A comparison of adaptive psychometric procedures based on the theory of optimal experiments and Bayesian techniques: Implications for cochlear implant testing

JEREMIAH J. REMUS AND LESLIE M. COLLINS
Duke University, Durham, North Carolina

Numerous previous studies have focused on the development of quick and efficient adaptive psychometric procedures. In psychophysics, there is often a model of the psychometric function supported by previous studies for the task of interest. The theory of optimal experiments provides a framework for utilizing a model of the process to develop quick and efficient sequential-testing strategies for estimating model parameters, making it appropriate for developing adaptive psychophysical-testing methods. In this study, we investigated the application of sequential parameter search strategies based on the theory of optimal experiments and Bayesian adaptive procedures for measuring psychophysical variables. The results presented in this article suggest that more sophisticated psychometric procedures can expedite the measurement of psychophysical variables. Such techniques for quickly collecting psychophysical data may be particularly useful in cochlear implant research, where a large set of psychophysical variables are useful for characterizing the performance of an implanted device. It is to be hoped that further development of these techniques will make psychophysical measurements available to clinicians for tuning and optimizing the speech processors of individual cochlear implant patients.

Development of more efficient psychometric procedures has been the focus of numerous published studies. Many of these techniques are adaptive, utilizing the outcomes of previous trials to determine the next step in the experiment. Pelli (1987) proposed the ideal psychometric procedure, looking ahead over every possible trial sequence for the length of the experiment to maximize confidence in the outcome. Several psychometric procedures—for example, QUEST (Watson & Pelli, 1983), ZEST (King-Smith, Grigsby, Vingrys, Benes, & Supowit, 1994), PSI (Kontsevich & Tyler, 1999)—utilize a Bayesian formulation to include the outcomes of previous trials. Studies have suggested that these techniques are highly efficient for measuring psychophysical variables.

One particular application for which efficient psychometric procedures are relevant is cochlear implant research. Psychophysical variables are the primary metric for gauging the operation and performance of a cochlear implant *in vivo*, since more invasive procedures are not feasible. However, a comprehensive assessment of cochlear implant performance requires collecting large amounts of psychophysical data in a time-consuming process. Hence, the motivation for developing more efficient psychometric procedures to rapidly collect cochlear implant psychophysical data is particularly strong, but any efficient procedures have potential applications in a variety of fields that utilize psychophysical data.

A potential alternative framework to those cited above for developing quick and efficient psychometric methods is the theory of optimal experiments. Federov (1972) and Chernoff (1972) have written extensively about the theory of optimal experiments, and it has been developed fully by many others (MacKay, 1992; Whaite & Ferrie, 1997). The theory of optimal experiments has been extended to a number of fields with applications involving estimating parameters through sequential trials. On the basis of successes in other fields, the theory of optimal experiments could serve as a promising basis for an adaptive psychometric procedure, but it has not been investigated as such to date.

The theory of optimal experiments requires a model of the process under investigation. The parameters of the model can be estimated iteratively on the basis of the outcomes of sequential trials through a two-stage process. First, the parameters for the model are estimated to best fit the observed data. Then a trial is conducted at the stimulus value that will provide the most information about the model, using the most recent estimate of the parameter values. The two processes controlled by the investigator are how to estimate the parameters from the observed data and what measure of information to use for selecting the stimulus for the next trial. This is a fairly general strategy for testing and parameter estimation; many techniques that are not directly derived from the theory of optimal ex-

L. M. Collins, lcollins@ee.duke.edu

periments can be considered to follow the same two-stage process. However, by associating them with the theory of optimal experiments, the current and proposed techniques follow the same strategy, allowing easier and more informative comparisons.

The psychophysical and psychoacoustic measurement task is an appropriate application for the theory of optimal experiments, due to the existence of a model of the psychometric function. Extensive previous data provide strong support for a known form of the psychometric function. Since this model of the psychometric function is known, it can be exploited by methods based on the theory of optimal experiments. Because it incorporates knowledge about the model of the psychometric function into the sequential-testing procedure, the theory of optimal experiments may have an advantage over previously investigated adaptive psychometric techniques that do not leverage this information.

The psychometric function is modeled in this study using the logistic cumulative distribution function (cdf), which has the form shown in Equation 1:

$$p(r; \lambda, x) = \frac{1}{1 + e^{-s(x-m)}}. \quad (1)$$

In Equation 1, r is the probability of successful response ($r = 1$), x is the stimulus intensity, and $\lambda = \{m, s\}$ is the parameter set containing the “threshold” or median m of the underlying logistic probability distribution function (pdf) and the parameter s controlling the slope of the psychometric function. The logistic cdf was selected as the model for the psychometric function since the logistic cdf equation is easy to manipulate for the various mathematical measures applied in the theory of optimal experiments. Psychometric functions in simulations testing psychometric procedures are often represented using integrated probability distributions. Other frequently used cdfs include the Weibull cdf and the Gaussian cdf (Klein, 2001). Each of these distributions contains two parameters: a distribution median parameter that controls the threshold in the psychometric function and a distribution variance parameter that controls the slope of the psychometric function. Estimating two parameters of the psychometric function introduces additional difficulties for previously developed Bayesian adaptive psychometric procedures. Whereas the threshold parameter is directly observable (i.e., the stimulus values and threshold are on the same parameter scale), this is not true for the slope parameter. Thus, a psychometric procedure dedicated to measuring multiple parameters is desired. The psychometric function chosen in this study requires two parameters; however, extension to tasks in which additional parameters are estimated is a possible application of the techniques to be considered.

One procedure capable of measuring both the slope and the threshold of the psychometric function is the PSI method. The PSI method, developed by Kontsevich and Tyler (1999), is based on a theory similar to that for the ZEST method, with a 2D parameter space containing a slope parameter and a threshold parameter that are updated after each trial, using the mean likelihood estimates. Levitt (1971) showed that conducting trials at the threshold value provides the minimum variance estimate of the threshold.

However, when the slope of the psychometric function is considered simultaneously, a modified criterion for selecting the next trial stimulus is necessary. Repeated trials at the threshold do not provide sufficient information for identifying the slope; indeed, Levitt again indicates that the stimulus values for minimum variance estimates of the slope of the psychometric function are at 6% and 94% (for the *yes/no* task). However, these points cannot be identified without knowledge of the entire psychometric function, which would require knowing the slope, as well as the threshold. Therefore, the issue of where to place the stimuli to provide data from which the slope can be estimated requires a more sophisticated solution. The PSI method selects the next stimulus value resulting in the most significant expected reduction in uncertainty about both parameters (threshold and slope) in the parameter space. Stimulus values are selected adaptively on the basis of the locations and outcomes of previous trials, and the parameters of the psychometric function are estimated quickly and efficiently.

As an alternative to the PSI method, adaptive psychometric procedures based on the theory of optimal experiments might provide an advantage, in terms of performance, by using a different criterion for selecting the stimulus values. The PSI method does not take into account the model of the psychometric function when selecting the stimulus values. If the model of the psychometric function is accurate, it may allow a psychometric procedure based on the theory of optimal experiments to select more informative stimulus values.

In this study, four adaptive psychometric procedures for estimating multiple parameters are considered within the framework of the theory of optimal experiments. The investigated psychometric procedures consist of two stages: estimation of parameters from the previous trials and selection of the next stimulus value, using the most recent estimates of the parameter values. Two techniques for each of the stages, one based on the PSI method and one proposed from previous implementations of the theory of optimal experiments, were considered in this study and were evaluated through simulations of psychophysical testing, using the logistic cdf to model the psychometric function. These psychometric procedures will be compared with other adaptive psychometric procedures commonly used in psychophysical and psychoacoustic research.

PSYCHOMETRIC METHODS

The PSI method (Kontsevich & Tyler, 1999) is a Bayesian adaptive technique that estimates the parameter values from posterior distributions calculated at points in a 2-D parameter space. The PSI method uses the probabilities assigned to the points in the 2-D parameter space and the probability of each response at all possible stimuli intensities to calculate the expected entropy after the next trial. The next stimulus level is selected to minimize the calculated entropy in the probability distribution of the slope and threshold parameter values. Minimizing the expected entropy produces a probability distribution of parameter values with high certainty in a small set of values, which are then used to calculate the mean parameter estimate.

The second sequential-testing procedure serving as a foundation for the investigated psychometric procedures is a parameter search strategy motivated by the theory of optimal experiments that was implemented by Liao and Carin (2004) for electromagnetic induction sensing of buried targets. The method applied by Liao and Carin uses the Levenberg–Marquardt method of gradient descent to search the parameter space for the values that minimize the squared error between the observed data and the model. With the “best” estimates of the parameter values, the stimulus for the next trial is located at the point in the stimulus space that maximizes the gain in Fisher information. Calculations of the gain in Fisher information take into account the assumed model of the underlying psychophysical process, as well as the available information about observed points in the model. The PSI method and the Liao–Carin implementation provide two techniques for each of the two stages in a sequential parameter search strategy for psychometric testing. The parameter values can be estimated using a Bayesian procedure (e.g., the PSI method) or an iterative search procedure (e.g., Levenberg–Marquardt, used by Liao & Carin, 2004). A divide-and-conquer iterative parameter estimation (IPE) technique was used in this study as an alternative to the Levenberg–Marquardt method, due to preliminary investigations using gradient search optimization techniques that indicated some instability arising from the binary nature of observations in psychophysical testing. The information gained at each possible value of the next stimulus, used to select the stimulus value at the next trial, can be quantified by minimizing the expected entropy (e.g., the PSI method) or using the gain in Fisher information (e.g., Liao & Carin, 2004). Each of these procedures for parameter estimation and information gain will be described below.

Parameter Estimation: Bayesian Procedure

The parameters, λ , of the psychometric function can be estimated using a Bayesian procedure. A prior probability distribution, $p(\lambda)$, is generated over an n -dimensional parameter space, where n is the number of psychometric function parameters to be estimated. The model selected for the psychometric function is used to generate likelihood functions for sample points throughout the parameter space. The likelihood functions, $p(r|\lambda, x)$, express the probability of successful response (response $r = 1$) or failure (response $r = 0$) at a certain stimulus intensity x if the psychometric function has the given parameter values λ . Note that $p(r = 0|\lambda, x) = 1 - p(r = 1|\lambda, x)$. The posterior probability distribution of each set of model parameters can then be calculated using Bayes’s rule, shown in Equation 2. This form of Bayes’s rule results when it is assumed that the prior on λ is independent of x —that is, $p(\lambda|x) = p(\lambda)$:

$$p(\lambda|r, x) = \frac{p(r|\lambda, x)p(\lambda)}{\sum_{\lambda} p(r|\lambda, x)p(\lambda)}. \quad (2)$$

After the i th trial is conducted, providing response r_i to the stimulus x_i , the posterior distribution corresponding to the stimulus value and the observed response can be used

to update the prior distribution, according to Equation 3. The updated prior, which now incorporates information from the most recent trial, is used to calculate the model probabilities for subsequent trials:

$$p(\lambda)_{i+1} = p(\lambda|r = r_i, x = x_i). \quad (3)$$

Parameter Estimation: Iterative Search (IPE)

The parameters of the psychometric function can also be estimated using an iterative search of the parameter space to find the values that minimize the error between the observed data and the model of the psychometric function. The n -dimensional parameter space is the same as that used in the Bayesian parameter estimation procedure. This iterative search technique is motivated by the particular error surface encountered when the logistic function with two parameters (threshold and slope) is fit to the observed data. The gradient of the error space along one parameter is nearly constant across the other dimension in the region of error surface minimum. Therefore, from an initial estimate of the parameter values, the parameters can be independently and alternately updated through successive iterations. After locating the set of parameter values resulting in the minimum error at a given grid spacing in the parameter space, the resolution of the grid can be increased, and the process can be continued to refine the parameter estimates. This is continued until increasing the resolution of the grid spacing only negligibly reduces the error.

To implement the iterative search, the initial value for the threshold parameter m_0 is calculated using the Spearman–Kärber method (Ulrich & Miller, 2004). The algorithm then iterates through five steps.

1. Holding m fixed at $m = M$, evaluate the weighted error between the observed data and psychometric function at N points ($S_1, S_2, \dots, S_N$) in the slope parameter dimension.
2. Select from ($S_1, S_2, \dots, S_N$) the slope parameter value $s = S_{\min}$ that minimizes the error, given $m = M$.
3. Holding s fixed, evaluate the weighted error at N points ($M_1, M_2, \dots, M_N$) in the threshold parameter dimension.
4. Update the threshold parameter $m = M_{\min}$. Repeat Steps 1–4 until Step 5 occurs.
5. Once $S_{\min}$ and $M_{\min}$ are found, focus the iterative search on the neighborhood of $S_{\min}$ and $M_{\min}$ by reducing the spacing between the N search points centered around $S_{\min}$ and $M_{\min}$.

The weighted squared error between the observed data and the modeled psychometric function at N equally spaced points in the slope (s) dimension of the parameter space is calculated while the value of the threshold is held fixed. The value of the slope parameter s that minimized the fitting error is then held fixed, the m dimension is sampled at N equally spaced points, and an updated estimate of the threshold that minimizes the fitted error is found. Once the search has focused on an area in the parameter space, the sampling resolution is increased by a factor of four; that is, the distance between the N points is 1/4 the previous spacing. This reduces the search to 1/16 of the

Figure 1. (Top) Observed data and the underlying psychometric function. (Bottom) Error surface for observed data shown in the top panel, including the search path for the psychometric function parameter values found using iterative parameter estimation that best fit the observed data.

previous search space. Since the dimensions are considered individually, only vector operations are necessary (as opposed to matrix operations to consider both parameters simultaneously), and the progressively increasing resolution allows the iterative search to run efficiently.

The motivation for using an iterative search technique, rather than a traditional gradient search technique such as Levenberg–Marquardt, is based on preliminary simulation results. In several situations, the sequence of trials and observed responses produced error surfaces with local minima that were not satisfactory estimates of the psychometric function parameters. Due to the binary outcomes of each trial in psychophysical testing, single observations at points on the psychometric function frequently result in error spaces with many local minima. The error surface is improved by weighting each data point by a confidence based on the number of trials at that point, but the problem is not eliminated. This issue of local minima can be exac-

erbated by certain sets of observed data resulting in initial values for the parameters on the “wrong” side of a saddle point. The iterative search technique is able to mitigate this issue by approaching parameter estimation from an initially high-level perspective and progressively focusing on an area in the error space, as opposed to the gradient search technique that starts at a point and searches outward.

Figure 1 illustrates how the iterative parameter search is conducted. At the top, the collected data are shown, along with the psychometric function truth with unknown parameter values. The collected data, shown in Figure 1 as $\times$, provide measurements of the psychometric function calculated from grouped trials at each sample point. The resolution of these measurements of the psychometric function is based on the number of trials conducted at each stimulus value. At the bottom is the error space identifying how well the data at the top are fit by the psychometric function, using each set of parameter values in the param-

eter space. This is the error surface that is searched using the iterative technique to find the parameter values that correspond to the minimum error between the observed data and the psychometric function. The initial point in the parameter search is shown on the error surface, as well as the true parameter values corresponding to the psychometric function plotted at the top. The resulting parameter values from the first three updates are shown, and they proceed toward the center of the innermost contour. Each point on the search path is an updated estimate of the parameters and results from independent steps in each of the two search space dimensions. The vertical dashed line identifies the values of the slope parameter tested to minimize the error after initializing of the mean parameter, using the Spearman–Kärber method (Ulrich & Miller, 2004). The value of the slope parameter that minimizes the error is chosen as the new value of the slope parameter. The horizontal line identifies the values of the mean parameter tested after the slope parameter has been updated. This iterative process is continued until it converges on a point in the error space. The grid is then resampled with higher resolution, and the process continues.

Information Gain: Minimum Expected Entropy [min(H)]

One criterion that can be used to measure the information gained by the next stimulus is the expected entropy of $p(\lambda)$, the probability distribution of the psychometric function parameters. Using this criterion, the stimulus value that minimizes the expected entropy provides the maximum gain in information. To find the stimulus value for the next trial that minimizes the expected entropy, the conditional entropy is first calculated. The conditional entropy, $H_r(x)$, is the entropy in the parameter probability distribution that will result after a trial at x produces response r . The conditional entropy is calculated according to Equation 4:

$$H_r(x) = -\sum_{\lambda} p(\lambda | x, r) \log p(\lambda | x, r). \quad (4)$$

After the conditional entropy has been calculated for each of the two possible responses (success and failure), the expected entropy can be calculated from Equation 5:

$$E[H(x)] = H_{r=1}(x)p(s|x) + H_{r=0}(x)p(f|x). \quad (5)$$

The expected entropy is a function of the possible stimulus values, and the next stimulus value, x_{i+1} , is selected as the argument that minimizes the expected entropy:

$$x_{i+1} = \arg \min_x E[H(x)]. \quad (6)$$

Information Gain: Fisher Information Gain (FIG)

The information gained by any possible stimulus value can also be measured using the Cramer–Rao bound and the gain in Fisher information (Cover & Thomas, 1991). The Cramer–Rao bound identifies the minimum variance for the estimates of the psychometric function parameters, assuming an unbiased estimator. The next sample is selected to maximize the Fisher information matrix, which is inversely related to the Cramer–Rao bound.

Maximizing the Fisher information effectively minimizes the Cramer–Rao bound, allowing lower bounds on the variance of the estimates based on the observed data. To explain the calculation of the incremental increase in the Fisher information, some notation will first be introduced. Let $X = \{x_1, x_2, \dots, x_m\}$ be the set of possible stimulus values—that is, the next stimulus $x_{i+1} \in X$ —and $g(\lambda, O)$ be the psychometric function measured at stimulus values O using parameters λ . Then, the information gain can be calculated using Equations 7–9 (Liao & Carin, 2004):

$$F(m) = \frac{\partial g(\lambda, X_m)}{\partial \lambda}, \quad (7)$$

$$B = \sum_i \frac{1}{w_i} \left[\frac{\partial g(\lambda, O_i)}{\partial \lambda} \right] \left[\frac{\partial g(\lambda, O_i)}{\partial \lambda} \right]^T, \quad (8)$$

and

$$\text{Infogain}(m) = \log \left[1 + F(m)B^{-1}F(m)^T \right] \quad (9)$$

In the calculations of the information gain, B is a two by two matrix, and F is an M by two matrix, with $F(m)$ a two-element column vector. Once the information gain has been calculated, the next stimulus value, x_{i+1} , is selected as the argument that maximizes the information gain:

$$x_{i+1} = \arg \max_x \text{Infogain}(x). \quad (10)$$

This formulation of the gain in Fisher information was derived by Liao and Carin (2004) for the case of additive white Gaussian noise. In the present study, each trial has a binary outcome: Each response is either correct or incorrect. The statistics for the noise, defined as the difference between the outcome of each trial and the model of the psychometric function, are very difficult to formulate and make the subsequent analyses intractable. One solution is to combine the outcomes of all trials at each stimulus value to estimate the psychometric function at each stimulus value.

Defining the noise as the error between the measurement, calculated from the pooled results of all previous trials at the same stimulus value, and the psychometric function, there are two sources for the observed noise. First, some quantization error will be present when the number of trials at a stimulus value is not sufficient for fine resolution. The second source of noise is standard random sampling error. As the number of trials increases, the quantization error will become negligible, leaving only random sampling error. Figure 2 shows the results of an illustrative simulation to examine the characteristics of the noise when defined as the error between the measurements at each stimulus value and the model of the psychometric function. In the plot at the top are four arbitrarily selected points along the psychometric function. The values of each of these points were estimated in 1,000 trials, where each trial consisted of 20 binary observations. For measurements based on 20 observations, the resolution is .05; that is, the measurements of the psychometric functions will be multiples of .05. Shown at the bottom in Figure 2 are histograms of the noise at each

Figure 2. Histograms of observation noise. The model of the psychometric function and the estimates of the psychometric function calculated from pooled results of trials at the same stimulus values are shown at the top. At the bottom are histograms of the noise, defined as the difference between the measurement and the model, at the four points shown in the plot of the psychometric function. For all but the highest probability, where ceiling effects limit one tail of the distribution, the noise follows a Gaussian distribution.

of the four points collected over the 1,000 trials. For the three lower probability points on the model that were examined [$p(r) = .53$, $p(r) = .68$ and $p(r) = .82$], the noise exhibits Gaussian characteristics. As was expected, the distribution of the noise at points of high probability in the psychometric function [e.g., $p(r) = .96$] is single-tailed,

due to ceiling effects. However, for most of the points in the model, as the number of trials run at each stimulus value increases, the distribution of the noise appears to follow a Gaussian distribution. Hence, the gain in Fisher information with additive Gaussian noise derived by Liao and Carin (2004) is an appropriate approximation of the information gain in the present study.

Four adaptive psychometric procedures were investigated in this study and are listed in Table 1: PSI [or Bayesian min(H)], IPE FIG, Bayesian FIG, and IPE min(H). The PSI method was developed by Kontsevich and Tyler (1999) for vision research and is included as a current technique for estimating two parameters through sequential testing. The IPE FIG method is an alternate two-parameter estimation procedure inspired by the method used by Liao and Carin (2004) for subsurface sensing of buried targets. Neither of these techniques has been applied to the measurement of psychophysical data in hearing. The Bayesian FIG and IPE min(H) methods are two proposed hybrid techniques that combine elements from the two previous sequential-testing procedures.

SIMULATIONS

In order to generate simulated results pertinent to some real application, simulations relevant to the cochlear implant psychophysics problem were developed. There are several psychophysical variables of interest for describing a cochlear implant patient's ability to resolve different stimuli in one of several domains—for example, temporal or spectral. In several studies of cochlear implant psychophysics, channel-specific anomalies were observed in some measurements of psychophysical variables. Examples of observed psychophysical anomalies or channel interactions include abnormal forward-masking levels (Chatterjee & Shannon, 1998; Shannon, 1983; Throckmorton & Collins, 1999), indiscriminable electrodes (Busby & Clark, 1996; Zwolan, Collins, & Wakefield, 1997), or electrodes evoking pitch percepts that do not follow the expected monotonic ordering from low to high frequency (Collins, Zwolan, & Wakefield, 1997; Townshend, Cotter, Van Compernelle, & White, 1987). The presence of channel interactions and psychophysical anomalies suggests that the percepts elicited through electrical simulation of the nerve fibers local to each electrode may vary across both the electrode array and individual cochlear implant patients. In many of these studies, the time required for exhaustive testing of all the parameters and electrodes is not feasible, given the constraints of working within the schedule of the patient and clinician. Researchers often resort to experiments in which only a subset of the parameters or electrodes is considered, due to time constraints and test complexity. The difficulty associated with collecting a comprehensive set of relevant psychophysical variables inhibits researchers' and clinicians' ability to utilize this information for tuning the speech processor of each patient for improved speech recognition. Thus, efficient psychometric procedures are of significant interest, and the simulations considered here are slanted toward the cochlear implant problem.

Table 1
Psychometric Methods Investigated in This Study

Method	Parameter Estimation		Information Gain	
	Bayesian	Iterative Search	Minimize Entropy	Fisher Information Gain
PSI	X		X	
IPE FIG		X		X
Bayesian FIG	X			X
IPE min(H)		X	X	

Note—The PSI method was developed by Kontsevich and Tyler (1999) for vision research and is sometimes referred to as “Bayesian min(H)” in the present article. The IPE FIG method is inspired by Liao and Carin’s (2004) parameter search procedure applied to subsurface sensing. Bayesian FIG and IPE min(H) are two proposed hybrid techniques that combine elements from the two previous sequential-testing procedures.

To investigate the proposed adaptive psychometric procedures on a task related to cochlear implant psychophysics, a threshold detection experiment was simulated, using typical values from psychophysical studies. Simulated experiments were run using the logistic cdf to model the ability of cochlear implant subjects to detect a stimulus of given intensity in a two-interval forced choice (2IFC) test. The 2IFC task was chosen for the test simulations because it is commonly used by clinicians when testing cochlear implant patients. In a 2IFC task, the subject observes two intervals, one containing the target signal and the other serving as a reference. With two intervals, the probability of correctly identifying the target signal through random guessing is 50%; this is an unfortunate property of the 2IFC task that results in higher variability in the parameter estimates, particularly with smaller sample sizes (McKee, Klein, & Teller, 1985). Using a greater number of intervals would reduce the probability of correct response due to guessing; however, it would also lengthen the test and force the subject to mentally track a greater number of stimuli, which can be confusing to some subjects.

The logistic cdf for the 2IFC test has the form

$$p(r; \lambda, x) = .5 + \frac{1 - 2\gamma}{2 + 2e^{-s(x-m)}}, \quad (11)$$

where r is the probability of a correct response, x is the stimulus intensity, and λ is the parameter set containing the median or “threshold” m of the logistic pdf [$p(r; \lambda, x) = .75$], the parameter s controlling the slope of the psychometric function, and γ , the asymptotic probability of a Type 1 error (negative response to a detectable stimulus). The values, or range of values when appropriate, for each parameter in the psychometric function for the simulations are shown in Table 2. The range of input current intensity values x were chosen to reflect a typical range of possible values when the thresholds of a subject with a Nucleus cochlear implant are tested (Skinner et al., 1991). The available current intensity values range are integer values with a maximum possible input value of 239; however, due to the variability of surviving neural populations in cochlear implant patients, the maximum stimulus level that produces a comfortable stimulus is dependent on subject and electrode. The maximum input value of 140 is consistent with the maximum comfortable levels seen in Zwolan et al. (1997). The possible values of m , from 80 to 110, are again consistent with the subject data in Zwolan

et al. There is little published data to support the values of s appropriate for modeling psychometric functions for threshold testing in cochlear implant subjects. The range of values chosen for the simulations in this study is reasonable and broad enough to cover the most likely true values of slope parameters. During simulations, the values of m and s were chosen from a continuous uniform distribution over the range of values. The value of γ is comparable to values used by Kontsevich and Tyler (1999).

Each trial in the simulated experiments consists of a simulated response at the input stimulus value determined by the psychometric procedure. A random number, uniformly distributed between zero and one, is compared with the simulated psychometric function at the stimulus value for that trial. If the random number exceeds the value of the psychometric function, a correct response is recorded; otherwise, an incorrect response is returned. A sequence of 300 trials produces a complete simulated experiment, or one iteration, with a sufficient number of trials to measure the asymptotic performance of each psychometric method under investigation. A total of 5,000 iterations were used to gather a significant number of test cases for generating statistics.

The parameter estimation techniques contain several control variables that were set to fixed values for all 5,000 iterations. For the Bayesian parameter estimation algorithm, the probabilities were computed at fixed points in the 2-D parameter space, where m ranged from 0 to 140 (the entire range of input values) in 2.5-unit increments and s ranged from 0.2 to 5 in 0.3-unit increments. Hence, there were 969 pairs of possible parameter values. Reducing the number of points in the parameter space by using larger increments significantly reduces the simulation time by decreasing the scale of matrix operations for calculating the posterior distribution. When the likelihood function was calculated, $p(r|\lambda, x)$, γ was assumed to equal zero. For the

Table 2
Parameters of the Psychometric Function Used in Simulations Testing Adaptive Psychometric Procedures

Parameter	Value(s)
x	0–140
m	80–110
s	0.85–1.35
γ	0.02
Iterations	5,000

Figure 3. Bias (top) and variance (bottom) of the threshold parameter estimates. Results are shown for four adaptive psychometric procedures: PSI [i.e., Bayes min(H)], IPE FIG, Bayes FIG, and IPE min(H). Bias and variance are calculated from 5,000 iterations (i.e., simulated experiments), each 300 trials in length. The goal is to achieve zero bias (0% error) with the lowest possible variance.

parameter value iterative search, the bounds set for m and s were 0–140 and 0.2–5, respectively. These bounds are the same as the limits for the 2-D parameter search space with the Bayesian parameter estimation technique, so both methods are searching over the same space and can take the same maximum and minimum values. For each level of the iterative search, the search space is divided into 100 steps ($N = 100$). The range of values for the next level is 25% of the range for the previous step, so the resolution increases by a factor of four at each subsequent level and the

2-D parameter space is 1/16 the previous size. The iterative search continues until the reduction in resolution changes the norm of the error by less than 10^{-10} .

Two existing adaptive psychometric methods were implemented to provide a baseline for comparison. The transformed Levitt (1971) procedure, commonly used for measuring psychophysics in cochlear implant patients, was started with an initial stimulus value of 71, the same initial stimulus value as that used with the investigated techniques. The initial step size was 5 units until three reversals had been recorded and then 2 units for the rest of the simulated experiment. The threshold estimate was calculated using the last eight reversals. The ZEST method used the same threshold parameter space as the Bayesian parameter estimation (m ranged from 0 to 140 in 2.5-step increments). The slope parameter for ZEST was fixed to the mean of the true slope parameter distribution. The probability distribution of the threshold parameter values was initialized using a uniform distribution. For an initial stimulus value, the ZEST method calculates the expected value of the threshold, using the initial distribution of the parameters. With the uniform distribution, this initial stimulus value is also 71, the midpoint of the range of parameter estimates.

RESULTS

The performance of the four psychometric techniques [PSI, IPE FIG, Bayesian FIG, and IPE min(H)] was evaluated in terms of the bias and variance of the threshold and slope parameter estimates. The bias and variance will be presented as a percentage of the “true” parameter value. The median value of each range of parameter values in the simulations was used to calculate the percentage of error. Therefore, bias and variance of the threshold parameter will be reported as a percentage of 95, and the slope statistics as a percentage of 1.1. When the bias and variance were calculated, the actual parameter values for each iteration were used to determine the performance of the estimator. Figure 3 shows the bias and variance for threshold estimation with each of the four methods. The best performance, as determined by the bias and variance, was obtained with the Bayesian FIG and PSI methods, both of which utilize the Bayesian procedure for parameter estimation. Achieving equivalent levels of performance with either IPE FIG or IPE min(H) requires a significantly greater number of trials. Figure 4 shows the bias and variance for the estimation of the slope parameter. Plotted as percentage error, the slope parameter is generally more difficult to estimate than the threshold parameter. As can be seen for the error surface plotted in Figure 1, the gradient in the direction of the slope parameter is much shallower. Therefore, the fit of the psychometric function to the observed data is much less sensitive to the slope parameter value. Three of the methods exhibit a similar bias for estimating the slope parameters. The bias for the procedures tends to be negative, since the true value of γ is unknown when the response probabilities are calculated (with Bayesian parameter estimation) or when the error between the observed data and the estimated psychometric function is calculated (with

iterative parameter estimation). Although IPE FIG has the lowest bias after 300 trials and appears to be converging on the true parameter values, rather than exhibiting a negative bias, extended simulations have shown that it will also cross the zero bias point (~ 400 trials) and asymptote near the other methods. Thus, it is actually lagging the other methods. The Bayesian FIG and PSI methods have similar variance curves and produce the lowest variance estimates of the slope parameter. The IPE FIG method asymptotes at a similar level but requires a larger number of trials for the variance to decrease. In Figures 3 and 4, the IPE min(H) technique displays higher variance than do the other procedures and much greater bias in the threshold parameter estimates.

The bias and variance curves for the threshold and slope parameters provide a quantitative measure of estimator performance. To better understand the behavior of the different psychometric procedures considered in this study, histograms of the threshold and slope parameter estimates after 250 trials in each of the 5,000 simulated experiments are shown in Figure 5. Focusing first on the threshold estimates generated by each of the four psychometric procedures, the histograms for the two procedures that utilize the Bayesian parameter estimation technique [Bayesian FIG and PSI/Bayesian min(H)] exhibit a comb-like character. This behavior is a function of the sampling resolution in the parameter space for the nonparametric probability distribution of the parameter values. Due to the selected sequence of trials and the error between the observed data and the model of the psychometric function, these estimates tend to cluster around the sampled points in the parameter space; that is, the parameter probability distribution has very low entropy. Comparing the histograms of the PSI/Bayesian min(H) method and the Bayesian FIG method suggests that this effect is lessened when the gain in Fisher information is used, rather than minimizing the entropy, to select the next sample point. The procedures using the iterative parameter estimation (IPE) technique generate a more continuous range of estimates. In some instances, the iterative search finds local minima in the error space that are well outside the range of any parameter values used in the simulations. This occurs more frequently when the minimized expected entropy is used to select the sequence of trials. The histograms of the slope parameter estimates after 250 trials for each of the four proposed psychometric procedures should ideally be uniformly distributed between 0.85 and 1.35. With the IPE FIG method, the slope estimates follow an approximate Gaussian distribution centered at 1.1, the mean of the distribution of the true slope values. The slope parameter histograms for the two methods using Bayesian parameter estimation have similar shapes, but again, the PSI/Bayesian min(H) method displays a comb-like structure in the histogram, with spikes centered at the sampled values in the parameter search space (e.g., 0.2, 0.5, 0.8). The sequence of trials that results when the gain in Fisher information is used produces more intermediate estimates by assigning nearly equal probability to two neighboring points in the parameter value probability distribution.

Figure 4. Bias (top) and variance (bottom) of the slope parameter estimates. Results are shown for four adaptive psychometric procedures: PSI [i.e., Bayes min(H)], IPE FIG, Bayes FIG, and IPE min(H). Bias and variance are calculated from 5,000 iterations (i.e., simulated experiments), each 300 trials in length. The goal is to achieve zero bias (0% error) with the lowest possible variance.

Figure 6 shows scatterplots of the error in the threshold and slope parameter estimates for each of the 5,000 simulated experiments after 250 trials, using each of the four investigated methods. The X at (0,0) in each plot identifies correct estimates of both the slope and the threshold parameter values. Parameter estimate error is plotted as a percentage of the true parameter values for each simulated experiment. From comparison of the two methods using IPE, the divergent estimates are apparent. Looking at the two Bayesian methods, it can be seen that the PSI

Figure 5. Histograms of the threshold and slope parameter estimates for each of the four investigated psychometric procedures. In each histogram of threshold estimates (left), the x-axis indicates the estimated value of the threshold parameter (true values range from 80 to 110). In each histogram of slope estimates (right), the x-axis indicates the estimated value of the slope parameter (true values range from 0.85 to 1.35).

method has a greater spread of parameter estimates. The parameter estimates with the Bayesian FIG method are more symmetric about the (0,0) point than are the estimates with the PSI method. The asymmetry of the parameter estimates with the PSI method results in the higher calculated bias, and the skew away from (0,0) is an effect of the more prominent comb-like characteristic exhibited in the histograms of the parameter estimates with the PSI method.

From the comparison of the four methods, the proposed Bayesian FIG method appears to perform with the lowest bias and variance for threshold estimation. The Bayesian FIG method also performed very well for slope estimation. Of the four methods compared in Figures 3–6, the Bayesian FIG method appears to have the most promise for providing fast estimates with low bias and variance. To compare the Bayesian FIG method with previously investigated and current methods used in psychometrics, the bias and variance for Bayesian FIG is plotted in Figure 7 versus the threshold bias and variance for the transformed Levitt (1971) procedure and the ZEST method. The results show that the Bayesian FIG method is able to outperform both techniques on threshold estimation, while simultane-

ously estimating the slope parameter. Bias and variance are plotted out to 300 trials in order to show asymptotic performance of the proposed Bayes FIG procedure. However, cochlear implant psychophysical tests in a research setting typically run between 30 and 60 trials. Even at these points, the Bayes FIG procedure estimates the parameters with lower bias and variance. Depending on the stimulus values, the slope parameter may or may not be estimable from the data collected using the transformed Levitt procedure. As was mentioned previously, the ZEST method assumes a value for the slope parameter (in these simulations, $s = \text{mean}(s_{\text{TRUE}})$, the mean value in the probability distribution from which the true slope parameters are selected) and proceeds to estimate only the threshold parameter. The arbitrary value of the slope parameter used in these simulations was chosen to provide moderate performance and assumes that the experimenter has enough knowledge about the task to reasonably estimate the slope parameter. Simulations at different assumed slope values indicated that performance is only marginally affected by small shifts in the assumed value of the slope parameter. The asymptotic limit of the threshold estimation bias observed with the ZEST method is due, in part, to the sampling resolution in the threshold

Figure 6. Scatterplots of parameter estimate error for all four investigated psychometric methods. Each point plots the parameter estimate error after 250 trials for one of the 5,000 iterations. Error is plotted as a percentage of the true value for each iteration. Points at (0%, 0%) identify iterations that achieved zero bias in both the threshold and the slope parameters.

dimension of the parameter space. The sampling resolution affects the threshold estimate bias, since the ZEST method frequently assigned high probabilities to a single threshold estimate, rather than assigning equal probabilities to two neighboring estimates. This is the same behavior that was observed with the PSI method, although the effect is more pronounced in the results shown in Figure 7. The Bayesian FIG method makes better decisions about where to locate the stimulus for the next trial, resulting in lower bias and variance, as is shown in Figures 3, 4, and 7.

DISCUSSION

In this study, more sophisticated psychometric techniques based on Bayesian procedures and the theory of optimal experiments were evaluated through simulations to determine their effectiveness for measuring psychometric variables. Four psychometric procedures, two based on previous studies and two developed as hybrids from those two procedures, use two different approaches to estimating the parameters from the observed data and selecting the stimulus value for the next trial that will provide the most information. On the basis of these simulations, using a Bayesian technique to estimate the parameter values, combined with the gain in Fisher information to measure the information provided by each stimulus value and se-

lect the stimulus for the next trial, provides the best estimation performance, as measured by the bias and variance of the estimates. This approach has not previously been evaluated in the psychophysical literature. In addition, these techniques provide a substantial improvement in efficiency over current techniques.

The results shown in Figures 3 and 4 for the threshold and slope parameter estimates, respectively, suggest that the combination of iterative parameter estimation and selecting the next stimulus value through minimizing the expected entropy produces a poor sequential-testing procedure. If the iterative parameter estimation finds poor estimates of the parameters, due to an unfortunate or improbable but possible series of observations, selecting the next stimulus to minimize the expected entropy can actually involve rejecting stimulus values that may correct the parameter estimates. Instead, a stimulus value may be chosen that is uninformative with respect to the true parameter values but succeeds in reinforcing the confidence in the poor parameter estimates.

This leads to a more general observation regarding the underlying principle of minimizing the expected entropy of the parameter probability distribution and two examples of when it may serve as a liability. First, as was mentioned previously, incorrect parameter estimates can be reinforced by stimulus values that neither provide

Figure 7. Comparison of threshold parameter estimates measured using the transformed Levitt procedure, ZEST method, and Bayesian FIG method. The threshold bias (top) and variance (bottom) were calculated using 5,000 iterations (i.e., simulated experiments) and are plotted over 300 trials.

information about the true parameter values nor rigorously examine the current parameter estimates. Second, when a nonparametric probability distribution is used for the parameters (e.g., the Bayesian parameter estimation technique), it may be preferable to assign nonzero probabilities to several neighboring points in the probability distribution, so that the mean estimate is approximately equal to the observed truth. However, in minimizing the expected entropy, stimulus values are sought that potentially reduce the parameter probability distribution to a single nonzero value. Using the gain in Fisher information to select the next stimulus value does not depend on the parameter probability distribution; instead, it uses the

psychometric function model and the number of previous trials at each possible stimulus value to minimize the variance of observations of the most informative points on the model. Therefore, in the two scenarios described above, the gain in Fisher information provides a more robust stimulus selection technique than does minimizing the expected entropy.

There are many factors to consider when the results provided by the simulations conducted in this study are interpreted. The choice of γ , equivalent to the probability of a Type 1 error, has some effect on the performance of the proposed psychometric methods. The inaccuracy of the γ parameter accounts for some of the bias in the parameter estimation and the sharp dip in the bias of the threshold estimates. Kontsevich and Tyler (1999) performed simulations using various assumed values of γ . They found that the effects of γ on the threshold estimate diminish as the experiment progresses but some influence remains on the slope parameter estimate. The importance of γ and its assumed value in the psychometric procedure was illustrated during examination of the Bayesian parameter estimation procedure. When the true value of γ is unknown and an incorrect value is assumed, which would seem to be the case under most circumstances, the Bayesian parameter estimation tends to assign a very high probability to one of possible threshold values in the probability distribution of the threshold estimates. For example, if the true value of the threshold parameter is 88.5, ideally $\text{prob}(m = 88) = .5$ and $\text{prob}(m = 89) = .5$, so that the expected value of the threshold is 88.5. If the true value of γ is known and the stimulus values are selected properly, this will be the outcome. However, if γ is incorrect, it is much more likely for one of the probabilities to be near unity, increasing the bias. This is due to the effect of γ on the likelihood $p(r|\lambda, x)$.

The second factor to consider is the computational complexity and execution time. These psychometric procedures will be used to measure psychophysical variables in human subjects. Therefore, parameter estimation and selection of the next stimulus value need to occur in a reasonable time frame, to prevent delaying or interrupting the experiment. The psychophysical procedures as simulated execute in sufficient time for use with human subjects. One of the variables with the most influence on the execution time of any of the proposed psychometric techniques was the sampling resolution of the parameter space for Bayesian parameter estimation. Increasing the resolution decreases the bias, so better results than those shown can be obtained with the trade-off of increased computational complexity. In addition, in more general applications in which more parameter dimensions are added for larger experiments, computational complexity will become a greater issue.

The third factor related to the use of the proposed psychometric procedures is the complexity of the implementation. The popular transformed Levitt procedure is a very simple method that relies on a small set of rules for conducting the experiment and does not assume any form for the psychometric function. However, implementing the gain in Fisher information calculation, for example, requires mathematical manipulations of the psychometric

model and special numerical implementation techniques to prevent overflow errors that will corrupt the calculations. Due to these issues, the proposed psychometric methods require additional effort to implement and set up. However, the results presented in this study suggest that these methods can potentially expedite the measurement of psychophysical variables. The relative value of the subject's time, versus the one-time investment in implementing the procedures, supports the use of more complicated, more sophisticated techniques to improve efficiency.

To summarize, the Bayesian parameter estimation technique appears to be the better choice for estimating the parameters of the psychometric function. The issue raised earlier about parameter space sampling resolution is only a concern when the Bayesian technique is compared with an alternate method of parameter estimation; however, the results presented here seem to clearly indicate better performance when the Bayesian procedure is used. The discussion then focuses on the best way to select the stimulus value for the next trial. The results presented in this study indicate that the gain in Fisher information, as used by Liao and Carin (2004), provides better selections of the stimuli for subsequent trials than does minimizing the entropy, as used in the PSI method. The outcomes of the computer-simulated trials for psychometric techniques based on Bayesian procedures and the theory of optimal experiments encourage further examination and adaptation to testing with human subjects.

CONCLUSIONS

1. More sophisticated psychometric techniques based on the theory of optimal experiments and Bayesian adaptive procedures should be able to measure cochlear implant psychophysical variables with greater efficiency than do standard staircase techniques.

2. The Bayesian parameter estimation provides stable and accurate estimates of the psychometric function parameters, but performance is dependent on the selected stimulus values for the sequential trials. The gain in Fisher information appears to outperform the expected entropy when the next stimulus values are selected.

AUTHOR NOTE

This research was supported by NIH Grant 1-R01-DC007994-01. The authors thank the anonymous reviewers for their helpful feedback. Correspondence concerning this article should be addressed to L. M. Collins, Department of Electrical and Computer Engineering, Duke University, Box 90291, Durham, NC 27708 (e-mail: lcollins@ee.duke.edu).

REFERENCES

- BUSBY, P. A., & CLARK, G. M. (1996). Electrode discrimination by early-deafened cochlear implant patients. *Audiology*, **35**, 8-22.
- CHATTERJEE, M., & SHANNON, R. V. (1998). Forward masked excitation patterns in multielectrode electrical stimulation. *Journal of the Acoustical Society of America*, **103**, 2565-2572.
- CHERNOFF, H. (1972, December). *Sequential analysis and optimal design*. Paper presented at the 8th Regional Conference of the Society for Industrial and Applied Mathematics, Philadelphia.
- COLLINS, L. M., ZWOLAN, T. A., & WAKEFIELD, G. H. (1997). Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects. *Journal of the Acoustical Society of America*, **101**, 440-455.
- COVER, T. M., & THOMAS, J. A. (1991). *Elements of information theory*. New York: Wiley.
- FEDOROV, V. V. (1972). *Theory of optimal experiments*. New York: Academic Press.
- KING-SMITH, P. E., GRIGSBY, S. S., VINGRYS, A. J., BENES, S. C., & SUPOWIT, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: Theory, simulations, experimental evaluation and practical implementation. *Vision Research*, **34**, 885-912.
- KLEIN, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics*, **63**, 1421-1455.
- KONTSEVICH, L. L., & TYLER, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, **39**, 2729-2737.
- LEVITT, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, **49**, 467-477.
- LIAO, X., & CARIN, L. (2004). Application of the theory of optimal experiments to adaptive electromagnetic-induction sensing of buried targets. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **26**, 961-972.
- MACKAY, D. J. C. (1992). Information-based objective functions for active data selection. *Neural Computation*, **4**, 589-603.
- 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.
- PELLI, D. G. (1987, May). *The ideal psychometric procedure*. Paper presented at the Investigative Ophthalmology and Visual Science, Sarasota, FL.
- SHANNON, R. V. (1983). Multichannel electrical stimulation of the auditory nerve in man: II. Channel interaction. *Hearing Research*, **12**, 1-16.
- SKINNER, M. W., HOLDEN, L. K., HOLDEN, T. A., DOWELL, R. C., SELIGMAN, P. M., BRIMACOMBE, J. A., & BEITER, A. L. (1991). Performance of postlingually deaf adults with the wearable speech processor (WSP III) and mini speech processor (MSP) of the nucleus multi-electrode cochlear implant. *Ear & Hearing*, **12**, 3-22.
- THROCKMORTON, C. S., & COLLINS, L. M. (1999). Investigation of the effects of temporal and spatial interactions on speech-recognition skills in cochlear-implant subjects. *Journal of the Acoustical Society of America*, **105**, 861-873.
- TOWNSHEND, B., COTTER, N., VAN COMPERNOLLE, D., & WHITE, R. L. (1987). Pitch perception by cochlear implant subjects. *Journal of the Acoustical Society of America*, **82**, 106-115.
- ULRICH, R., & MILLER, J. (2004). Threshold estimation in two-alternative forced-choice (2AFC) tasks: The Spearman-Kärber method. *Perception & Psychophysics*, **66**, 517-533.
- WATSON, A. B., & PELLI, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, **33**, 113-120.
- WHAITE, P., & FERRIE, F. P. (1997). Autonomous exploration: Driven by uncertainty. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **19**, 193-205.
- ZWOLAN, T. A., COLLINS, L. M., & WAKEFIELD, G. H. (1997). Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects. *Journal of the Acoustical Society of America*, **102**, 3673-3685.

(Manuscript received October 6, 2005;
revision accepted for publication June 28, 2006.)