[4] When completed, the process was repeated, with the three procedures being tested in the reverse order.

Psychometric functions for individual listeners were estimated from the data collected for each procedure, one function for each repetition, yielding six psychometric functions per listener. This was done using the psignifit routine developed by Witchmann and Hill (2001a, 2001b). Flat priors were used for all parameters. The ranges of the parameters were from –20 to 20 for α, from 0.1 to 10 for β, and from 0 to .3 for λ. To provide a best estimate of the true underlying psychometric function, all of the data collected from each listener were pooled (1,200 trials), and the parameter estimates were calculated using the psignifit routine.

For each procedure and each of the α, β, and λ parameters, let ϕ_k denote the best parameter estimate using the pooled data for the *k*th listener, and let $\phi_{r,k,n}$ denote the parameter estimate obtained from the *k*th listener in the *r*th repetition and after the *n*th trial. The goodness of the parameter estimate for the *k*th listener, the *r*th repetition, and after n = 200 trials was quantified by a deviation $|\phi_{r,k,200} - \phi_k|$. When

$$\left|\phi_{r,k,200} - \phi_k\right| > 0.5\phi_k, \tag{2}$$

the parameter estimate for the *k*th listener in the *r*th repetition was considered poor. Besides the accuracy of the parameter estimates, two additional aspects of the experimental procedures—repeatability and rate of convergence—were also estimated. To quantify the repeatability, an across-repetition deviation R (at the end of 200 trials and averaged across listeners) was calculated as

$$R = \sqrt{\frac{\sum_k \left(\phi_{1,k,200} - \phi_{2,k,200}\right)^2}{4}}. \tag{3}$$

Smaller values of R indicated better repeatability. To investigate the rate of convergence, the root-mean-squared

[2] The expected variance of λ estimates was a monotonically decreasing function of x. Therefore, the λ sweet point did not correspond to a unique signal strength. Instead, it was defined at the upper limit of the stimulus parameter space.

[3] According to this design, the order in which the three procedures were tested could have been the same, by chance, for two or more listeners. However, this did not occur in either Experiment 1 or 2.

[4] The data collection for the up–down procedure consisted of four tracks of 50 trials, while for the entropy and UML procedures, data were collected in tracks of 200 trials. Four tracks were used with the up–down procedure (1) to represent the common practice of averaging threshold estimates from multiple up–down tracks, and (2) to increase stimulus sampling at long gap durations, usually occurring at the beginning of each up–down track. Frequent stimulus sampling at long gap durations is expected to improve the estimation of the lapse parameter of the psychometric function (Shen & Richards, 2012).

(RMS) deviation from the best estimate after the nth trial was calculated as

$$D_n = \sqrt{\frac{\sum_r \sum_k \left(\phi_{r,k,n} - \phi_k\right)^2}{8}}. \tag{4}$$

Note that D_n was defined for each trial, averaged across listeners and repetitions. The rate of convergence was reflected in how rapidly the value of D_n dropped with increasing numbers of trials.[5]

Results and discussion

The best parameter estimates (from the pooled data) and the estimates from the two repetitions of the three procedures are listed in the different columns of Table 1. The results are shown for each individual listener and for the three parameters of the psychometric function, in rows. Previous works have suggested that for broadband noise carriers, the gap detection threshold was about 2 ms (for 71 % correct; e.g., Forrest & Green, 1987), which corresponds to 6 dB on the stimulus parameter scale used in the present study. The α estimates obtained here are approximately the same.

Using the criterion specified in Eq. 2, the poor parameter estimates are indicated in Table 1 by hash marks. Comparing across the three procedures, poor parameter estimates occurred less frequently for the up–down procedure (four out of 24 occasions) than for the entropy (eight out of 24 occasions) and UML (eight out of 24 occasions) procedures. All three procedures provided fairly reliable estimates of the α parameter, for which no poor estimates were observed. The up–down and entropy procedures seemed to provide better estimation of the β parameter than did the UML procedure, while the up–down and UML procedures outperformed the entropy procedure in terms of the λ estimation.

Table 2 lists the values of R (Eq. 3) for α, ln β, and λ from the three procedures. Recall that R is a summary statistic, and smaller values of R mean better test–retest reliability. The values of R were comparable across different procedures, suggesting similar repeatability for the three procedures tested. Figure 1 plots the RMS deviation from the best estimate, D_n, as a function of trial number. For the α parameter (left panel), fast convergence of the estimates over the first 100 trials was observed. The rates of convergence were comparable across the three procedures. In contrast, the value of ln β converged gradually. Here, the rates of convergence were initially similar among the three procedures; after 100 trials, the up–down procedure began to converge more rapidly than the other two procedures, and ultimately provided the best estimate of β. For the λ parameter, the RMS deviations did not decrease with the trial number in a systematic fashion for the up–down and entropy procedures. On the other hand, a generally monotonic decreasing D_n was observed for the UML procedure.

Figure 2 illustrates the differences in stimulus placement for the three procedures. Each panel of Fig. 2 plots histograms of the gap durations presented to one of the listeners. The vertical dashed lines mark the sweet points according to the best parameter estimate for this listener. These sweet points are the optimal places to sample in order to minimize the variances in the threshold, slope, and lapse rate estimates, assuming a logistic psychometric function (e.g., Shen & Richards, 2012). For all three procedures, significant numbers of trials had signal strengths near the α and β sweet points (the leftmost three dashed lines). In this regard, the distributions of the gap durations were similar across procedures and listeners, except that listeners S2 and S4 showed a more concentrated distribution for the UML procedure than for the other two procedures. Moreover, all three procedures visited the λ sweet point (the rightmost dashed line), though the entropy and UML procedures spent more trials at the λ sweet point than did the up–down procedure. It is worth pointing out that although the λ sweet point was the best place (within the defined parameter space) to sample the stimuli for the estimation of λ, all gap durations associated with high percentages correct (see the labels above the dashed lines) contributed to the λ estimate. Although the up–down procedure visited the λ sweet point less frequently than did the other two procedures, it spent a significant proportion of trials at other gap durations in the high-percentage-correct region. Therefore, a reasonable estimate of λ was achieved using the up–down procedure, even though it did not specifically sample the stimuli at the λ sweet point.

The results from the present experiment suggested that although the algorithms for updating the stimulus placement were different for these procedures, the resulting distributions for the stimulus presentation were strikingly similar (Fig. 2). The up–down procedure, despite of its simplicity, provided better estimates of the β parameter than did the UML procedure, and better estimates of the λ parameter than did the entropy procedure. The success of the up–down procedure was consistent with the findings of Leek et al. (1992). These authors showed that when the lapse rate was assumed to be zero, the transformed up–down procedure provided accurate estimates of the psychometric function threshold and slope using 200 experimental trials.

For both the entropy and UML procedures, their performance would likely to be improved if appropriate prior parameter distributions were implemented. Informative prior distributions might help prevent the placement of the stimuli at extreme signal strengths during the entropy procedure, enhancing its efficiency. It is not clear, however,

[5] Note that, for the slope parameter, the calculations of R and D_n were based on ln β.

Table 1 The threshold (α), slope (β), and lapse rate (λ) parameters estimated for individual listeners for the two repetitions of Experiment 1. Parameters were estimated using three different procedures: (1) the up–down staircase procedure (up–down), (2) the entropy-based Bayesian procedure (entropy), and (3) the updated maximum-likelihood procedure (UML). The hash marks indicate the poor estimates (see the criterion of Eq. 2)

		Pooled	Up–Down		Entropy		UML	
			Rep1	Rep2	Rep1	Rep2	Rep1	Rep2
α	S1	5.90	5.78	5.84	3.77	7.32	4.37	5.61
	S2	4.92	5.41	5.05	4.88	5.28	3.69	5.18
	S3	6.96	8.00	5.72	6.97	6.40	9.00	6.42
	S4	6.81	5.25	7.86	4.57	7.70	5.85	7.17
β	S1	0.53	0.84#	0.32	2.44#	0.67	0.32	1.94#
	S2	0.63	0.91	0.43	0.48	0.68	1.80#	0.94#
	S3	0.64	0.57	0.96	1.24#	0.83	1.10#	0.50
	S4	0.78	0.81	1.27#	0.54	0.71	0.58	1.22#
λ	S1	.12	.06	.08	.25#	.15	.11	.12
	S2	.03	.02	.03	.00#	.02	.04#	.03
	S3	.04	.02	.03	.00#	.13#	.02	.08#
	S4	.04	.09#	.00#	.00#	.00#	.07#	.03

whether the introduction of priors would cause the entropy procedure to outperform the UML procedure, or vice versa. A systematic investigation of the effect of the prior distribution would be needed to explore this question.

In the present experiment, the quality of the parameter estimates was evaluated by comparing individual estimates to the best estimate ϕ_k, based on the pooled data across the three procedures (see Eqs. 2–4). However, if ϕ_k provided a biased estimate of the true psychometric function, the usefulness of these quality measures could be undermined. As pointed out by Kaernbach (2001), biases in parameter estimates could be a consequence of the sequential dependency inherent in adaptive procedures. Therefore, it is important to check whether ϕ_k agrees with the estimates from procedures in which the sampling of stimuli is independent of responses on previous trials. For this purpose, the estimation of the psychometric function was repeated for one of the listeners (S4) using the method of constant stimuli. Five blocks were run, each of which contained 60 trials. Within the 60 trials, six gap durations (3, 5, 7, 9, 11, and 13 dB) were tested in quasirandom manner, with ten responses being collected at each of the gap durations. Following the data collection, 300 trials of data were used to estimate the psychometric function using the psignifit routine. The resulting estimates were 6.48 for α, 0.75 for β, and .01 for λ. These estimates using the method of constant stimuli matched very closely to the best estimates from the pooled data listed in Table 1 for listener S4. No obvious bias was observed, except that the λ estimate was smaller using the method of constant stimuli. The close agreement between procedures with and without sequential dependency in stimulus sampling provided support for the validity of the best estimates ϕ_k.

Table 2 The across-repetition deviations R for the up–down, entropy, and updated maximum-likelihood (UML) procedures in Experiment 1. Smaller values indicate smaller difference in parameter estimates across replicates

	Up–Down	Entropy	UML
α	1.74	2.39	1.74
ln β	0.30	0.31	0.48
λ	.05	.08	.03

Experiment 2: Effect of inattention on estimates of the psychometric function

When estimating the threshold, slope, and lapse rate simultaneously, one of the major difficulties faced by the estimation algorithm (such as the psignifit routine) is that a shallow slope is easily confused with a high lapse rate, causing a bimodal instability. This occurs frequently when the lapse rate is high. This problem severely prevents the reliable measurement of the psychometric function in subjects who typically exhibit high lapse rates, such as naive subjects, subjects from clinical populations, infants, young children, and laboratory animals. The present experiment was designed to investigate whether this difficulty of estimating the psychometric function associated with the lapse of attention could be alleviated by the sampling strategies used by the three procedures.

Method

The same four listeners participated in Experiment 2. The stimuli and procedure used in the present experiment were identical to those of Experiment 1, except that on one fourth

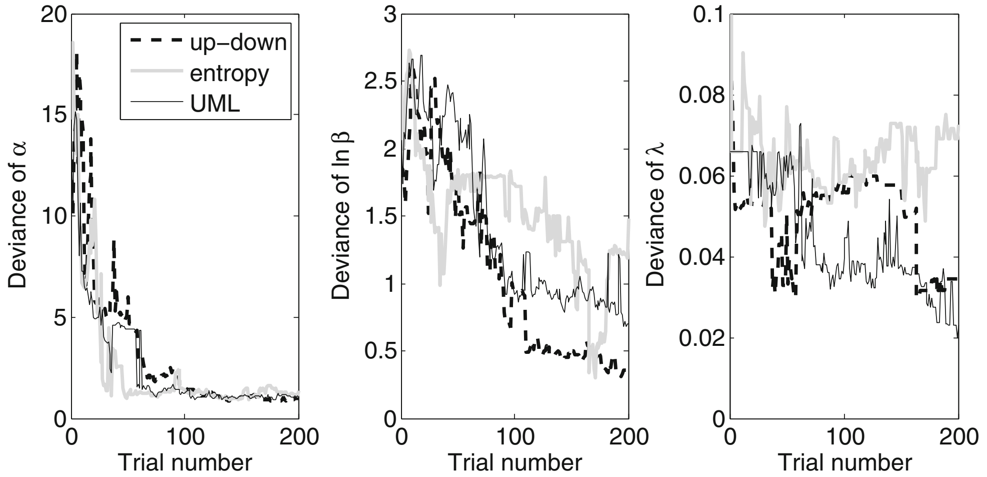

Fig. 1 RMS deviations between the parameter estimates and the best parameter estimates of the psychometric function, as a function of number of trials for Experiment 1. Changes in the functions indicate the rates at which the parameters converge for three parameters: α, ln β, and λ (*left to right panels*). The up–down staircase procedure, the entropy-based Bayesian procedure, and the updated maximum-likelihood (UML) procedure are plotted with different line styles

of the trials, determined at random, the listeners' responses were discarded and random responses were assigned. This manipulation was introduced as a simulation of frequent inattention during the experiment. Because during these inattention trials the correctness of responses was determined at random, the maximum proportion correct was bounded by .875 instead of 1. Therefore, the λ parameter was expected to be at least .125.

For each listener, best estimates of the psychometric function parameters were calculated, which were used as references to assess the accuracy provided by the three procedures. In contrast to Experiment 1, these best estimates were not obtained using the pooled data across procedures and repetitions. Because the expected lapse rates were very high, even pooling all of the collected data for each listener (1,200 trials) would not guarantee a reliable estimate of the psychometric function. When fitting the logistic psychometric function to the pooled data from Experiment 2 using the psignifit routine, the confidence intervals of the parameters were sometimes extremely large. On the other hand, when performing the same analyses on the pooled data from Experiment 1, much narrower confidence intervals were obtained. Therefore, the best parameter estimates of Experiment 2 were derived from the best estimates obtained using the pooled data from Experiment 1. Let p_1 be the best estimated psychometric function in Experiment 1; the best psychometric function estimate in Experiment 2 then was given by

$$p_2(\alpha_2, \beta_2, \lambda_2, x) = .75 \cdot p_1(\alpha_1, \beta_1, \lambda_1, x) + .25 \cdot .5. \qquad (5)$$

Consequently, the best parameter estimates for the two experiments followed the relationship $\alpha_2 = \alpha_1$, $\beta_2 = \beta_1$, and $\lambda_2 = .75\lambda_1 + .125$. To assess the goodness, repeatability, and rate of convergence of the estimates, $|\phi_{r,k,200} - \phi_k|$, R, and D_n were calculated, respectively, following the same procedure as in Experiment 1.

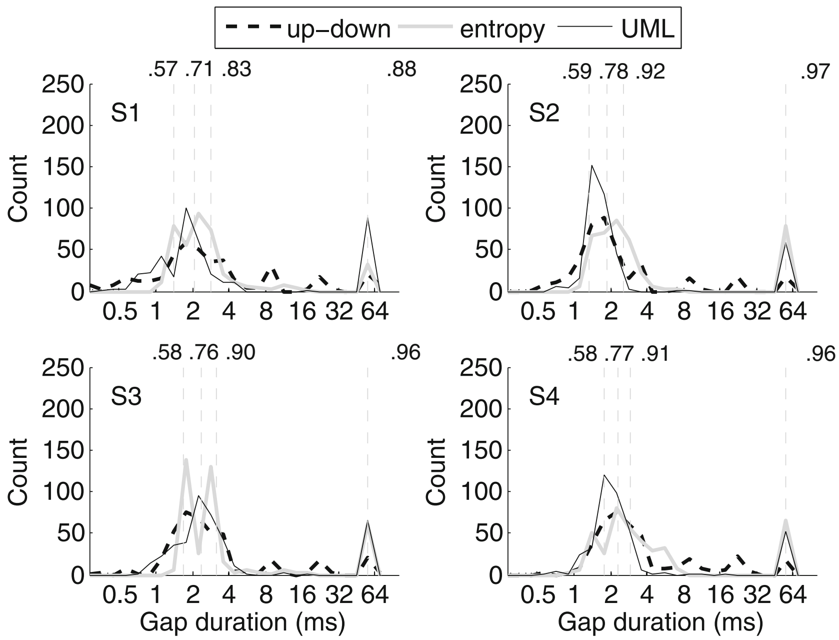

Fig. 2 Histograms of the gap durations tested for the up–down staircase procedure, the entropy-based Bayesian procedure, and the updated maximum-likelihood procedure (UML) for Experiment 1. Results for the individual listeners are plotted in separate panels. Within each panel, the *vertical lines* mark the sweet points derived from the best estimate of the psychometric function using the pooled data across procedures and repetitions. The proportions correct at the sweet points are labeled *above the vertical lines*

Results and discussion

The derived best parameter estimates and the estimates from the three procedures for the two repetitions are listed in the different columns of Table 3. The results are arranged as in Table 1. Comparing across the three procedures, poor parameter estimates (hash marks) occurred less frequently for the UML procedure (four out of 24 occasions) than for the entropy (six out of 24 occasions) and up–down (12 out of 24 occasions) procedures. The UML procedure was more successful in estimating the α and β parameters, as compared to the other two procedures. Relative to the results from Experiment 1 (Table 1), the parameter estimates were more variable, and the total number of the poor estimates was higher in Experiment 2. For example, for listener S1, the estimates provided by the up–down procedure in the present experiment did not at all resemble the best estimates. The α estimates were about 20 and 13 dB in the two repetitions, while the expected α value based on the results from Experiment 1 was approximately 6 dB. These results suggested that introducing frequent lapses of attention brought difficulties to the procedures. Among the three procedures tested, the UML procedure seemed to be the most resistant to inattention.

Table 4 lists the across-repetition deviations R for Experiment 2. The UML procedure exhibited smaller values of R, and hence better repeatability, for the α and λ parameters, whereas the up–down procedure had smaller values of R for ln β. Figure 3 plots the values of D_n as a function of trial number. For the α parameter (left panel), fast convergence of the estimates over the first 100 trials was observed for all three procedures. However, only the α estimate in the UML procedure converged to a value that was close to the expected value. The up–down and entropy procedures gave biased estimates for the α parameter, and these biases were largely associated with a single listener (S1). For this listener, the α estimates obtained from the up–down procedure and from the first repetition of the entropy procedure were much larger than the best estimates predicted using the data from Experiment 1 (see Table 3). The convergence of ln β was fairly unstable for all of the procedures, but for the UML procedure, the convergence was somewhat more consistent: After 200 trials, the UML procedure provided the best estimates of β. For the λ parameter, all three procedures showed rapid convergence, with the values of D_n being comparable across procedures throughout the 200 trials.

Table 4 The across-repetition deviations R for the up–down, entropy, and updated maximum-likelihood (UML) procedures in Experiment 2

	Up–Down	Entropy	UML
α	3.64	6.75	1.88
ln β	0.42	1.00	0.61
λ	.12	.08	.03

Figure 4 illustrates the differences in stimulus placement for the three procedures tested. For the UML procedure, the distributions of gap durations were very similar to those obtained in Experiment 1 (see Fig. 2). The stimuli were concentrated into two areas: One was near the α and β sweet points, and the other was at the upper limit of the gap durations (56.2 ms). The stimulus distributions for the entropy procedure were similar to those for the UML procedure for three of the listeners (S2, S3, and S4). For listener S1, however, the stimuli were broadly distributed across all gap durations. This was also the listener who had exhibited

Table 3 The threshold (α), slope (β), and lapse rate (λ) parameters estimated for individual listeners for the two repetitions of Experiment 2. Parameters were estimated using three different procedures: (1) the up–down staircase procedure (up–down), (2) the entropy-based Bayesian procedure (entropy), and (3) the updated maximum-likelihood procedure (UML). The hash marks indicate the poor estimates (see the criterion of Eq. 2)

		Predicted	Up–Down		Entropy		UML	
			Rep1	Rep2	Rep1	Rep2	Rep1	Rep2
α	S1	5.90	19.90$^{\#}$	13.28$^{\#}$	19.93$^{\#}$	6.63	3.45	6.94
	S2	4.92	5.02	7.60$^{\#}$	4.86	5.16	5.01	4.90
	S3	6.96	8.04	8.95	4.82	5.85	4.63	6.03
	S4	6.81	6.80	8.12	6.96	9.00	8.86	9.01
β	S1	0.53	0.10$^{\#}$	0.10$^{\#}$	0.10$^{\#}$	0.27	0.67	7.18$^{\#}$
	S2	0.63	0.76	4.94$^{\#}$	5.30$^{\#}$	0.44	0.59	1.96$^{\#}$
	S3	0.64	0.24$^{\#}$	0.15$^{\#}$	7.29$^{\#}$	0.27$^{\#}$	0.38	0.49
	S4	0.78	1.25$^{\#}$	1.71$^{\#}$	6.45$^{\#}$	1.02	1.11	2.55$^{\#}$
λ	S1	.21	.01$^{\#}$	.22	.21	.19	.21	.26
	S2	.14	.08	.14	.19	.08	.10	.14
	S3	.16	.12	.00$^{\#}$	.22	.10	.16	.14
	S4	.15	.15	.12	.16	.13	.10	.08$^{\#}$

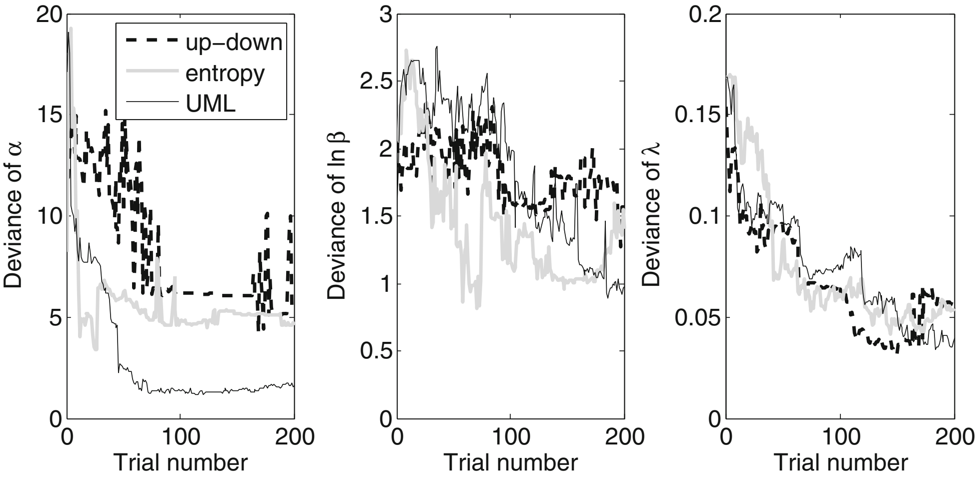

Fig. 3 RMS deviations between the parameter estimates and the best parameter estimates of the psychometric function, as a function of number of trials for Experiment 2. Changes in the functions indicate the rates at which the parameters converge for three parameters: α, ln β, and λ (*left to right panels*). The up–down staircase procedure, the entropy-based Bayesian procedure, and the updated maximum-likelihood (UML) procedure are plotted with different line styles

the largest lapse rate in Experiment 1. Therefore, it seems possible that listeners who naturally have high lapse rates would be more likely to exhibit suboptimal sampling of stimuli and poor psychometric function estimates when additional lapses of attention are artificially introduced. In such situations, the dynamic range of the psychometric function, from the chance performance level to the upper asymptote, would be very narrow, which would make the identification of the optimal place to sample the stimuli extremely difficult. However, it is not clear why a large lapse rate would affect the entropy procedure more than the UML procedure. Broad distributions of gap durations were observed for the up–down procedure for all four listeners. Listeners S2, S3, and S4 had the highest concentrations of stimuli near the α and β sweet points, while for listener S1, almost all stimuli were presented at gap durations above the α and β sweet points. These results indicated that when the lapse rate was high, it would take large numbers of trials for the stimuli in an up–down track to approach its targeted percentage correct (e.g., 70.7 % correct for a two-down, one-up track; Levitt, 1971).

In summary, the present experiment introduced random responses in order to simulate lapses of attention. Among the three procedures tested, the UML procedure seemed to be less affected by this manipulation, demonstrating a similar accuracy, repeatability, and rate of convergence as compared to Experiment 1, where no artificial inattention was introduced. On the other hand, frequent lapses of attention had detrimental effects on performance for the up–down procedure. Poor estimates of the α and β parameters and poor convergence of ln β were found in the present experiment. The entropy procedure gave reasonable estimates of the psychometric function, except for one listener, who also had the highest lapse rate. For the up–down and entropy

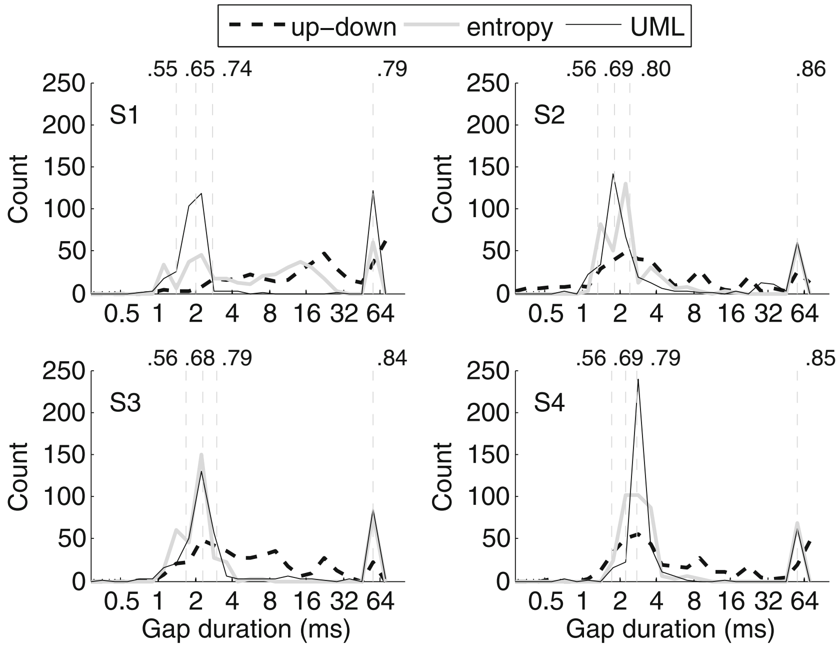

Fig. 4 Histograms of the gap durations tested for the up–down staircase procedure, the entropy-based Bayesian procedure, and the updated maximum-likelihood procedure (UML) for Experiment 2. Results for the individual listeners are plotted in separate panels. Within each panel, the *vertical lines* mark the sweet points derived from the best estimate of the psychometric function using the pooled data across procedures and repetitions. The proportions correct at the sweet points are labeled *above the vertical lines*

procedures, the failures in the estimation of the psychometric function were associated with poor sampling strategies.

Conclusions

Three adaptive procedures were compared against one another as means of estimating the psychometric function for an auditory gap detection task. The psychometric function was modeled as a logistic function that was described by three parameters: threshold, slope, and lapse rate. The results from four listeners showed that the up–down staircase procedure (up–down procedure; Levitt, 1971), the entropy-based Bayesian procedure (entropy procedure; Kontsevich & Tyler, 1999), and the updated maximum-likelihood procedure (UML procedure; Shen & Richards, 2012) performed similarly in estimating the threshold of the psychometric function for this task. The up–down procedure provided more efficient estimation of the slope and the lapse rate than did the other two procedures. When the lapse rates of the listeners were elevated via experimental manipulations, the up–down procedure gave poor estimates of the threshold and slope of the psychometric function, presumably because it failed to optimize the stimuli for estimating these parameters. The UML procedure was less sensitive to the increased occurrences of inattention as compared to the entropy procedure. Therefore, if low lapse rates are expected, the up–down procedure could be a simpler and slightly superior method for estimating the parameters of the psychometric function simultaneously. However, when high lapse rates are expected or the lapse rates are unknown, the UML procedure is more likely to provide reliable estimates of the psychometric function.

Author Note This research was supported by National Institutes of Health Grant No. R21 DC010058 awarded to Virginia M. Richards. The author thanks Theodore Lin and Andrew Silva for their assistance in data collection and in preparation of the manuscript.

References

Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *Journal of the Acoustical Society of America, 111,* 2801–2810.

Derman, C. (1957). Non-parametric up-and-down experimentation. *Annals of Mathematical Statistics, 28,* 795–798.

Durham, S. D., & Flournoy, N. (1995). Up-and-down designs I: Stationary treatment distributions. In N. Flournoy & W. F. Rosenberger (Eds.), *Adaptive designs: Papers from the Joint AMS–IMS–SIAM Summer Conference held at Mt. Holyoke College, South Hadley, MA, July 1992* (pp. 139–157). Hayward, CA: Institute of Mathematical Statistics.

Fitzgibbons, P. J., & Wightman, F. L. (1982). Gap detection in normal and hearing-impaired listeners. *Journal of the Acoustical Society of America, 72,* 761–765.

Forrest, T. G., & Green, D. M. (1987). Detection of partially filled gaps in noise and the temporal modulation transfer function. *Journal of the Acoustical Society of America, 82,* 1933–1943.

Green, D. M. (1990). Stimulus selection in adaptive psychophysical procedures. *Journal of the Acoustical Society of America, 87,* 2662–2674.

Green, D. M. (1993). A maximum-likelihood method for estimating thresholds in a yes–no task. *Journal of the Acoustical Society of America, 93,* 2096–2105.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up–down method. *Perception & Psychophysics, 49,* 227–229. doi:10.3758/BF03214307

Kaernbach, C. (2001). Slope bias of psychometric functions derived from adaptive data. *Perception & Psychophysics, 63,* 1389–1398.

King-Smith, P. E., & Rose, D. (1997). Principles of an adaptive method for measuring the slope of the psychometric function. *Vision Research, 37,* 1595–1604.

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

Leek, M. R., Hanna, T. E., & Marshall, L. (1992). Estimation of psychometric functions from adaptive tracking procedures. *Perception & Psychophysics, 51,* 247–256.

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

Penner, M. J. (1975). Persistence and integration: Two consequences of a sliding integrator. *Perception & Psychophysics, 18,* 114–120.

Plomp, R. (1964). Rate of decay of auditory sensation. *Journal of the Acoustical Society of America, 36,* 277–282.

Shailer, M. J., & Moore, B. C. (1983). Gap detection as a function of frequency, bandwidth, and level. *Journal of the Acoustical Society of America, 74,* 467–473.

Shen, Y., & Richards, V. M. (2012). A maximum-likelihood procedure for estimating psychometric functions: Thresholds, slopes, and lapses of attention. *Journal of the Acoustical Society of America, 132,* 957–967.

Taylor, M. M., & Creelman, C. D. (1967). PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America, 41,* 782–787.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33,* 113–120. doi:10.3758/BF03202828

Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63,* 1293–1313. doi:10.3758/BF03194544

Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics, 63,* 1314–1329. doi:10.3758/BF03194545