

The single interval adjustment matrix (SIAM) yes–no task: an empirical assessment using auditory and gustatory stimuli

Daniel Shepherd · Michael J. Hautus ·
Miriam A. Stocks · Siew Y. Quek

Published online: 2 May 2011
© Psychonomic Society, Inc. 2011

Abstract The SIAM yes–no task is an efficient bias-free adaptive procedure for estimating absolute thresholds, though it arguably requires further evaluation prior to its adoption into mainstream psychological research. We report two experiments undertaken in the auditory and gustatory modalities designed to assess the accuracy and efficiency of the SIAM method. In the first experiment, estimates of absolute thresholds for 1000-Hz tones obtained using a two-alternative forced choice adaptive procedure were compared to those obtained using both the SIAM yes–no task and a modification of the SIAM task incorporating the method of free response, the SIAM-rapid. In [Experiment 2](#), we compared absolute thresholds for sucrose in solution obtained with either a two-alternative forced choice adaptive procedure or the SIAM yes–no task. Both experiments demonstrated the efficiency and validity of the SIAM approach, with SIAM thresholds proving to be equivalent to those obtained in the two-alternative forced choice tasks and to those reported in the literature.

Keywords Psychometrics/testing · Audition · Gustation

D. Shepherd
Department of Psychology, AUT University,
Auckland, New Zealand

M. J. Hautus · M. A. Stocks · S. Y. Quek
Department of Psychology and Department of Chemistry,
University of Auckland,
Auckland, New Zealand

D. Shepherd (✉)
University of Auckland,
Private Bag 92019, Auckland, New Zealand
e-mail: daniel.shepherd@auckland.ac.nz

The determination of a stimulus level along the psychometric function that conforms to some predefined performance criterion (e.g., 75% correct in a single-interval task) defines a detection task. Taylor (1971) and, later, Treutwein (1995) describe three primary considerations when selecting threshold estimation procedures. First, will the task produce valid and reliable estimates? Second, is the procedure easy to implement? Third, is the task the most efficient, given the context? To these considerations, we add a fourth: Is the procedure simple and immediately intuitive to the participant? In the last 50 years of psychophysical research, a class of procedures known as *adaptive procedures* have emerged as best satisfying these four criteria. A variety of validated adaptive methods are on offer (see Treutwein, 1995), which Macmillan and Creelman (2005) categorise as maximum likelihood methods (e.g., QUEST and Best PEST), nonparametric methods (e.g., PEST), and staircase methods. The last category, also nonparametric in nature, includes the ubiquitous two-down one-up and three-down one-up staircase procedures. Substantial effort has been expended in evaluating and comparing these methods (e.g., Gu & Green, 1994; Hesse, 1986; Macmillan & Creelman, 2005, chap. 11; Stillman, 1989; Taylor, Forbes, & Creelman, 1983), and for a given context, each method has its supporters and detractors. Generally, the two-interval adaptive methods enjoy greater popularity than single-interval methods, even though the two-interval procedures are inherently lower in statistical efficiency and higher in statistical bias than single-interval tasks (Kershaw, 1985; Klein, 2001; Madigan & Williams, 1987; McKee, Klein, & Teller, 1985; Ulrich & Vorberg, 2009). In this report, we assess, using gustatory and auditory stimuli, a relatively new variant of the single-interval adaptive procedures, the single-interval adjustment matrix yes–no (SIAM YN) task (Kaernbach, 1990a, 1990b). Additionally, in the auditory context we modify the SIAM

procedure to incorporate aspects of the method of free response (Egan, Greenber, & Schulman, 1961), so as to encourage further efficiency gains.

Kaernbach (1990a, 1990b) noted that, for a standard yes–no task, a point in receiver operating characteristic (ROC) space (see Fig. 1) is determined by two factors (Green & Swets, 1974): first, the ability of the participant to detect the signal (i.e., sensitivity), and second, the locality of their response criterion (i.e., response bias). Different methods exist to influence criterion placement, one example being the use of a payoff matrix to classify response consequences as rewards or punishments. Response payoffs can be experimentally manipulated to induce participants to adopt different response criteria—for example, a low (bias to saying “yes”), neutral (i.e., equally inclined to respond “yes” or “no”), or high (i.e., bias to responding “no”) criterion. Kaernbach, with remarkable insight, proposed that an adjustment matrix, adjusting the signal intensity, can be substituted in place of a payoff matrix, producing an adaptive yes–no task that converges on a bias-free estimate of the absolute threshold at a target performance specified by the investigator. Assuming that the probability of the signal being presented on any one trial is chance (i.e., $p = .5$), and for the moment leaving target performance (t) unspecified, the adjustment matrix, according to which stimulus level is adjusted given a specific outcome from the previous trial, is given in Table 1.

Target performance (i.e., t), as defined by Kaernbach (1990b), is the maximum distance between the ROC curve and the major diagonal. This distance, representing the maximal difference between hit (H) and false alarm (FA) rates, is termed the *maximum reduced hit rate* (MRHR). Figure 1 displays the MRHR, which is the vertical

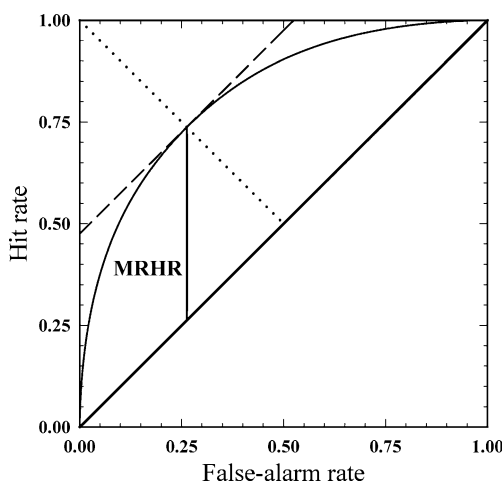


Fig. 1 A receiver operating characteristic (ROC) curve, contained in ROC space, showing the maximum reduced hit rate (MRHR), which is the vertical distance between the point on the ROC curve possessing a slope of 1 (dashed line) and the major diagonal (thick line)

Table 1 Payoff matrix for the SIAM yes-no task when the desired target performance (t) is unspecified

	Responds “Yes”	Responds “No”
Trial contained signal: ($p = .5$)	Hit –1	Miss $t/(1 - t)$
Trial contained noise: ($p = .5$)	False Alarm $1/(1 - t)$	Correct Rejection 0

difference between the point on the ROC curve possessing a slope of 1—for a symmetrical ROC curve, this point must fall along the minor diagonal—and the major diagonal; for this case, $H = 1 - FA$, and MRHR is therefore equal to $2H - 1$. The SIAM YN procedure involves “shifting” a participant’s operating point in ROC space to the position corresponding to the MRHR (i.e., falling on the minor diagonal for a symmetrical ROC) while simultaneously adjusting the level of the target stimulus to match a target performance level, $MRHR = t$. This entails inducing the participant to adopt a neutral response criterion that, Kaernbach (1990b) claimed, can be achieved using trial-by-trial feedback. Importantly, a failure to maintain a neutral response criterion can only ever produce higher threshold estimates. Thus, a key assumption of the SIAM approach is that participants are inherently motivated to maximise performance, so that a decrease in level becomes a reinforcer and an increase in level a punisher. A value of $t = .5$ was recommended as a suitable target level of performance, as this value is equivalent to 75% correct responses by an unbiased responder undertaking a yes–no task. The corresponding adjustment matrix is shown in Table 2, where all four values in the matrix can be scaled by a constant to better suit the experimental context.

Another assumption of the SIAM task that warrants further scrutiny is the shape of the ROC—for example, Figure 1, which is based on the Gaussian model. The area underneath the ROC (AUROC) is related to the MRHR, and while both are valid measures of sensitivity, the proportion correct corresponding to the MRHR varies with the form of the ROC. The signal detection model underlying the ROC determines the exact form of the relationship between AUROC and MRHR. Kaernbach (1990a, 1990b) demonstrated, using theoretical relationships and empirical data, that while different signal detection models can marginally distort a proportional relationship between AUROC and MRHR, empirical data manifest a proportional relationship between the two. Further to this, in Monte Carlo simulations, Kaernbach (1990a, 1990b) showed that the systematic error of the SIAM procedure was not related to the assumed form of the

Table 2 Payoff matrix for the SIAM yes-no task when the desired target performance (t) equals 0.5

	Responds “Yes”	Responds “No”
Trial contained signal: ($p = .5$)	-1	$t/(1-t) = .5/(1-.5) = 1$
Trial contained noise: ($p = .5$)	$1/(1-t) = 1/(1-.5) = 2$	0

ROC, whereas the same is not true for the von Békésy tracking procedure.

A commonly used performance index associated with symmetrical ROC curves is the signal detection theory measure d' . If $t = \text{MRHR} = .5$, then $FA = .25$ and $H = .75$, yielding $d' = z(.75) - z(.25) = 1.35$ (Green & Swets, 1974). After the application of Green's area theorem (Green, 1964), and assuming an unbiased responder, a d' of 1.35 is shown to be equivalent to 83% correct in a two-alternative forced choice (2AFC) task, which is higher than the performance level conventionally targeted as the threshold (i.e., $d' = 1$; 76% correct in 2AFC). For the more general case of an asymmetrical ROC curve, assuming that $t = .5$, the MRHR equates to between 75% and 87.5% correct in a two-interval forced choice task, with this range determined by the underlying signal detection model (Kaernbach, 1990b). Despite these issues, we adhere to Kaernbach's (1990b) recommended definition of threshold (i.e., $t = .5$), on the basis that threshold is an operationally defined concept dependent upon context, and this value yields 75% correct responses in the context of the yes–no task, even if only when the ROC is symmetrical.

A commonly used method of threshold estimation, the staircase method (Cornsweet, 1962), is popular in experimental psychology due to its efficiency and reliability. This procedure derives threshold estimates by having the stimulus level traverse a participant's threshold multiple times during a block of trials. The starting level of the stimulus is either well above or well below threshold and increases or decreases in discrete steps (i.e., the step size). The progression rule is predetermined by the experimenter and is frequently either two-down one-up or three-down one-up. With a three-down one-up regime, three correct responses in a row are required in order to reduce the stimulus level by some amount. An incorrect response at any stage leads to an increase in level. The transformed up–down method (Levitt, 1971) tracks a participant's performance in terms of the percentage of correct responses and allows threshold to be defined as a specific percentage correct, usually 70.7% (two-down one-up) or 79.4% (three-down one-up) correct in a 2AFC task. Using the example of a three-down one-up procedure, the level of the stimulus converges to a point at which the probability of the level decreasing equals the probability of the level increasing. At threshold, the probability of three correct responses in a row must then be equal to .5—that is, $p \cdot p \cdot p = .5 = p^3 = .5$. Thus, the percentage correct level to which a three-down

one-up procedure converges must be $(.5)^{1/3} = .794$, which corresponds approximately to the 79%-correct point on the psychometric function (Levitt, 1971).

In the present study, absolute thresholds for 1000-Hz tones and for sucrose in aqueous solution are estimated using both a 2AFC staircase procedure and the SIAM YN task. The 2AFC staircase task is ubiquitous in sensory and perception research and can be considered a standard to which other procedures can be benchmarked (Gu & Green, 1994; Kaernbach, 1990a; Macmillan & Creelman, 2005; Stillman, 1989). The 2AFC procedure is generally considered to be less affected by bias than the yes–no task and provides relatively simple data for analysis, though it is made less efficient by the requirement for two observation intervals, as opposed to one. Several researchers have used the SIAM YN task in areas such as visual feature search with hemineglect patients (Brooks, Wong, & Robertson, 2005; List et al., 2008), brightness matching (Brown & Rudd, 1998), and auditory filter measurements (Leeuw & Dreschler, 1998). However, 20 years have passed since the description of the SIAM YN task was published (i.e., since Kaernbach, 1990b), and these studies appear to be the only reported use of the SIAM task. That the SIAM task is so rarely utilized can be explained by an insufficient degree of evaluation that would give researchers the confidence to use the technique, or perhaps by competition from another single-interval procedure, the maximum likelihood estimation yes–no task (MLE YN; Green, 1990). We note that, though Kaernbach's (1990b) SIAM YN task is endowed with a strong theoretical backbone, it still requires further empirical validation, especially since Kaernbach's (1990b) own assessment utilised only 4 participants in one sensory context, audition. Hautus, Stocks, and Shepherd (2010) employed the SIAM YN task in the gustatory context (sucrose detection) but didn't compare the task against other adaptive procedures. In this article, we report a set of sterner tests designed to assess the veracity of the SIAM approach in the auditory (Exp. 1) and gustatory (Exp. 2) modalities.

Experiment 1: pure-tone detection

Experiment 1 assesses the suitability of the SIAM YN task to efficiently and accurately estimate absolute thresholds for 1000-Hz tones. Two versions of the SIAM YN task are

employed; the first is described by Kaernbach (1990a, 1990b), in which the presentation of the next experimental trial is contingent upon response. The second version, which we call the SIAM-rapid task, is similar to the go/no-go task in that it does not require an indefinite pause while awaiting a participant's response and incorporates aspects of the method of free response (Egan et al., 1961). Instead, the SIAM-rapid paradigm extracts information from both a response (e.g., a buttonpress) and a nonresponse (i.e., no action) and uses this information to adjust the stimulus magnitude for the next trial. The aim of regulating the response period is to create a more temporally efficient task, and we will compare the SIAM YN and SIAM-rapid tasks in terms of trial and block times.

To evaluate the two versions of the SIAM task, the validation protocol used by Gu and Green (1994), who assessed the efficiency of an MLE YN task, is adopted. Gu and Green compared absolute thresholds derived using maximum likelihood estimation with those obtained using a three-down one-up 2AFC procedure. In the present context, it would be expected that thresholds estimated using the SIAM YN task would be statistically indistinguishable from those collected using the 2AFC staircase task, although there might be a slight trend toward higher SIAM thresholds, due to SIAM's more demanding performance regime—that is, about 83% correct in the SIAM task versus 79.4% correct in the staircase task. In practical terms, however, we would expect empirical differences to be minor.

Method

Participants

A group of 19 experimentally naive participants, 8 males and 11 females between 21 and 36 years of age, participated in Experiment 1. The participants were recruited from a pool of postgraduate students and were offered a monetary reward to participate. None reported any current or historical hearing pathology, and all listened monaurally with their left ear.

Materials

Tones A personal computer running LabVIEW 8.1 (National Instruments, NI) and housing a digital-to-analog conversion card (NI, PXI 4461) produced 1000-Hz tones at a sampling rate of 44.1 kHz, the levels of which were controlled by a programmable attenuator (Tucker Davis Technologies, TDT, PA5). The tones were 10 ms in duration with 1-ms ramps (\cos^2) and were transmitted to a monaural earpiece (Telephonics, TDH-49P) via a headphone buffer (TDT HB7). The participant, seated in a sound-attenuating chamber (Amplaid, Model E) in front of

a panel of LED feedback lights, responded using one of two buttons (2AFC and SIAM YN tasks) or a single button (SIAM-rapid).

Procedure

In Experiment 1, we applied three detection procedures to estimate absolute thresholds for short-duration 1000-Hz tones. For all three tasks, the initial level of the tone was set at 30 dB SPL, and this level tracked up or down contingent on response. Full instructions were given to participants prior to commencing a block of trials, and trial-by-trial feedback was given. Participants undertook 10 blocks of trials per task, and each block terminated after 15 turnarounds. Absolute thresholds, calculated for each block, were taken as the average of the last 12 turnarounds, with the first 3 turnarounds being disregarded in order to control for statistical bias. The three tasks, described now, were presented in a random order across the experimental series, and participants undertook a minimum of 1 block of each task during any experimental session. Each experimental session consisted of 5 blocks of trials, and participants attended six sessions.

2AFC detection task In the 2AFC detection task, each trial consisted of two observation intervals, with the tone randomly assigned to one or the other of the intervals with equal probability ($p = .5$). If a participant made an incorrect response, the level was increased by 1 dB, while after three consecutive correct responses the level of the tone was decreased by 1 dB. Each trial began with the illumination of a centrally located light emitting diode (LED) for 400 ms, followed by a pause of 400 ms, and then the first and second observation intervals (10 ms each) were presented, separated by a 400-ms interstimulus interval. Each observation interval was accompanied by the illumination of the same LED. A response interval ensued in which the participant was required to report which interval contained the tone. Response was by button (left button = "1st interval," right button = "2nd interval"), which triggered feedback in the form of a left-of-centre LED (tone was in the 1st interval) or right-of-centre LED (tone was in the 2nd interval).

SIAM YN task In the SIAM YN task, only a single observation interval was presented per trial, which either did or did not contain the tone ($p = .5$). On any one trial, a warning LED was flashed for 400 ms, followed by a 400-ms pause, and then a single 10-ms observation interval, accompanied by an LED flash. The participant indicated whether or not the observation interval contained the tone using either a green- ("yes") or red- ("no") coloured button.

A response interval of unconstrained duration preceded feedback in the form of a left-of-centre LED (tone was present) or a right-of-centre LED (tone was absent). The level of the tone was adjusted on a trial-by-trial basis using the adjustment matrix corresponding to $t = .5$, with a hit decreasing level by 1 dB, a correct rejection leaving the tone unchanged, and a miss or false alarm increasing the level of the tone by 1 or 2 dB, respectively.

SIAM-rapid For the SIAM-rapid task, participants were asked to press a button only if they judged a tone to have coincided with the flash of a green LED. The presentation of the next trial was only partially controlled by the participant; that is, the depression of the button resulted in immediate feedback and the initiation of the next trial, while no response incurred a wait of 1,000 ms, after which feedback was provided. If a tone was present and the participant indicated as much by depressing the button, the outcome was classified as a hit (i.e., level reduced by 1 dB), whereas if the participant offered no response, the outcome was a miss (i.e., level increased by 1 dB). If the tone was absent but the participant depressed the button, the outcome was a false alarm (i.e., level increased by 2 dB). Otherwise, no response constituted a correct rejection (i.e., no change to signal level). All other procedural parameters were identical to those employed in the SIAM YN task.

Results and discussion

Figure 2 displays tracks for each of the three tasks, with each plot being the last block of trials of each task undertaken by Participant 1 (see Table 3 for the means). Note that, for the SIAM YN (Fig. 2b) and SIAM-rapid (Fig. 2c) tasks, the levels that are plotted are always those that would apply had the trial contained the target. Thus, the false alarms and correct rejections (i.e., blank trials) in Fig. 2b and c are not plotted as 0 dB SPL, which in our opinion would provide a degraded representation of the stimulus level converging around threshold. Figure 3a (top panel) plots mean absolute thresholds (dB SPL), accompanied by 95% confidence intervals, as a function of block number and task. Each data point is the average threshold of the 19 participants. For any one experimental block, the greatest difference in mean threshold estimates between any two of the three tasks is 3.4 dB SPL, with a mean difference of 2.14 dB SPL across all tasks and blocks. A repeated measures ANOVA was undertaken using two within-groups factors, Task (three levels) and Block (10 levels). If the Mauchly's test of sphericity statistic was significant, the Greenhouse–Geisser correction was then utilised. No significant differences between the mean 2AFC thresholds ($M_{2AFC} = 22.7$ dB SPL, $SD = 1.04$), SIAM YN thresholds

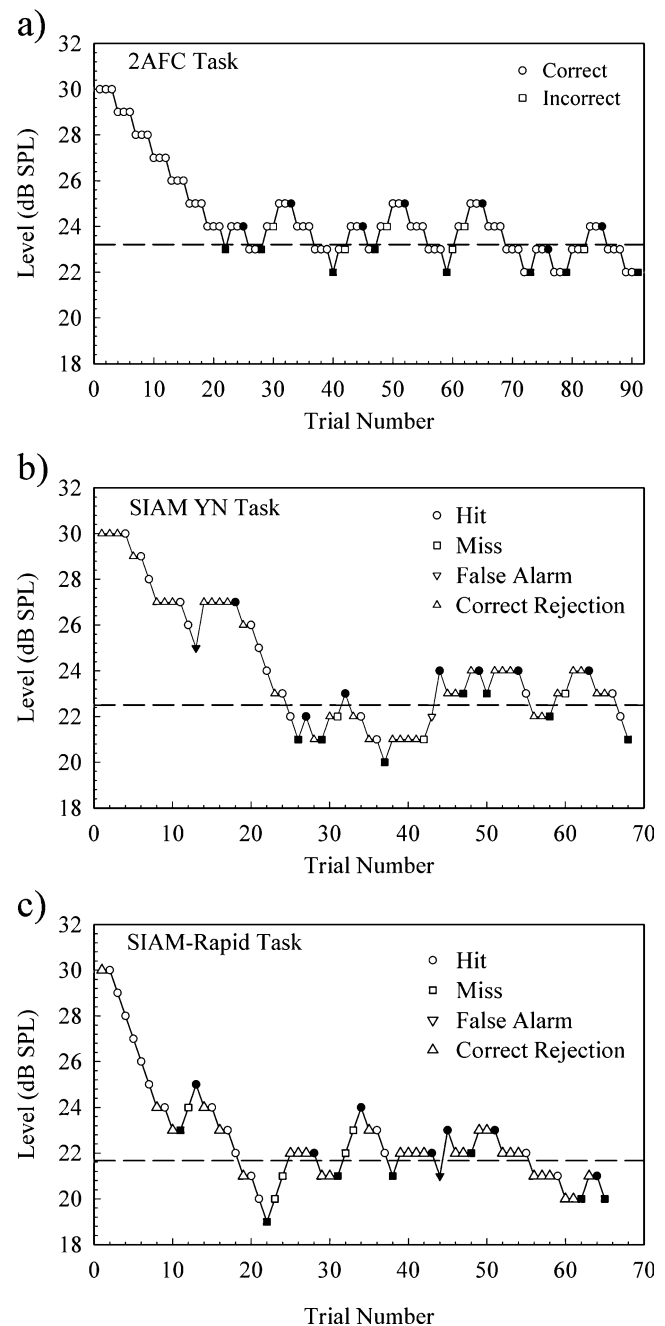


Fig. 2 Level as a function of trial number for three tasks: (a) the 2AFC three-down one-up adaptive task, (b) the SIAM YN task, and (c) the SIAM-rapid task. In panel (a), the circles correspond to correct responses and squares to incorrect responses. In plots (b) and (c), hits (circles), misses (squares), false alarms (inverted triangles), and correct rejections (triangles) are indicated. For all three plots, the dashed horizontal line indicates the final threshold calculation for that block, and solid symbols represent turnarounds

($M_{SIAM} = 22.98$ dB SPL, $SD = 3.67$), and SIAM-rapid thresholds ($M_{Rapid} = 23.871$ dB SPL, $SD = 4.68$) were found [$F(2, 36) = 2.339$, $p = .144$, $\epsilon = .795$]. Additionally, no main effect of block was noted [$F(9, 162) = 1.319$, $p = .278$, $\epsilon = .279$], nor a significant interaction effect between

Table 3 Absolute thresholds (dB SPL) for 1000-Hz tones estimated using the 2AFC, SIAM YN, and SIAM-rapid tasks

Participant	2AFC Adaptive Task		SIAM YN Task		SIAM-Rapid Task	
	Mean	SD	Mean	SD	Mean	SD
1	19.877	1.579	23.298	2.019	23.101	1.497
2	21.769	2.466	22.883	2.184	22.867	2.486
3	21.654	3.000	23.794	2.722	21.161	3.492
4	22.385	1.492	25.342	1.678	26.607	1.274
5	18.608	2.112	23.820	2.318	23.985	2.946
6	17.885	3.667	20.547	1.959	19.940	1.654
7	15.692	2.258	13.034	0.873	13.614	0.663
8	25.556	2.780	28.770	2.357	27.474	5.937
9	27.308	3.927	28.900	2.478	36.018	5.079
10	19.131	2.748	18.179	2.305	18.454	5.142
11	29.977	2.931	29.118	2.043	30.268	6.260
12	24.577	2.988	22.733	3.705	25.184	5.180
13	26.287	2.368	21.771	2.747	21.291	1.829
14	24.060	1.085	23.403	2.573	22.113	1.555
15	28.964	3.929	24.413	1.492	26.388	1.327
16	21.473	1.363	22.628	1.347	24.716	1.567
17	21.726	1.377	22.020	2.645	22.366	1.669
18	20.720	1.397	20.757	1.155	22.337	1.213
19	23.644	3.466	23.327	3.514	25.660	1.636
Mean	22.700	2.470	23.091	2.217	23.871	2.754
SD	3.792	0.913	3.730	0.726	4.680	1.825

Each threshold estimate is the average of ten blocks, and is accompanied by a standard deviation (SD)

task and block [$F(18, 324) = 1.121, p = .278, \epsilon = .392$]. Follow-up polynomial contrasts indicated no significant linear or quadratic effects across blocks within a task. Table 3 presents mean threshold values for each participant and each task. Differences across participants are noted and expected, and will arise in part from individual differences in auditory acuity. Also presented are standard deviations, and though these are over 3 dB in some cases, it must be remembered that only naive participants were recruited. The mean standard deviation across participants is slightly higher for the SIAM-rapid task than for either the 2AFC or SIAM YN task, though not significantly so [$F(2, 36) = 1.282, p = .29$].

The grand means, averaged across all 10 blocks, for the 2AFC and the SIAM tasks are consistent with reports in the literature (e.g., Florentine, Fastl, & Buus, 1988), with Shepherd and Hautus (2009) reporting thresholds for 1000-Hz 10-ms tones ranging from 14.5 to 23.6 dB SPL. Inspection of Fig. 3 reveals that, over the 10 experimental blocks, there is considerable agreement between the absolute thresholds estimated using the 2AFC and SIAM YN tasks. As anticipated, the mean SIAM YN threshold was marginally higher than the mean 2AFC threshold, though the difference was slight (0.12 dB) and not statistically significant. However, the SIAM-rapid task produced mean thresholds greater than both the 2AFC

(23.87 vs. 22.7 dB) and SIAM YN (23.87 vs. 22.98 dB) tasks, and though these differences were not significant, an explanation is warranted. When faced with ambiguous information during a 2AFC or SIAM YN trial, the participant is not constrained temporally to make a judgment. However, such a constraint does occur in the SIAM-rapid task, so there is the possibility that the procedure will extract a decision of “tone absent” when the participant might, having been granted more time, have responded differently.

The concordance of the thresholds derived using the 2AFC task and the SIAM YN task can be further scrutinised by calculating, for each individual, the absolute differences between the thresholds estimated in both tasks. With reference to Table 3, the mean absolute difference between the 2AFC and SIAM YN tasks was, across the 19 participants, 2.1 dB. The absolute difference between the 2AFC and the SIAM-rapid thresholds was greater, with a mean of 2.5 dB, while the difference between the two SIAM tasks was 1.47 dB. Scatterplots and Spearman's correlation coefficients further support the higher congruence between the 2AFC and the SIAM YN task (Fig. 4), and based on these results we argue that the continuity of threshold estimates is marginally better between the 2AFC and SIAM YN tasks than between the 2AFC and SIAM-rapid tasks. Gu and Green (1994) compared thresholds for

Fig. 3 Comparison of the 2AFC (open circles), SIAM YN (open squares), and SIAM-rapid (open triangles) tasks across 10 experimental blocks for three performance indicators: (a) absolute threshold, (b) mean block completion time, and (c) mean number of trials used in the calculation of thresholds. Each point is the average of 19 participants, and the horizontal lines represent the grand means across participants and blocks for the 2AFC (long-dashed lines), SIAM YN (solid lines), and SIAM-rapid (short-dashed lines) tasks. The dotted vertical lines denote block boundaries, and whiskers represent 95% confidence intervals

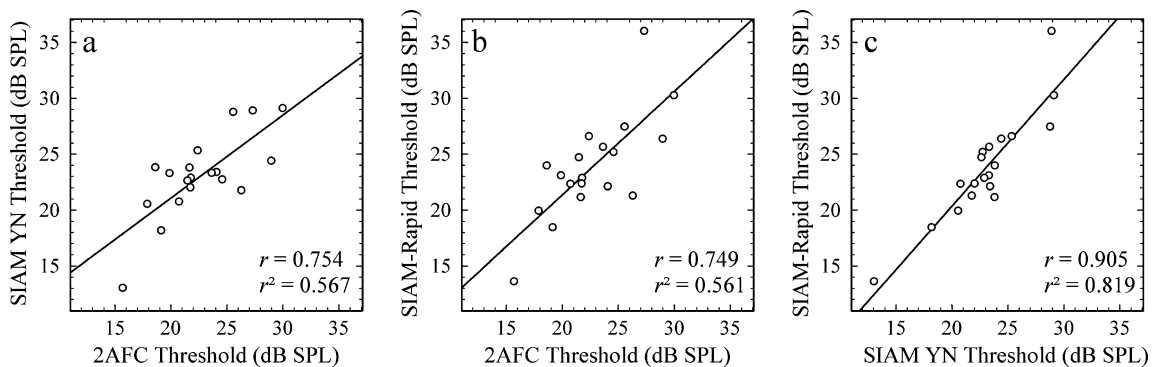
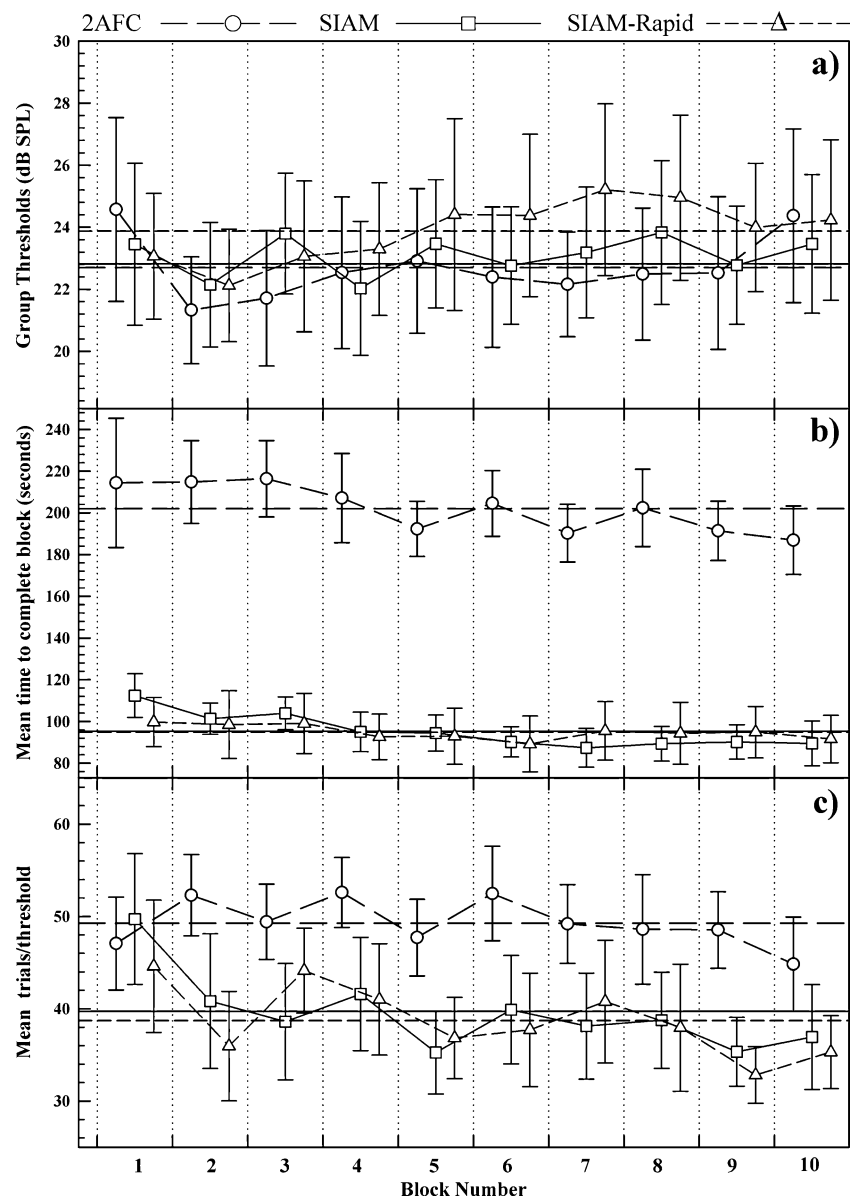


Fig. 4 Scatterplots exhibiting the association between the 2AFC and SIAM YN tasks (a), the 2AFC and SIAM-rapid tasks (b), and the SIAM YN and SIAM-rapid tasks (c). The solid lines represent the best linear least-squares fits, and Pearson’s correlation coefficients (r) are included

1000-Hz tones of 100-ms duration, obtained using either an MLE YN procedure or a three-down one-up 2AFC adaptive procedure. They obtained strong associations (around $r = .9$) between thresholds obtained in the two tasks, generally higher than those reported in this study between the 2AFC and SIAM tasks, which were around $r = .75$ (see Fig. 4). The discrepancy between the two studies is likely explained by the fact that in Gu and Green's study 16 of the 22 participants were highly practiced listeners, rather than the group of inexperienced listeners used in this study.

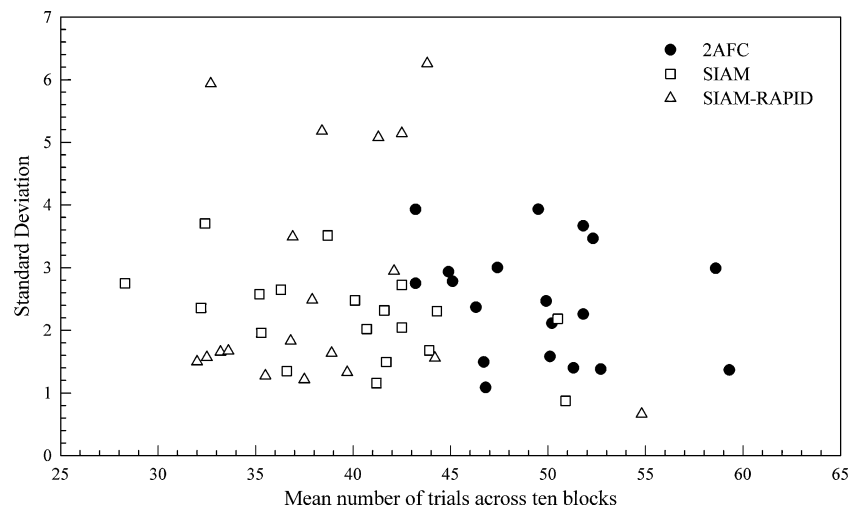
The standard deviations reported in Table 3 index reliability (or, as conceptualised by Taylor, 1971, and Kollmeier, Gilkey, & Sieben, 1988, precision), where an inverse relationship between standard deviation and reliability exists. Inspection of Table 3 reveals that standard deviations averaged across participants are greater in the 2AFC task than in the SIAM YN task, while the SIAM-rapid task, which as we described above suffers from procedural problems, had the largest standard deviation of the three. A repeated measures ANOVA revealed that these differences in standard deviations across the three tasks were not statistically significant [$F(2, 36) = 4.282, p = .286$]. Kershaw (1985) and McKee et al. (1985), reporting the results of a series of computer simulations, found greater variability in the predicted 2AFC staircase estimates than in estimates predicted from the yes–no task. Kaernbach (1990b), also using computer simulations, found the SIAM YN thresholds to be less variable than those from 2AFC staircase procedures, while Gu and Green (1994) reported greater standard deviations for an MLE YN task than a 2AFC three-down one-up task. Block-by-block variability is also of interest, since a systematic decrease can indicate learning or practice effects, while a systematic increase is indicative of fatigue or other methodological issues, all of which compromise reliability (Kollmeier et al., 1988). A repeated measures ANOVA failed to find any effect of block on standard deviations [$F(9, 162) = 2.617, p = .123$], nor was there a task by block interaction [$F(18, 324) = 1.170, p = .284$]. The absence of unwanted effects such as practice or fatigue across the three tasks is evident in the stability of both the central tendency and the variability of the threshold estimates; that is, no decline, abrupt or gradual, across the 10 blocks. We attribute this finding to the comprehensive instructions given to the participants prior to commencing an experimental session. Threshold consistency as a function of block has likewise been reported by Stillman (1989), who demonstrated that stable threshold estimates across blocks and sessions can be obtained from naive participants by providing careful instructions on a block-by-block basis. Thus, we find that, for the auditory modality, the SIAM tasks are equally as reliable as the 2AFC adaptive task.

To gauge efficiencies across the three tasks, further comparisons of group data were undertaken. Figure 3b plots,

for each task and each block, the group mean of the time taken to complete a block of trials, with the mean 2AFC completion times ($M_{2AFC} = 202.07$ s, $SD = 21.77$) higher than those for the SIAM YN ($M_{SIAM} = 95.37$, $SD = 14.21$) and SIAM-rapid ($M_{Rapid} = 94.82$, $SD = 26.23$) tasks. Additionally, the average time to complete a single trial was calculated for each block by dividing the number of trials contained within the block into the time taken to complete the block as a whole (in seconds). Here, the SIAM YN ($M_{SIAM} = 1.32$, $SD = 0.15$) and SIAM-rapid ($M_{Rapid} = 1.31$, $SD = 0.2$) tasks were equivalent, whereas a 2AFC trial ($M_{2AFC} = 2.82$, $SD = 0.16$) takes longer to complete. Kaernbach (1990a) reported trial lengths of 3.2 s for a 2AFC staircase procedure and 2.1 s for the SIAM YN task. The incongruence between our measurements and those of Kaernbach (1990a) can be explained by differences in stimulus duration: Our target stimulus of 10 ms afforded a shorter observation interval than did Kaernbach's 600-ms stimuli. The efficiency gain of the SIAM over the 2AFC is greater in our study than in Kaernbach's (1990a), with approximately two SIAM trials being completed for every one 2AFC trial. Figure 3b also gives the impression that participants increased the speed of their responding across the 10 blocks of trials. However, for both mean block completion time and mean trial completion time, there was no main effect of block, nor a task by block interaction ($p > .05$).

Figure 3c plots, for each task and each block, the average number of trials occurring between the 3rd and 15th turnarounds. These trials are those that contribute to the calculation of the final threshold estimate for a block, and this range of trials was chosen to control for acuity differences between participants. The 2AFC task had the highest mean number of trials ($M_{2AFC} = 49.26$, $SD = 4.67$), followed by the SIAM YN ($M_{SIAM} = 39.71$, $SD = 6.15$) and SIAM-rapid ($M_{Rapid} = 38.7$, $SD = 5.82$) tasks. The difference between the two SIAM tasks and the staircase methods arises from the fact that a decrease in level can occur based on the outcome of a single trial for the SIAM tasks, but requires the combined outcomes of three trials for the 2AFC task. Thus, once the threshold region has been reached, the SIAM procedures are evidently more efficient in estimating threshold than is the 2AFC task, though it must be acknowledged that this difference would diminish had the 2AFC task employed been the more commonly used two-down one-up regime, or variations thereof (e.g., Kaernbach, 1991), so turnarounds are only an approximate index of efficiency. Kollmeier et al. (1988) suggested that efficiency is better assessed graphically by plotting the accuracy of the estimate (i.e., the standard deviations of multiple threshold estimates; see Table 3) as a function of effort (i.e., number of trials) than by using a single numerical index. Figure 5 demonstrates that while the 2AFC task and the SIAM YN task have equivalent standard

Fig. 5 Standard deviations of threshold estimates plotted as a function of mean trial number for 19 participants. Filled circles represent the 2AFC task, while open symbols represent the SIAM YN (squares) and SIAM-rapid (triangles) tasks



deviations, the latter is associated with fewer trials. The same relationship holds between the SIAM YN and SIAM-rapid tasks, though the latter is associated with some alarmingly high standard deviations.

Finally, a comparison of the orthodox SIAM YN task and the SIAM-rapid task indicates that there is no benefit in incorporating the method of free response into the SIAM YN procedure. The finding that both SIAM tasks are equivalent in terms of block completion and trial completion times, and given the higher variability of estimates evident in the SIAM-rapid task, leads us to recommend the orthodox version as originally described by Kaernbach (1990a, 1990b). From Fig. 3a, it is clear that the group threshold estimates obtained in the two tasks are not equivalent, and even though the difference is not statistically significant, there are procedural explanations as to why it occurred. Furthermore, even though the correlation between individual thresholds derived from the two tasks is high (see Fig. 4), this does not excuse the potential inaccuracies that may arise from these procedural issues.

Experiment 2: aqueous sucrose detection

Of all the sensory modalities, the chemical senses (i.e., gustation, olfaction) are arguably the most challenging to assess, both in terms of controlling the proximal stimulus and managing the capacity of the participant to receive the stimulus. The quantity of the stimulus and the frequency of sample presentations must be carefully considered, since participants become rapidly fatigued or satiated during tasting sessions. For these reasons, it is common to fix the number of stimulus presentations within a block, rather than let them vary until some stopping rule has been satisfied. Methods that provide the most accurate estimates of sensory performance for the least number of stimulus presentations are therefore of great utility. Experiment 2

applies sucrose in solution to assess the SIAM YN task relative to a 2AFC adaptive procedure. A key difference between Experiments 1 and 2 is that, because of the limited number of trials that can be credibly undertaken using stimuli of this sort, a 2AFC two-down one-up procedure was used in place of the three-down one-up procedure. This modification allowed for a sufficient number of turnarounds to be obtained from the limited number of trials to calculate absolute thresholds.

Method

Participants

A group of 23 undergraduate students, 5 males and 18 females ($M = 21.4$ years, $SD = 2.19$), volunteered to participate in Experiment 2, none of whom had participated in Experiment 1. Participants had no previous experience with gustatory detection tasks, and all reported good health and no past history of oral disease. Participants were instructed to refrain from eating and drinking (excluding water) at least 1 h before the commencement of testing. Testing involved four sessions, one week apart, each comprising one block of 2AFC and one block of SIAM trials.

Materials

Stimuli For both types of experimental procedures (i.e., 2AFC and SIAM YN), the same set of stimuli was utilised. Within the set, stimuli were categorised as “blank” (i.e., water alone) or “target” (i.e., water + sucrose). Target stimuli, measured in units of molar concentration (M), consisted of 10 sucrose solutions varying in concentration, ranging from -3.25 to -1.00 log M (i.e., ~ 0.00055 to 0.1 M), in -0.25 log M increments. All stimuli were prepared the day before testing, refrigerated until required, and then equilibrated to room temperature.

Other materials The equipment consisted of 10 laboratory-grade beakers containing 10 concentrations of laboratory-grade sucrose, and 10 measuring spoons, 1 for each beaker of solution. An additional beaker contained filtered water. Stimuli were presented to the participant in small plastic disposable cups. The participant had a spittoon for expelling solution (i.e., the sip-and-spit method) and a glass of water for rinsing between trials (SIAM YN and 2AFC tasks) and/or stimuli (2AFC only). The participants were also provided with laminated mats containing trial-by-trial instructions on the order of stimulus tasting, when to rinse, and their response options. A wooden partition meant that the experimenter and the participant were out of sight of each other, but could still communicate verbally when needed.

Procedure

2AFC detection task In the 2AFC adaptive task, each trial consisted of two observation intervals, with a target (sucrose solution) and a blank (water) being randomly assigned to one or the other of the intervals with equal probability ($p = .5$). The participant was required to judge whether the target was presented in the first or the second interval. The sucrose level of following trials was determined according to the participant's response history. If a participant made an incorrect response, the concentration was increased by one step; two correct responses in series resulted in the concentration being decreased one step. The initial sucrose level was $-2.66 \log M$. The 2AFC task began with a warning interval during which the participant rinsed with water. Then the first sample (first interval) and the second sample (second interval) were presented for consumption. Participants were instructed to rinse their mouth with water between samples and, after consuming the second sample, to verbally report their decision to the experimenter. Finally, the experimenter provided feedback to participants: either correct or incorrect. This process was repeated until 30 trials elapsed, and the threshold was the average sucrose level for all but the first two (or three, if there was an even number) turnarounds.

SIAM YN task In the SIAM YN task, the sole observation interval contained either the sucrose solution (i.e., the target) or water (i.e., the blank), with the participant indicating which of the two the observation interval contained. Each SIAM block consisted of 60 trials, half of them containing blank samples. The blank samples were randomly distributed within the block. As with the 2AFC task, the starting sucrose level was $-2.66 \log M$, with subsequent levels calculated by the experimenter according to the participant's response history. As with the SIAM YN

task described for [Experiment 1](#), target performance (t) was fixed to .5. Thus, if a participant was presented a target and he/she responded "Yes" (a hit), the concentration level would be decreased by $-0.25 \log M$, whereas a "no" response (a miss) would increase the level by $-0.25 \log M$. As with the 2AFC staircase task, the SIAM YN task began with a warning interval consisting of the experimenter instructing the participant to rinse with water. Next, the participant was given the sample to consume, and then asked to judge whether or not the sample was the target. Finally, the experimenter provided feedback in the form of "correct" or "incorrect." This sequence of events repeated until 60 trials had elapsed, which took an amount of time equivalent to completing 30 2AFC trials. Block thresholds were calculated using all but the first two (or three, if there was an even number) turnarounds. The order of the SIAM YN and 2AFC tasks was counterbalanced across participants and sessions.

Results and discussion

Figure 6 presents absolute threshold values averaged across participants for four blocks of trials, with 3 of the 4 blocks having higher SIAM YN thresholds relative to the staircase thresholds. To ascertain whether the mean thresholds for the 2AFC ($M_{2AFC} = -2.25$, $SD = .331$) and SIAM YN ($M_{SIAM} = -2.17$, $SD = .346$) tasks differed, a repeated measures ANOVA (sphericity assumed) was undertaken with Task (two levels) and Block (four levels) as within-group factors. No main effects of task [$F(1, 22) = 2.791$, $p = .109$] or block [$F(3, 66) = .931$, $p = .431$] were found, nor was a task by block interaction [$F(3, 66) = 1.192$, $p = .320$]. Additionally, for each task across the 10 experimental blocks, all linear and higher-order polynomial contrasts were not significant ($p > .05$).

Table 4, presenting individual mean thresholds, exhibits both the within- and between-groups variability associated with gustatory testing. Sucrose thresholds reported in the literature range from -2.8 to $-1.22 \log M$, and our 2AFC ($M_{2AFC} = -2.25$) and SIAM YN ($M_{SIAM} = -2.17$) thresholds fall into this range. For a paired-comparison forced choice task, James, Laing, and Oram (1997) reported sucrose thresholds of $-2.75 \log M$ for males and $-2.10 \log M$ for females. Kunka, Doty, and Settle (1981), using a two-down one-up 2AFC adaptive procedure, reported mean sucrose thresholds of $-2.8 \log M$, while Shepherd, Quek and Pathirana (2008) reported thresholds estimated in professionally designed sensory booths ($-2.8 \log M$) and standard laboratory tables ($-2.39 \log M$). Fukunaga, Uematsu, and Sugimoto (2005), using Asian tasters, obtained a mean threshold of $-1.77 \log M$. Hautus et al. (2010), in an unrelated experiment, reported sucrose thresholds of $-2.16 \log M$ for

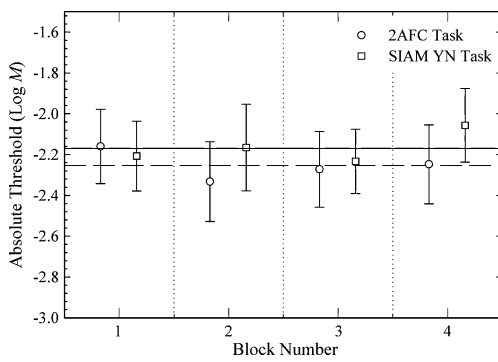


Fig. 6 Absolute threshold (log M) as a function of block number for thresholds estimated using a 2AFC staircase (open circles) or SIAM YN (open squares) task. The horizontal lines represent the grand means across participants and blocks for the 2AFC (dashed line) and SIAM YN (solid line) tasks, and the vertical bars are 95% confidence intervals

the SIAM YN task, this being the first demonstration of the SIAM procedure in the gustatory modality. Thus, the SIAM YN data we report here demonstrate convergent validity with preexisting threshold measures.

Table 4 Mean absolute thresholds (log M) and mean turnarounds per block arranged by task (2AFC or SIAM YN) for each participant

Taster	Sucrose Thresholds					Turnarounds per Block				
	2AFC Task		SIAM Task		Difference	2AFC Task		SIAM Task		Difference
	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
1	-2.09	0.23	-1.88	0.42	-0.21	8.25	2.22	19.75	3.10	-11.5
2	-2.43	0.57	-2.69	0.10	0.26	12.00	2.94	20.50	1.29	-8.5
3	-2.40	0.29	-2.29	0.36	-0.11	9.25	1.89	15.75	1.71	-6.5
4	-2.57	0.27	-2.40	0.30	-0.17	9.00	1.63	22.25	3.86	-13.25
5	-2.19	0.36	-2.42	0.36	0.23	7.75	1.50	17.25	2.75	-9.5
6	-2.09	0.60	-1.96	0.38	-0.13	10.00	0.82	21.00	2.16	-11
7	-2.50	0.12	-2.17	0.37	-0.33	5.50	2.38	14.00	2.71	-8.5
8	-1.19	0.07	-1.47	0.27	0.28	9.25	2.63	20.00	2.16	-10.75
9	-2.67	0.15	-2.72	0.32	0.05	9.50	1.00	17.25	6.70	-7.75
10	-2.11	0.60	-2.36	0.27	0.25	9.00	2.45	19.75	1.71	-10.75
11	-1.80	0.45	-1.53	0.24	-0.27	10.75	0.96	21.00	2.58	-10.25
12	-2.42	0.17	-2.22	0.22	-0.2	8.25	1.50	20.25	3.69	-12
13	-2.63	0.21	-2.33	0.26	-0.3	11.00	1.83	19.75	1.71	-8.75
14	-2.36	0.27	-2.44	0.27	0.08	11.75	1.89	23.50	1.91	-11.75
15	-2.39	0.48	-2.21	0.26	-0.18	9.00	2.16	17.75	4.03	-8.75
16	-2.03	0.31	-1.67	0.46	-0.36	9.50	3.42	23.25	2.22	-13.75
17	-2.12	0.05	-2.18	0.12	0.06	8.75	2.06	20.25	4.79	-11.5
18	-2.62	0.22	-1.85	0.31	-0.77	8.25	0.96	17.50	3.70	-9.25
19	-2.33	0.08	-2.32	0.42	-0.01	8.75	1.71	20.75	3.30	-12
20	-2.54	0.30	-2.43	0.51	-0.11	10.25	3.30	20.50	3.70	-10.25
21	-1.91	0.41	-1.64	0.20	-0.27	9.75	2.75	17.25	3.40	-7.5
22	-2.27	0.49	-2.22	0.31	-0.05	9.75	1.26	17.25	2.06	-7.5
23	-2.15	0.72	-2.39	0.37	0.24	10.75	2.75	19.00	1.41	-8.25
Mean	-2.25	0.32	-2.17	0.31	-0.08	9.39	2.00	19.37	2.90	-9.98
SD	0.33	0.19	0.35	0.10	0.02	1.40	0.75	2.32	1.26	1.92

Standard deviations (SD) are included, as are the differences between the means associated with each task (i.e., Difference = $M_{2AFC} - M_{SIAM}$)

Table 4 also indicates that the SIAM YN task has lower levels of variability compared to the 2AFC task. The mean standard deviation across participants was significantly lower [$F(1, 22) = 8.978, p = .007$] for the SIAM YN task ($M_{SIAM} = .10$) than for the 2AFC task ($M_{2AFC} = .19$). This indicates that threshold measures are more reliable when using the SIAM YN task. Additionally, there was no main effect of block [$F(3, 66) = .537, p = .658$] on the magnitude of the standard deviations, nor was there a task by block interaction [$F(3, 66) = 1.508, p = .221$]. Furthermore, scrutiny of Fig. 6 fails to marshal evidence of learning, practice, or fatigue effects for either task. As with Experiment 1, the participants were given thorough instructions prior to commencing a block of trials and were issued an instruction sheet that could be consulted on a trial-by-trial basis. The implication for both tasks is that naive participants can adapt to tasks rapidly if properly instructed (Stillman, 1989), thus negating the need for laborious training regimes.

The difference in the numbers of turnarounds per block between the two tasks is also of interest, given the constraints on the number of trials when using gustatory

stimuli. Taylor and Creelman (1967) asserted that threshold variability decreases as more trials or turnarounds are used to compute it. The numbers of turnarounds per block (Table 4) differed significantly [$t(22) = 24.528, p < .001$] across the two tasks, even though the time taken to complete a block of either task was approximately the same. Keeping in mind that the values presented in Table 4 include three turnarounds that were subsequently excluded from the calculation of threshold, and that the number of turnarounds in the 2AFC task largely resides in the lower end of Macmillan and Creelman's (2005) suggested range for adaptive procedures (i.e., 5–10 turnarounds). On average, twice as many turnarounds were obtained using the SIAM YN than using the 2AFC task, and with more turnarounds resulting in less measurement error, this is a positive finding for the SIAM YN task. Hautus et al. (2010), also measuring sucrose thresholds with the SIAM YN task, reported an average of 8.5 turnarounds for 30 trials across three blocks and 15 participants, and our finding of an average of 19.4 turnarounds across 60 trials is comparable.

General discussion

Our assessment of the SIAM in both auditory and gustatory contexts suggests that the SIAM procedure can produce valid and reliable threshold estimates. Specifically, we demonstrated convergent validity between the auditory and gustatory thresholds obtained from SIAM procedures and a gold standard (i.e., 2AFC adaptive procedures), as well as results comparable to those reported in the literature. Additionally, the standard deviations calculated from 10 (auditory) or 4 (gustatory) blocks of trials indicated that the SIAM estimates were as reliable as those obtained in the 2AFC adaptive tasks. The SIAM procedures were also easy to implement, in both the auditory and gustatory modalities, and we failed to note any specific procedural difficulties from either an experimenter's or a participant's perspective. The final criterion raised in the introduction relates to task efficiency. Our data overwhelmingly demonstrated that thresholds could be obtained with fewer trials and in quicker time using the SIAM tasks than when a 2AFC staircase procedure was employed.

The SIAM YN task offers an alternative method with which to estimate absolute thresholds. Its merits lie in the use of trials containing a single interval, which enhances efficiency, and its claim to be a procedure uncontaminated by response bias, which improves validity. How then does the SIAM YN task compare to other single-interval procedures currently available? The more popular MLE YN procedure (Green, 1990, 1993) is the only single-interval adaptive procedure currently affording meaningful

comparison to the SIAM YN task. It has been argued that the MLE YN task is more efficient than the SIAM (Green, 1993); however, after showing early promise, the merits of the procedure have increasingly been challenged, with findings questioning the assumptions, efficiency, and mechanics of the MLE approach. The MLE YN task mostly assumes an underlying psychometric function based on the logistic function (Green, 1990). This assumption has not been found to be robust (e.g., Lecluyse & Meddis, 2009), though Gu and Green (1994) argued this finding away, claiming that because the threshold estimates produced by the procedure are consistent with others reported in the literature, it is a moot point whether these assumptions are upheld. However, Lecluyse and Meddis reported that psychometric functions do differ across clinical groups, and a universal form should not be assumed; this stance was supported by Leek (2001). Amitay, Irwin, Hawkey, Cowan and Moore (2006) reported that with novice participants, the form of the underlying psychometric function is prone to change due to learning effects. However, it should be noted that the key assumptions underlying the SIAM approach remain to be tested, and furthermore that the testing of these assumptions provides a stern methodological challenge (Green, 1993).

In relation to efficiency, Green (1993) initially proposed that reliable estimates of auditory threshold could be collected in 8 trials using the MLE YN task. However, Amitay et al. (2006), reporting on naive participants using a two-interval MLE procedure in an auditory study, argued that a minimum of 30 trials are needed due to intra- and interblock learning effects. Leek (2001) recommended 24 MLE YN trials, of which 8 should be designated catch trials. The use of catch trials introduces significant inefficiencies, as they have no impact on stimulus levels or threshold estimates, and if a participant is "caught out," this may result in the block being abandoned and restarted, again compromising efficiency.

A final consideration about the MLE YN task is the reliability and validity of its threshold estimates. It soon became apparent after Green's (1990) initial description of the MLE YN task that threshold estimates could be significantly different from those reported in the literature (Green, 1993). Since then, it has been consistently demonstrated (e.g., Baker & Rosen, 2001; Green, 1995; Gu & Green, 1994; Lecluyse & Meddis, 2009; Leek, 2001) that the MLE YN task is extremely unforgiving of errors that occur early in a track, especially in the first five trials (Gu & Green, 1994). Such mistakes are more likely to be made with naive participants than experienced ones, and because of this issue the MLE YN task cannot be relied on to give estimates of true threshold with untrained participants. These misleading estimates were modelled by Green (1995) as lapses in attention, and he concluded that the

MLE YN task can severely underestimate or overestimate true thresholds, depending on when the attentional lapses occur. However, Lecluyse and Meddis (2009) demonstrated that these erroneous threshold estimates can be explained by processes intrinsic to the MLE YN task itself, rather than by factors associated with the participant. Though Gu and Green offer a correction for unintended errors, the number of trials necessary to use the correction effectively eliminates any efficiency gains and introduces additional complexity into the procedure.

Conclusion

In summary, our findings indicate that the SIAM YN task can provide estimates of absolute threshold equivalent to those obtained by “gold standard” procedures or those reported in the literature. Both auditory and gustatory experiments have demonstrated the efficiency and accuracy of the SIAM approach, and from our data, we are confident in recommending Kaernbach’s (1990b) SIAM approach to estimate absolute thresholds. Additionally, the SIAM YN task may emerge as a viable alternative to the MLE YN task, which is vulnerable to biased threshold estimates. These results also suggest that the SIAM YN task can be applied in the clinical setting, where the use of two- (or more) interval procedures is not always practical, and the increase in redundant presentation time (i.e., producing information equivalent to a single-interval task, but with increased trials) undesirable. The forced choice procedure is often unpopular with patients because they are reluctant to guess if they fail to detect a signal in any of the intervals (Green, 1993; Kaernbach, 2001). A single-interval task avoids this difficulty, as patients report only whether (or not) they detected the target stimulus. Thus, our data support the use of the SIAM YN task as a potentially fast, reliable, and accurate procedure for the measurement of absolute thresholds in experimental and clinical contexts.

Author Note Parts of this study were presented at the 25th Annual Meeting of the International Society for Psychophysics (2009), Galway, Ireland.

References

- Amitay, S., Irwin, A., Hawkey, D. J. C., Cowan, J. A., & Moore, D. R. (2006). A comparison of adaptive procedures for rapid and reliable threshold assessment and training in naive listeners. *The Journal of the Acoustical Society of America*, *119*, 1616–1625.
- Baker, R. J., & Rosen, S. (2001). Evaluation of maximum-likelihood threshold estimation with tone-in-noise masking. *British Journal of Audiology*, *35*, 43–52.
- Brooks, J., Wong, T., & Robertson, L. (2005). Crossing the midline: Reducing attentional deficits via interhemispheric interactions. *Neuropsychologia*, *43*, 572–582.
- Brown, L. G., & Rudd, M. E. (1998). Evidence for a noise gain control mechanism in human vision. *Vision Research*, *38*, 1925–1933.
- Cornsweet, T. N. (1962). The staircase-method in psychophysics. *The American Journal of Psychology*, *75*, 485–491.
- Egan, J. P., Greenber, G. Z., & Schulman, A. I. (1961). Operating characteristics, signal detectability, and the method of free response. *The Journal of the Acoustical Society of America*, *33*, 993–1007.
- Florentine, M., Fastl, H., & Buus, S. (1988). Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking. *The Journal of the Acoustical Society of America*, *84*, 195–203.
- Fukunaga, A., Uematsu, H., & Sugimoto, K. (2005). Influences of aging on taste perception and oral somatic sensation. *Journals of Gerontology*, *60A*, P109–P113.
- Green, D. M. (1964). General prediction relating yes–no and forced-choice results. *The Journal of the Acoustical Society of America*, *36*, 1042.
- Green, D. M. (1990). Stimulus selection on adaptive psychophysical procedures. *The 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. *The Journal of the Acoustical Society of America*, *93*, 2096–2105.
- Green, D. M. (1995). Maximum-likelihood procedures and the inattentive observer. *The Journal of the Acoustical Society of America*, *97*, 3749–3760.
- Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics*. New York: Wiley. Original work published 1966.
- Gu, X., & Green, D. M. (1994). Further studies of a maximum-likelihood yes–no procedure. *The Journal of the Acoustical Society of America*, *96*, 93–101.
- Hautus, M. J., Stocks, M., & Shepherd, D. (2010). The single interval adjustment matrix (SIAM) yes–no task applied to the measurement of sucrose thresholds. *Journal of Sensory Studies*, *25*, 940–955.
- Hesse, A. (1986). Comparison of several psychophysical procedures with respect to threshold estimates, reproducibility & efficiency. *Acustica*, *59*, 263–266.
- James, C. E., Laing, D. G., & Oram, N. (1997). A comparison of the ability of 8–9-year-old children and adults to detect taste stimuli. *Physiology & Behavior*, *62*, 193–197.
- Kaernbach, C. (1990a). SIAM: Adaptive yes–no task testing with the single-interval adjustment matrix. In F. Müller (Ed.), *Proceedings of the sixth annual meeting of the international society for psychophysics* (pp. 88–93). Würzburg: International Society for Psychophysics.
- Kaernbach, C. (1990b). A single-interval adjustment-matrix (SIAM) procedure for unbiased adaptive testing. *The Journal of the Acoustical Society of America*, *88*, 2645–2655.
- Kaernbach, C. (1991). Simple adaptive testing with the weighted up–down method. *Perception & Psychophysics*, *49*, 227–229.
- Kaernbach, C. (2001). Adaptive threshold estimation with unforced-choice tasks. *Perception & Psychophysics*, *63*, 1377–1388.
- Kershaw, C. D. (1985). Statistical properties of staircase estimates from two interval forced choice experiments. *The British Journal of Mathematical and Statistical Psychology*, *38*, 35–43.
- Klein, S. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics*, *63*, 1421–1455.
- Kollmeier, B., Gilkey, R. H., & Sieben, U. K. (1988). Adaptive staircase techniques in psychoacoustics: A comparison of human data and a mathematical model. *The Journal of the Acoustical Society of America*, *83*, 1852–1862.

- Kunka, M., Doty, R. L., & Settle, R. G. (1981). An examination of intertrial interval and gender influences on sucrose detection thresholds established by a modified staircase procedure. *Perception, 10*, 35–38.
- Lecluyse, W., & Meddis, R. (2009). A simple single-interval adaptive procedure for estimating thresholds in normal and impaired listeners. *The Journal of the Acoustical Society of America, 126*, 2570–2579.
- Leek, M. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics, 63*, 1279–1292.
- Leeuw, A. R., & Dreschler, W. A. (1998). The relation between otoacoustic emissions and the broadening of the auditory filter for higher levels. *Hearing Research, 126*, 1–10.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America, 49*, 467–477.
- List, A., Brooks, J. K., Esterman, M., Flevaris, A. V., Landau, A. N., Bowman, G., Stanton, V., Vanvleet, T.M., Robertson, L.C., & Schendel, K. (2008). Visual hemispatial neglect, re-assessed. *Journal of the International Neuropsychological Society, 14*, 243–256.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Hillsdale: Erlbaum.
- Madigan, R., & Williams, D. (1987). Maximum-likelihood psychometric procedures in two-alternative forced-choice: Evaluation and recommendations. *Perception & Psychophysics, 42*, 240–249.
- McKee, S. P., Klein, S. A., & Teller, D. Y. (1985). Statistical properties of forced-choice psychometric functions: Implications of probit analysis. *Perception & Psychophysics, 37*, 286–298.
- Shepherd, D., & Hautus, M. J. (2009). Negative masking and the units problem in audition. *Hearing Research, 247*, 60–70.
- Shepherd, D., Quek, S. Y., & Pathirana, S. (2008). Sucrose detection and the stability of the 2-AFC procedure in the presence of a confounding variable. *Journal of Sensory Studies, 23*, 600–613.
- Stillman, J. A. (1989). A comparison of three adaptive psychophysical procedures using inexperienced listeners. *Perception & Psychophysics, 46*, 345–350.
- Taylor, M. M., & Creelman, C. D. (1967). PEST: Efficient Estimates on Probability Functions. *The Journal of the Acoustical Society of America, 41*, 782 - 787
- Taylor, M. M. (1971). On the efficiency of psychophysical measurement. *The Journal of the Acoustical Society of America, 49*, 505–508.
- Taylor, M. M., Forbes, S. M., & Creelman, C. D. (1983). PEST reduces bias in forced choice psychophysics. *The Journal of the Acoustical Society of America, 74*, 1367–1374.
- Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research, 35*, 2503–2522.
- Ulrich, R., & Vorberg, D. (2009). Estimating the difference limen in 2AFC tasks: Pitfalls and improved estimators. *Attention, Perception, & Psychophysics, 71*, 1219–1227.