

# Bayesian adaptive estimation of the auditory filter

Yi Shen<sup>a)</sup> and Virginia M. Richards

Department of Cognitive Sciences, University of California, Irvine, 3151 Social Science Plaza, Irvine, California 92687-5100

(Received 3 December 2012; revised 12 June 2013; accepted 17 June 2013)

A Bayesian adaptive procedure for estimating the auditory-filter shape was proposed and evaluated using young, normal-hearing listeners at moderate stimulus levels. The resulting quick-auditory-filter (qAF) procedure assumed the power spectrum model of masking with the auditory-filter shape being modeled using a spectrally symmetric, two-parameter rounded-exponential (*roex*) function. During data collection using the qAF procedure, listeners detected the presence of a pure-tone signal presented in the spectral notch of a noise masker. Dependent on the listener's response on each trial, the posterior probability distributions of the model parameters were updated, and the resulting parameter estimates were then used to optimize the choice of stimulus parameters for the subsequent trials. Results showed that the qAF procedure gave similar parameter estimates to the traditional threshold-based procedure in many cases and was able to reasonably predict the masked signal thresholds. Additional measurements suggested that occasional failures of the qAF procedure to reliably converge could be a consequence of incorrect responses early in a qAF track. The addition of a parameter describing lapses of attention reduced the likelihood of such failures. © 2013 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4812856>]

PACS number(s): 43.66.Ba, 43.66.Yw [TD]

Pages: 1134–1145

## I. INTRODUCTION

One of the fundamental features of the auditory system is its tonotopic organization. The auditory periphery acts as a frequency analyzer, mapping different frequency components of sounds to specific locations along the basilar membrane. Functionally, this process can be modeled as a bank of band-pass filters, namely the auditory filters (AF). The shape of the AF, in particular its bandwidth, is highly predictive of perceptual phenomena such as masking. Not surprisingly, the characterization of the AF was an important topic in early psychophysical studies of hearing (e.g., Wegel and Lane, 1924; Fletcher, 1940).

Paradigms to estimate the AF shape are well established. A frequently adopted method for estimating AF shapes is the notched-noise masking experiment (e.g., Patterson, 1976; Patterson *et al.*, 1982; Rosen and Baker, 1994; Moore, 1995) in which the detection of a pure-tone signal presented within the spectral notch of a noise masker is measured. In most situations, the signal threshold can be well described by the amount of masker energy at the output of the AF centered at the signal frequency. This power spectrum model of masking predicts that as the width of the spectral notch increases, the signal threshold improves. Therefore, by measuring the signal threshold as a function of the notch width and then using the power spectrum model of masking to fit the data, the shape of the AF can be derived (e.g., Patterson *et al.*, 1982; Glasberg and Moore, 1990). Although the specific formulation of the auditory-filter shape and the experimental methods might vary from study to study, procedures for the estimation of the AF have adopted

a similar approach: Thresholds are measured, and then the parameters of the AF shape are estimated based on the threshold data. For this reason, we will refer to these traditional procedures as threshold-based procedures.

Threshold-based procedures can be very time consuming, often requiring threshold estimation using more than ten notch bandwidths for each AF estimate. This drawback of the threshold-based procedures has been partially overcome by reducing the total number of thresholds to be measured (Stone *et al.*, 1992; Leeuw and Dreschler, 1994), or by improving the time efficiency of the psychophysical procedures for threshold estimation (Leeuw and Dreschler, 1994; Nakaichi *et al.*, 2003; Hopkins and Moore, 2011).

In addition to efforts to more rapidly estimate frequency selectivity using the notched-noise paradigm, efforts to efficiently estimate the psychophysical tuning curve (PTC) have been described. The PTC represents the level of a narrow-band masker required to just mask a fixed-frequency, fixed-level, tonal probe as a function of the masker's center frequency. Traditionally, the PTC is collected by measuring the masked threshold of the probe tone for various masker center frequencies, which usually takes hours to complete. Søk *et al.* (2005), using a Békésy procedure, significantly improved the time required for the measurement of the PTC. In their procedure, the masker level at the detection threshold was adaptively tracked as the center frequency of the masker was slowly changed in frequency. As demonstrated using normal-hearing listeners (Søk *et al.*, 2005; Charaziak *et al.*, 2012) and school-age children (Malicka *et al.*, 2009), this modification to the traditional procedure reduced the time needed to measure a PTC to less than 10 min.

Here, we propose a computational procedure for the rapid estimation of frequency selectivity using the notched-noise method. In contrast to threshold-based procedures, in

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: shen.yi@uci.edu

the current study a “parameter-based” procedure is considered. The goal is to estimate the parameters that describe the AF shape directly rather than first estimating thresholds and then fitting the model. Because the number of parameters used to express the AF shape is small, parameter-based procedures have the potential to be very efficient.

While there have been no prior efforts to use the parameter-based approach for the estimation of AF shape, it has been widely used for the estimation of psychometric functions (e.g., Green, 1990; King-Smith and Rose, 1997; Kontsevich and Tyler, 1999; Leek *et al.*, 2000; Shen and Richards, 2012; Shen, 2013). The goal of parameter-based procedures for the estimation of the psychometric function is to adaptively search for the optimal stimulus sampling strategy as responses are collected. Compared to the psychometric function, the estimation of the AF requires adaptively manipulating stimuli in a two-dimensional stimulus space, i.e., the signal level and the notch width.

Lesmes *et al.* (2010) proposed an adaptive procedure for the estimation of the contrast sensitivity function (CSF) in vision. The CSF describes the ability to see a low-contrast grating as a function of the grating’s spatial frequency. To overcome the time consuming process for the estimation of the CSF, Lesmes *et al.* (2010) extended the Bayesian adaptive algorithm developed by Kontsevich and Tyler (1999) to estimate a four-parameter CSF by efficiently sampling the two dimensional stimulus parameter space (grating frequency and contrast). Simulation and psychophysical data suggested that this procedure achieved excellent accuracy using only 100 trials.

In a recent study, a parameter-based procedure was proposed by Shen and Richards (2013) for estimating the temporal modulation transfer function (TMTF). The TMTF describes the threshold for detecting sinusoidal amplitude modulation as a function of modulation rate. It is often modeled using a first-order, low-pass function and is thought to reflect the resolution of auditory temporal processing. To enable efficient estimation of the TMTF, Shen and Richards (2013) adaptively varied two stimulus parameters, modulation depth and modulation rate, from trial to trial in a modulation detection task. A computational algorithm very similar to that described by Lesmes *et al.* (2010) was used to determine the optimal stimulus on each trial based on all previously collected responses. The information gain (or entropy loss) on each trial was maximized. Both simulation and behavioral data suggested that the parameter-based procedure estimated the two parameters of the TMTF (i.e., sensitivity and cutoff frequency) more efficiently than the traditional threshold-based procedure, reducing the time for data collection from a few hours to about 10 min.

In the current study, an adaptive parameter-estimation procedure was developed and tested to evaluate its efficiency in estimating the shape of the AF. The resulting procedure, the “quickAF” or “qAF” procedure was similar to that described by Shen and Richards (2013) for the estimation of the TMTF. In the following sections, a detailed description of the qAF procedure is given. Then, three experiments are presented which compare AF shapes for the qAF and threshold-based procedures (in experiment I), and the qAF procedure’s robustness to changes in the underlying model

assumptions (in experiment II) and to early response errors (in experiment III).

## II. AN ADAPTIVE BAYESIAN PROCEDURE FOR EFFICIENT ESTIMATION OF THE AUDITORY FILTER

The Bayesian qAF procedure was developed based on the work of Kontsevich and Tyler (1999), and parallels the procedures developed by Lesmes *et al.* (2006) and Shen and Richards (2013). The goal was to rapidly estimate the parameters of the function describing the shape of the AF using a notched-noise paradigm.

Patterson *et al.* (1982) suggested the two-parameter rounded-exponential (*roex*) function to describe the shape of the auditory filter. The *roex*( $r, p$ ) function is defined as

$$W(g) = (1 - r)(1 + pg)e^{-pg} + r, \quad (1)$$

where  $W$  is the transfer function of the filter,  $g$  is the normalized deviation in frequency from the center of the filter,  $p$  is related to the slopes of the filter, and consequently the auditory filter bandwidth, and  $r$  is the minimum value of the filter response. According to the power spectrum model of masking, the signal power at the detection threshold  $P_s$  and the auditory filter  $W$  are related in the following way:

$$P_s = 2K \int_0^{\infty} N(g)W(g)dg, \quad (2)$$

where  $N(g)$  represents the long-term power spectrum of the masker, and  $K$  describes the listener’s efficiency. Note that the above equation assumes the shape of the auditory filter to be symmetric in frequency, which is a reasonable assumption for low-to-moderate stimulus levels but not for high stimulus levels (e.g., Glasberg and Moore, 1990).

For a noise masker with a spectrum level of  $N_0$  and a spectral notch with a bandwidth  $2\Delta f$ , the detection threshold of a pure-tone signal presented at the center of the notch ( $f_0$ ) is predicted to be (e.g., Patterson *et al.*, 1982):

$$P_s(\Delta f/f_0) = 2KN_0f_0[-(1 - r)p^{-1}(2 + pg)e^{-pg} + rg]_{\Delta f/f_0}^{0.8}, \quad (3)$$

where 0.8 is the assumed integration limit. To derive the model parameters  $K$ ,  $r$ , and  $p$ , signal thresholds are typically measured for various values of  $\Delta f$ . Then, the function relating notch bandwidth and the corresponding signal levels at threshold is fitted using the function in Eq. (3).

Here, we propose to fit the parameters  $K$ ,  $r$ , and  $p$  directly from a listener’s trial-by-trial responses. The responses are modeled using a logistic psychometric function:

$$PC(x, \Delta f/f_0; K, r, p) = \gamma + (1 - \gamma)/[1 + e^{-\beta(x - 10\log P_s)}], \quad (4)$$

where  $PC(x, \Delta f/f_0; K, r, p)$  is the probability of a response being correct,  $x$  is the signal level on each trial in dB sound pressure level (SPL),  $\gamma$  is the chance performance level (i.e., 0.5 for a 2-alternative forced-choice task), and  $\beta$  is related to the slope of the psychometric function. The signal threshold

[Eq. (3)] is defined as being at the center of the range of the psychometric function; therefore PC is a function of the stimulus parameters  $x$  and  $\Delta f/f_0$  and the model parameters  $K$ ,  $r$ , and  $p$ .

The proposed Bayesian sampling procedure is similar to that described by Kontsevich and Tyler (1999) in which the stimulus parameters are chosen adaptively to enable efficient estimation of the model parameters. The steps are as follows. First the ranges for the stimulus parameters are chosen and the densities of the potential values within the ranges are defined. That is, the parameter space is discrete with a certain gradation. Next, the ranges of  $K$ ,  $r$ , and  $p$  and their gradation are set. A prior distribution is chosen for each of the parameters. During a qAF run, listeners detect the presence of a pure-tone signal (with a signal level given by  $x$ ) presented in a simultaneous notched-noise masker (with a notch width given by  $\Delta f/f_0$ ) in a two-interval, two-alternative forced choice task. Following the  $i$ th trial, the posterior distribution in the three-dimensional parameter space is updated according to the listener's response on the  $i$ th trial. Then, for each potential combination of  $x$  and  $\Delta f/f_0$ , the expected total entropy of the posterior parameter distribution is calculated based on possible outcomes for the  $(i+1)$ th trial (i.e., either a correct or an incorrect response from the listener). The stimulus pair of  $x$  and  $\Delta f/f_0$  that leads to a minimum expected entropy is chosen for stimulus presentation on the  $(i+1)$ th trial. This one-step-ahead search algorithm maximizes the information gain from each trial with regard to the uncertainty of the model parameters.<sup>1</sup> Details of the computational implementation of the algorithm have been described by Shen and Richards (2013, Sec. II A) for a comparable application. After data collection, final parameter estimates are derived as the mean of the posterior parameter distribution.

One advantage of the qAF procedure is the possibility of incorporating prior parameter distributions. Prior parameter distributions reflect an experimenter's *a priori* belief about the most likely parameter values. If the experimenter has reasonable knowledge about the potential distribution of a model parameter, an "informative" prior might be used. For example, the distributions of  $K$ ,  $r$ , and  $p$  for the young, normal-hearing population are readily available in the literature (e.g., Patterson *et al.*, 1982; Moore, 1987; Wright, 1996; Badri *et al.*, 2011); therefore, they can be implemented in the qAF procedure as the prior distributions. For data collection with a relative small number of trials, appropriately chosen priors promote rapid convergence of the parameter estimates.

For the current study, one of the main purposes was to compare the qAF procedure to the traditional threshold-based procedures. To enable relatively "fair" comparisons, only weakly informative priors were used for the qAF procedure, such that both the qAF and the traditional procedures were based on few assumptions regarding the likely values of the model parameters. For all of the experiments presented here, all potential values for each of the model parameters were, *a priori*, assumed to be equally likely. It is worth pointing out that this choice of the priors was based on the assumption that the probability of the parameters being outside of their defined ranges was zero; therefore, they were not strictly "uninformative" priors.

### III. EXPERIMENT I: THRESHOLD- VERSUS PARAMETER-BASED PROCEDURES FOR THE AUDITORY-FILTER ESTIMATION

This experiment compared AF-parameters estimated using threshold-based and parameter-based approaches. The two goals were to (a) compare estimates and reliability across the two methods, and (b) examine the test-retest reliability of the proposed qAF procedure.

#### A. Method

##### 1. Listeners

Six listeners participated, five of whom completed all conditions. All had absolute thresholds of 20 dB hearing level (HL) or better at audiometric frequencies from 250 to 8000 Hz, except S6, whose absolute threshold in his left ear was 25 dB HL at 6000 Hz. All listeners were aged between 20 and 23 yr, except S6, who was 42 yr of age. Three of the six listeners (S1-3) had previous experience in psychoacoustic experiments but had no experience with tone-in-noise masking experiments. The remaining three listeners were naive with regard to psychoacoustic experimentation, except for audiometric measurement completed prior to the experiment. The protocol for this study, including experiments II and III, was approved by the Institutional Review Board at the University of California, Irvine.

##### 2. Stimuli

A two-interval, two-alternative forced-choice procedure was used with a pure-tone signal being presented either in the first or second temporal interval with equal probability. Listeners were instructed to select which interval contained the signal tone. The two intervals were separated by a 400-ms period of silence. The signal to be detected was a 2000-Hz tone and the masker was a noise low-pass filtered at 8000 Hz with a spectral notch at 2000 Hz. The spectrum level of the masker was 30 dB SPL. The stimuli were digitally generated using a sampling frequency of 44.1 kHz on a PC, which also controlled the experimental procedure and data collection through custom-written software in MATLAB (The MathWorks, Inc.). The signal and masker durations were 372 ms, including 10-ms raised cosine onset and offset ramps. The masker was generated by taking the FFT of a random sample of Gaussian noise, setting the magnitudes and phases of the appropriate components to zero as required to generate the notch width desired, and then taking an inverse FFT of the resulting vector.<sup>2</sup> The stimuli were presented to the left ear via a 24-bit soundcard (Envy24 PCI audio controller, VIA technologies, Inc.), a programmable attenuator and headphone buffer (PA4 and HB6, Tucker-Davis Technologies, Inc.) and a Sennheiser HD410 SL headset. Each stimulus presentation was followed by visual feedback indicating the correct response.

##### 3. Procedures

The estimation of the AF shape was conducted using both threshold-based and qAF procedures. Each run of the

qAF procedure began with the initialization of the parameter space, followed by 150 experimental trials (three blocks of 50 trials). Listeners were encouraged to take short breaks in between blocks.

Four runs of the qAF procedure were conducted (except for S3), one of which preceded the threshold-based procedure. The other three were obtained after the threshold-based procedure was completed. For listener S3, only the first two runs of the qAF procedure were tested due to the limited availability of this listener.

The first two and last two runs had slight differences in terms of the initialization of the parameter space. For runs 1–2, the parameters  $\beta$  and  $\gamma$  in the assumed psychometric function [Eq. (4)] were treated as constant and took the values of 1 and 0.5, respectively. For the stimulus parameters, the potential values of  $\Delta f/f_0$  [see Eq. (3)] were 0.05, 0.1, 0.2, 0.4, and 0.6 ( $\Delta f = 100, 200, 400, 800,$  and  $1200$  Hz), and the potential signal levels  $x$  were between 30 and 80 dB SPL with 5-dB steps. Thus, the potential signal levels that could be tested were coarsely defined, and a very large range was tested. For the model parameters, the potential values of  $K$  ranged from 0.2 to 2 in steps of 0.2; the potential values of  $r$  were 0,  $1 \times 10^{-5}$ ,  $5 \times 10^{-5}$ ,  $1 \times 10^{-4}$ ,  $5 \times 10^{-4}$ ,  $1 \times 10^{-3}$ ,  $5 \times 10^{-3}$ ,  $1 \times 10^{-2}$ ,  $5 \times 10^{-2}$ ,  $1 \times 10^{-1}$ ; and the potential values of  $p$  ranged from 10 to 40 in steps of 5.

For runs 3–4, the parameters  $\beta$ ,  $\gamma$ , and  $\Delta f/f_0$  were set up as for the first two runs, but the potential values of  $K$  ranged from 0.4 to 2.4 in steps of 0.2, the potential values of  $r$  were 0,  $1 \times 10^{-6}$ ,  $5 \times 10^{-6}$ ,  $1 \times 10^{-5}$ ,  $5 \times 10^{-5}$ ,  $1 \times 10^{-4}$ ,  $5 \times 10^{-4}$ ,  $1 \times 10^{-3}$ ,  $5 \times 10^{-2}$ , and the potential values of  $p$  ranged from 10 to 65 in steps of 5. The potential signal levels  $x$  were between 10 and 60 dB SPL with 3.3-dB steps. In runs 3–4, the proportion correct produced by the model [as in Eq. (4)] was limited at an upper boundary of  $(1 - 1 \times 10^{-10})$ . This was implemented to prevent the occurrence of numerical errors for the following reason. The calculation of the expected entropy, which preceded each trial during the qAF procedure, involved taking the logarithm of  $(1 - PC)$ . In cases where the value of  $PC$  [Eq. (4)] was very close to 1, the logarithm of  $(1 - PC)$  approached infinity which undermined the stability of the procedure.

Following data collection, the trial-by-trial stimulus parameters and responses were used to generate the final estimates of the model parameters. For this purpose, a parameter space with much finer gradation than that used during data collection was implemented. The parameter  $K$  had 21 logarithmically spaced values between 0.3 and 3.7,  $r$  had 61 logarithmically spaced values between  $1 \times 10^{-6}$  and  $1 \times 10^{-1}$ , and  $p$  took 56 linearly spaced values between 10 and 65. The purpose of implementing finer grids for the final parameter estimation was to provide an accurate estimate of the posterior parameter distributions. As in the experiment, all possible combinations of the three parameters were initialized to have equal probability (flat priors were used). For each qAF run, the values of  $x$ ,  $\Delta f/f_0$  and the correctness of the responses from the 150 trials were used to calculate the posterior probability across the three-dimensional parameter space. Collapsing the posterior distribution onto the  $K$ ,  $r$ , or  $p$  dimensions, the final parameter estimates were derived,

according to the resulting marginal parameter distributions, as the means of  $10 \log(K)$ ,  $10 \log(r)$ , and  $p$ . The 2.5th and 97.5th percentiles of the marginal distributions were used as the 95% credibility limits of the parameters.

For the threshold-based procedure, detection thresholds for the 2-kHz signal were estimated using a 2-down, 1-up, staircase procedure (Levitt, 1971). The signal level was initially 70 dB SPL and was altered in 8-dB steps. The step size was changed to 5 dB after two reversals and to 2 dB after four reversals. The signal level was not allowed to exceed 80 dB SPL, nor fall below 0 dB SPL. During experiment I, the signal level reached the upper limit (80 dB SPL) on only one occasion for listener S4 and never reached the lower limit (0 dB SPL). The procedure was terminated after 50 trials. The threshold estimate for each set of 50 trials was calculated from the average of the reversals when the step size was 2 dB unless the number of reversals was odd, in which case the first value was removed from the average. If fewer than four reversals were available for the threshold estimate, as happened twice, the staircase was repeated.

Five values of  $\Delta f/f_0$  were included in the threshold-based procedure: 0.05, 0.1, 0.2, 0.4, and 0.6. Thresholds were estimated for each value of  $\Delta f/f_0$  in turn, but in random order. This process was repeated four times, and those values were averaged to generate overall thresholds. The average thresholds across the four replicates as a function of the value of  $\Delta f/f_0$  were fitted to Eq. (3) using *fminsearch* in MATLAB (The MathWorks, Inc.). A Monte-Carlo procedure was conducted to provide bootstrap estimates of the expected error in the parameter estimates. This procedure repeatedly fitted the model in Eq. (3) to resampled data 250 times. For each repetition, the resampled data were obtained by first drawing four threshold estimates from the original four thresholds with replacement (an original threshold estimate could be repeatedly sampled) and then averaging them for each value of  $\Delta f/f_0$ . The Monte-Carlo procedure provided 250 estimates for each of the model parameters ( $K$ ,  $r$ , and  $p$ ). The resulting distributions were not always normal for the  $K$  and  $r$  parameters, so the standard errors of the estimates for all three parameters were described in terms of the parameter values at the 2.5th and 97.5th percentiles of the bootstrap replicates.

The estimated AF parameters were analyzed in terms of the agreement among the multiple qAF runs, and the agreement between the qAF and threshold-based procedures. Additionally the parameter estimates obtained from the qAF procedure were used to predict the signal level required for 71% correct. The predicted thresholds were compared with the thresholds obtained using the threshold-based procedure.

## B. Results

### 1. Agreement among qAF runs

Figure 1 plots the stimulus and model parameters as functions of trial number during the fourth run of the qAF procedure for listener S1. As responses were collected, all three model parameters ( $K$ ,  $r$ , and  $p$ ) converged rapidly during the first 50 trials (right panels). The efficient estimation of these model parameters was achieved by the

computational optimization in stimulus sampling. For the results shown in Fig. 1, the stimuli were sampled across the full range of values for  $\Delta f/f_0$  (bottom left panel) and the signal levels were mainly sampled between 25 and 50 dB SPL (top left panel). The distribution of the stimuli tested is visualized in the two dimensional stimulus space in Fig. 2 for the same qAF run as shown in Fig. 1. In Fig. 2, the circles indicate the locations in the stimulus space that were visited and the sizes of the circles indicate the number of visits. The solid curves in the figure are the estimated iso-performance contours (at proportions correct of 0.6, 0.75, and 0.9) of the two-dimensional psychometric function [Eq. (4)] estimated after 150 trials. For most of the trials, the stimulus was presented near these iso-performance contours. That is, stimulus sampling avoided regions of the stimulus space where performance was expected to be 100% correct or near chance. The efficiency of the qAF procedure is supported by the fact that sampled stimuli are within the range of the psychometric function.

To obtain final parameter estimates from the raw trial-by-trial data, the stimulus parameters and responses collected from all trials were fitted using a model with fine-grid parameter space (see details in Sec. III A 3). This analysis also enabled a visualization of the posterior parameter distribution and an estimation of the credibility limits. The left panels of Fig. 3 plot the posterior parameter distributions after 150 trials for the same qAF run as shown in Figs. 1 and 2 for S1. The three panels, from top to bottom, correspond to the posterior distribution collapsed across the  $K$ ,  $r$ , and  $p$  dimensions. Darker areas in the figure indicate higher probability density. For parameters  $r$  and  $p$  (top left panel), the posterior distribution was fairly compact after 150 trials. The 95% Bayesian confidence limits, or credibility limits (indicated using dashed lines), for these two parameters were relatively close to the mean of the distribution, which was also the final parameter estimates (indicated using a cross). Moreover, a posterior covariance between  $10\log(K)$  and  $10\log(r)$  was observed, suggesting that the estimates of the  $K$

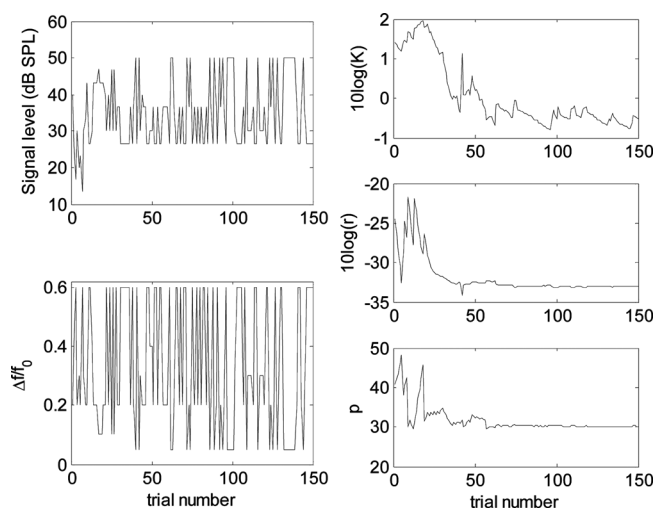


FIG. 1. The results of 150 qAF trials are plotted for S1. The left panel shows the values of signal strength ( $x$ , upper) and notch bandwidth ( $\Delta f/f_0$ , lower) tested on each trial. The right panels indicate the estimated value of  $K$ ,  $r$ , and  $p$  (top to bottom) following each trial.

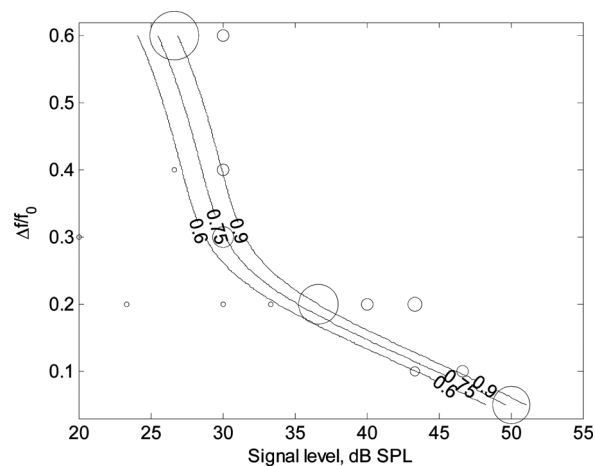


FIG. 2. A portion of the two-dimensional psychometric function estimated from the data of S1 is shown. Circles indicate locations in the signal-level  $\times$  notch-bandwidth stimulus space visited by the qAF procedure. Larger circles indicate a larger number of trials at that point in the stimulus space.

and  $r$  parameters were not independent of each other during the qAF procedure.

The right panels of Fig. 3 illustrate a examples of the posterior parameter distribution when the parameters occasionally failed to converge during a qAF run (the third qAF run for S4). In this case, the credibility limits were fairly broad, and the covariance among the parameters was larger compared to the examples in the left panels. The strong inter-parameter covariance can be interpreted as a consequence of sub-optimal sampling of the stimuli. That is, at very narrow notch widths, poor sensitivity could be explained by a small value of  $p$  or a large value of  $K$ , while at very large notch widths, poor sensitivity could be well captured by a high value of either  $r$  or  $K$ . Therefore, if the notch widths are only sampled at the two extremes (near  $\Delta f/f_0$ 's of 0 and 0.6 for the current experiments), the collected responses can be equally well fitted using models with large or small values of  $K$ . For the models with relatively high values of  $K$ , the estimated  $p$  would be relatively large and the estimated  $r$  would be relatively small, leading to posterior covariance among the parameters. On the other hand, if, besides the extreme notch widths, the stimuli are also sampled at the transition region between the sloping portion (controlled by the  $p$  parameter) and the flat portion (controlled by the  $r$  parameter) of the auditory filter, the values of  $p$  and  $r$  can be uniquely determined, leading to a unique solution of  $K$  when fitting the model to the data.

The example run summarized in Figs. 1 and 2, and the left column of Fig. 3, was representative of most qAF runs measured during runs 3 and 4. For runs 1 and 2, similar observations were made, except that the procedure sampled  $\Delta f/f_0$  mostly at small values ( $\leq 0.2$ ), potentially because the range of  $x$  was limited between 30 and 80 dB SPL during runs 1 and 2.

The parameter estimates obtained from the four runs of the qAF procedure are shown in Table I for all listeners but S3. For S3, only estimates from runs 1 and 2 are listed. Besides the three model parameters  $K$ ,  $r$ , and  $p$ , the AF bandwidth derived from the parameter  $p$  is also listed.<sup>3</sup> The

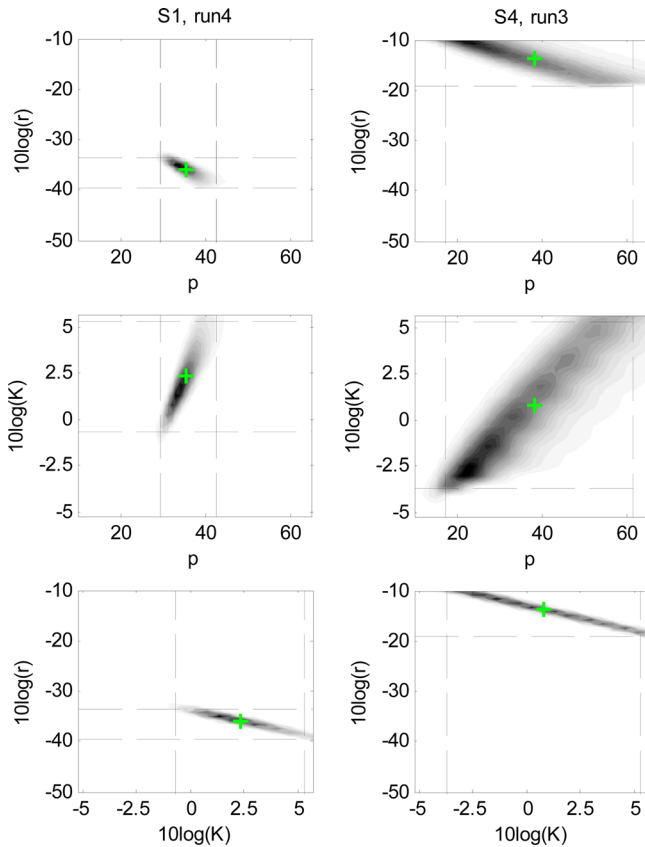


FIG. 3. (Color online) The posterior parameter distributions for the fourth qAF run for listener S1 (left) and the third qAF run for listener S4 (right). For the two runs shown here, the posterior distribution has been collapsed across the  $K$ ,  $r$ , and  $p$  dimensions in the top, middle, and bottom panels, respectively. The dashed lines mark the 95% credibility limits for the model parameters based on their marginal distributions and the cross indicates the mean (which is also the parameter estimate).

estimates of the AF bandwidth are expressed in terms of equivalent rectangular bandwidth (ERB). The values on the left side of the table indicate estimates based on the first 50 trials and the values on the right side are based on 150 trials. The values of  $10\log(K)$ ,  $10\log(r)$ , and  $p$  were near 0,  $-40$ , and  $30$ , after the first 50 trials and stayed relatively unchanged as more trials were run. These parameter estimates fell within the ranges of values expected for a young normal-hearing population (e.g., Wright, 1996).

When comparing the predominant parameter of interest, ERB, across the multiple runs, the ERB estimates were fairly consistent across runs for most of the listeners after 150 trials. The standard deviations of the ERB estimates across runs ranged from 9.7 Hz (for S6) to 68.6 Hz (for S5). Listeners S4 and S5 showed relatively more variable estimates. In particular, the ERB estimates during the first qAF run for S4 and second qAF run for S5 were much larger than the estimates obtained from the other runs for these listeners. The fact that the ERB estimates did not differ markedly across runs for most of the listeners suggests that the ERB estimates were not sensitive to whether the listener had previous listening experience, and/or that variations in the parameter initialization were inconsequential, and/or the variation in computational procedure was not important.

TABLE I. Values of  $K$ ,  $r$ , and  $p$  based on data collected using the qAF procedure with 50 (left) and 150 (right) trials. Results for four runs are shown (except for S3).

	First 50 Trials				150 Trials			
	$10\log(K)$	$10\log(r)$	$p$	ERB	$10\log(K)$	$10\log(r)$	$p$	ERB
S1	3.54	-47.6	35.1	228	3.59	-47.8	34.7	230
	2.68	-45.7	33.4	239	2.50	-44.0	34.6	232
	0.19	-36.7	24.0	334	1.26	-37.3	27.1	295
	2.09	-35.2	35.8	223	2.37	-36.0	35.4	226
S2	1.11	-41.6	28.4	282	0.72	-41.0	28.1	285
	2.28	-42.0	30.1	266	2.28	-41.4	29.5	271
	-2.85	-35.4	21.8	367	-0.97	-34.6	27.7	288
	-2.17	-39.2	25.7	311	-4.05	-37.3	24.5	326
S3	0.38	-47.1	39.3	203	3.26	-48.1	41.5	193
	1.72	-48.9	39.8	201	1.75	-48.6	39.0	205
S4	3.45	-41.5	23.8	336	0.45	-35.7	22.4	356
	2.11	-41.2	29.5	271	2.65	-42.0	29.6	271
	-0.62	-13.1	44.8	178	0.77	-13.8	38.2	210
	-0.95	-34.5	25.9	309	1.64	-36.3	27.7	289
S5	3.05	-46.7	34.2	234	3.77	-48.9	35.4	226
	2.42	-41.3	23.1	346	2.20	-40.1	22.9	350
	3.22	-48.0	37.7	212	4.05	-48.2	38.6	207
	1.78	-54.1	23.5	340	3.50	-52.6	25.5	314
S6	2.85	-45.8	33.4	240	2.62	-45.3	33.1	242
	-1.03	-43.8	31.3	256	1.45	-44.4	33.2	241
	-1.30	-43.8	31.5	254	-0.39	-43.4	35.3	227
	-0.81	-43.9	37.9	211	-3.55	-43.0	32.0	250

After 150 trials, the estimates of  $10\log(r)$  and  $10\log(K)$  were fairly consistent across runs. For the  $r$  parameter, the standard deviations of  $10\log(r)$  across runs was less than 10 for all listeners but S4, for whom a large value of  $10\log(r)$  was obtained from the third run ( $-13.8$ ). For the  $K$  parameter, the standard deviation of  $10\log(K)$  ranged from 0.82 (for S5) to 2.71 (for S2). For all listeners and all runs, the parameter estimates after the first 50 trials were highly predictive of the parameter estimates after 150 trial (e.g.,  $R^2=0.81$ ,  $p<0.01$  for the ERB estimates). This suggests that the qAF ERB estimates had converged after 50 trials.

In summary, the multiple runs of the qAF procedure provided similar estimates of the model parameters. Because the data from runs 3–4 were free of potential numerical errors, the results from these runs were used to compare the qAF and threshold-based procedures.

## 2. Agreement between the qAF and threshold-based procedures

Table II lists the values of  $K$ ,  $r$ ,  $p$ , and ERB, based on fits to the thresholds measured using the threshold-based procedure (bold), with the 95% confidence limits shown in parentheses. The values of  $K$  and  $r$  are expressed as  $10\log(K)$  and  $10\log(r)$ , which were translated after averaging and after the bootstrap simulation. The parameters estimated from qAF runs 3–4 (150 trials) are repeated from Table I with their 95% credibility limits given in parentheses. When a parameter estimate from the qAF procedure was outside of the confidence limits for the parameter derived using the

TABLE II. Values of  $K$ ,  $r$ , and  $p$  from data collected using the staircase procedure (Bold). Values in parentheses indicate the 95% confidence limits of the parameter estimates. Also listed are the corresponding estimates obtained from qAF runs 3 and 4 (not available for listener S3) with the 95% credibility limits in parentheses.

		10log( $K$ )	10log( $r$ )	$p$	ERB
S1		<b>-4.1 (-9.5 -0.69)</b>	<b>-32.8 (-36.3 -29.5)</b>	<b>26.4 (16.4 32.7)</b>	<b>303 (245 487)</b>
S1 (qAF)	run3	1.26 <sup>a</sup> (-0.8 2.8)	-37.3 <sup>a</sup> (-39.7 -35.8)	27.1 (23.4 30.2)	295 (265 342)
	run4	2.37 <sup>a</sup> (-0.7 5.3)	-36.0 <sup>a</sup> (-39.6 -33.6)	35.4 <sup>a</sup> (29.3 42.4)	226 <sup>a</sup> (189 273)
S2		<b>-2.1 (-4.8 -0.1)</b>	<b>-34.7 (-36.6 -33.2)</b>	<b>26.6 (20.7 31.0)</b>	<b>301 (258 386)</b>
S2 (qAF)	run3	-0.97 (-4.9 2.0)	-34.6 (-38.2 -31.5)	27.7 (22.7 31.8)	288 (251 353)
	run4	-4.05 (-5.2 -2.2)	-37.3 <sup>a</sup> (-40.0 -35.8)	24.5 (22.2 26.3)	326 (304 361)
S3		<b>2.0 (0.5 3.4)</b>	<b>-48.5 (-50.5 -46.8)</b>	<b>39.7 (37.5 42.2)</b>	<b>201 (190 213)</b>
S4		<b>6.6 (-6.7 11.9)</b>	<b>-33.8 (-169.7 -25.9)</b>	<b>40.1 (12.4 51.9)</b>	<b>199 (154 646)</b>
S4 (qAF)	run3	0.77 (-3.7 5.3)	-13.8 <sup>a</sup> (-19.0 -10.2)	38.2 (17.4 61.2)	210 (131 461)
	run4	1.64 (-0.5 3.5)	-36.3 (-38.7 -34.7)	27.7 (23.1 32.6)	289 (246 347)
S5		<b>5.9 (0.9 9.1)</b>	<b>-48.3 (-53.4 -44.7)</b>	<b>28.9 (24.5 33.0)</b>	<b>277 (243 326)</b>
S5 (qAF)	run3	4.05 (1.4 5.6)	-48.2 (-50.7 -46.1)	38.6 <sup>a</sup> (34.6 41.5)	207 <sup>a</sup> (193 231)
	run4	3.50 (0.8 5.5)	-52.6 (-55.5 -50.5)	25.5 (22.1 27.7)	314 (289 362)
S6		<b>0.1 (-3.3 1.4)</b>	<b>-44.5 (-46.2 -42.7)</b>	<b>31.9 (24.9 34.3)</b>	<b>251 (233 321)</b>
S6 (qAF)	run3	-0.39 (-4.8 3.7)	-43.4 (-48.2 -40.0)	35.3 <sup>a</sup> (30.0 40.0)	227 <sup>a</sup> (201 267)
	run4	-3.55 (-5.2 -0.5)	-43.0 (-46.8 -41.0)	32.0 (29.0 35.4)	250 (226 275)

<sup>a</sup>qAF parameter estimates that exceed the 95% confidence limits around the mean estimates based on threshold-based measures.

threshold-based procedure, a disagreement between the two procedures is indicated using an asterisk.

Comparing the ERB estimates (or equivalently estimates of  $p$ ) between the two procedures, disagreement occurred in three out of ten comparisons (the fourth qAF run for S1 and the third qAF runs for S5 and S6). For listeners S2 and S6, on an absolute frequency scale, the differences between the qAF and threshold-based estimates were never larger than 25 Hz. For listeners S1, S4, and S5, the differences between the two procedures were somewhat larger. The ERB estimates from these three listeners included wide confidence limits for the threshold-based procedure, wide credibility limits for the qAF procedure, and inconsistencies between the third and fourth qAF runs. As an example, the right panel of Fig. 3 plots the posterior parameter distribution in the third run for S4, where the 95% credibility limit for ERB was found to be the broadest among all qAF runs. None of the three parameters appear to have converged after 150 trials and the 95% credibility limits were, in most cases, very close to the defined boundaries of the parameter space. Note that among all listeners, this listener (S4) also had the largest 95% confidence limits in the ERB estimate for the threshold-based procedure. Further inspections of this listener's data suggested that this listener produced incorrect responses for very high signal levels and very wide notch bandwidths far more frequently than other listeners. As a consequence, unreliable estimates of ERB were obtained for both the threshold-based and qAF procedures. The response errors on trials with high signal levels and wide notch bandwidth, often occurring at the beginning of the adaptive tracks, may reflect frequent lapses of attention during the experiment. Potentially the qAF procedure is sensitive to attentional lapses and early response errors. This issue is investigated in greater detail in experiments II and III.

Besides the ERB estimates, disagreements across runs were detected for two out of ten comparisons for the parameter  $K$ . The largest difference between the 10log( $K$ ) estimates

from the two procedures ( $\sim 6.5$ ) occurred in the third qAF run for S1. For the parameter  $r$ , disagreements were observed for four out of ten comparisons. The difference between the 10log( $r$ ) estimates using the threshold-based and qAF procedures was always less than 5 dB except for the third run for listener S4 (see the posterior parameter distribution for this qAF run in the right panel of Fig. 3), where a difference of 20 dB was obtained.

To provide a second comparison between the results for the threshold-based and qAF procedures, the models estimated by the qAF procedure were used to calculate the signal level required for 71% correct, i.e., the model provided predictions of the thresholds obtained using the threshold-based procedure.

Figure 4 plots the resulting predicted thresholds (ordinate) as a function of the measured thresholds (abscissa). Results based on the models provided by the third and fourth qAF runs are shown in the left and right panels, respectively. Each panel contains measured (using the threshold-based procedure) and predicted (using the qAF procedure) thresholds for five different values of  $\Delta f/f_0$  for five listeners (all listeners in experiment I but S3). For each qAF run, results based on the first 50 and 150 trials are indicated using unfilled larger circles and smaller filled circles, respectively. Predicted and measured thresholds were significantly correlated, with  $R^2$  values of 0.80 ( $p < 0.01$ ) and 0.79 ( $p < 0.01$ ) after 50 and 150 trials for the third qAF runs, and with  $R^2$  values of 0.91 ( $p < 0.01$ ) and 0.93 ( $p < 0.01$ ) after 50 and 150 trials for the fourth qAF runs. The qAF procedure appeared to provide good agreement with the threshold-based procedure after only 50 trials.

The above comparison showed that fair agreement between the qAF and threshold-based procedures was achieved, despite the threshold-based procedure requiring more than 1000 trials of data collection and more than 2 h to finish, while each qAF run took approximately 10 min.

#### IV. EXPERIMENT II: ROBUSTNESS OF THE QAF PROCEDURE

Experiment II was conducted to test the robustness of the qAF procedure. Three factors potentially capable of influencing the AF estimates using the qAF procedure were studied. First, the effect of the range of the stimulus parameter  $x$  was investigated. When implementing the qAF procedure as described in Sec. II, the signal level at the beginning of a qAF run was always set at the center of the  $x$  parameter range. Therefore, when the  $x$  range was shifted, the initial signal level changed, which might affect the convergence of the model parameters ( $K$ ,  $r$ , and  $p$ ). In experiment I, for the first two qAF runs  $x$  was between 30 and 80 dB SPL, while for the last two runs  $x$  ranged from 10 to 60 dB SPL. The results from experiment I failed to suggest a systematic influence of the range of  $x$  (or the initial signal level) on the parameter estimates. However, since numerical errors could have influenced the results from the first two qAF runs of experiment I, an influence of range of  $x$  cannot be excluded. The qAF runs were repeated in conditions 1 and 2 of the current experiment, which shared identical parameter ranges and gradation except for the range for  $x$ . The effect of the range of  $x$  was studied by comparing the results for conditions 1 and 2.

Second, for the qAF procedure considered here, the slope of the psychometric function (the  $\beta$  parameter) is fixed. In experiment I, the value of  $\beta$  was 1 for all four runs. In practice, the actual psychometric-function slopes are often unknown and likely to deviate from the assumed value. It is not clear whether the assumed  $\beta$  value of 1 in experiment I was a good choice. Therefore, the assumed value of  $\beta$  was manipulated across three conditions (conditions 1, 3, and 4) of the current experiment. The results provided empirical guidelines regarding the choice of  $\beta$  in the model.

Last, it has been reported that Bayesian adaptive procedures, such as the qAF procedure developed here, could be sensitive to response errors made early in a track (K. Saberi, personal communication; Kontsevich and Tyler, 1999). The results of experiment I for S4 suggested that this may be true for the qAF procedure. This is mainly because the choice of the stimuli on a given trial depends on the stimuli and

responses for all previous trials. Computer simulations suggested that when incorrect responses occurred on the first few trials of a run, the convergence of the AF parameters could be extremely slow. During these first trials, the signal level typically rose well above threshold which led to large shifts in the interim parameter estimates, inefficient stimulus sampling, and consequently little improvement of the parameter estimates on subsequent trials. The current experiment included a condition (condition 5) in which incorrect responses were reported to the qAF algorithm during the first three trials regardless of the actual responses from the listener. By comparing results for conditions 1 and 5, the effect of early errors on the qAF procedure was investigated.

#### A. Methods

The listeners from experiment I, except for S3, participated in this experiment. A new listener S7 was recruited. This listener had substantial experience in psychoacoustic tasks, but no experience in tone-in-noise experiments. Her absolute thresholds were 20 dB HL or better for audiometric frequencies from 250 to 8000 Hz.

In condition 1, the stimuli, procedures, and the initialization of the parameters were identical to those for the qAF runs 3–4 in experiment I. In particular, the potential signal levels  $x$  ranged from 10 to 60 dB with 3.3 dB spacing, and  $\beta$  was set to 1. Condition 2 was identical to condition 1 except that the potential signal levels ranged from 30 to 80 dB. Therefore, the effect of changes in the range of  $x$  would be indicated as differences between the results for conditions 1 and 2. Conditions 3 and 4 were identical to condition 1 except that the values of  $\beta$  were 0.5 (relatively shallow) and 1.5 (relatively steep) for conditions 3 and 4, respectively. Therefore, the effect of the choice of  $\beta$  would be indicated as differences in the results for conditions 1, 3, and 4. Condition 5 was identical to condition 1 except that the responses from the first three trials of each qAF run were forced to be incorrect. Therefore, the effect of early errors would be indicated as differences in the results of conditions 1 and 5. In condition 5, although incorrect responses were provided to the pAF algorithm, accurate feedback was provided to the listener.

For all listeners, five qAF runs that corresponded to the five conditions were first tested in random order. Each qAF run contained 150 trials, blocked into three sets of 50 trials. Then, the process was repeated in the same order.

#### B. Results and discussion

Because the AF bandwidth was the predominant parameter of interest, data analyses focused on the ERB estimates. First consider conditions 1–4 (see Table III). The average ERB estimates across the six listeners and across the two runs were 280, 314, 269, and 280 Hz for conditions 1, 2, 3, and 4, respectively. Of the 48 ERB estimates (six listeners  $\times$  four conditions  $\times$  two runs), six appeared to be unreliable (indicated by asterisks in Table III). That is, they were more than 2 standard deviations from the ERB expected for a young, normal-hearing population at 2 kHz, according to the population mean and standard deviation of 255 and 46 Hz

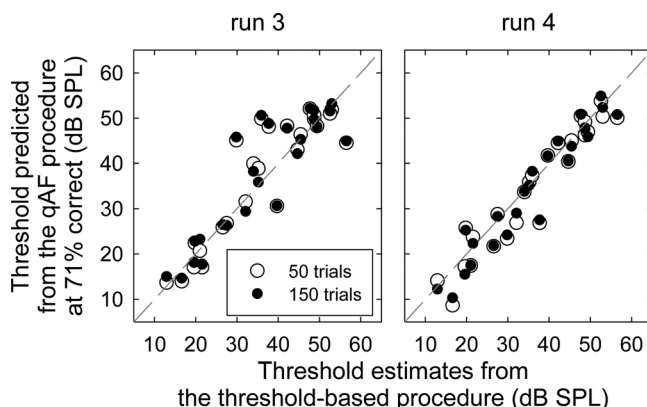


FIG. 4. Thresholds predicted based on the qAF model after 50 (unfilled circles) and 150 (filled circles) trials are plotted as a function of the observed thresholds. Only thresholds near or larger than 25 dB SPL were evaluated. The dashed line indicates equal estimates.



estimated by Wright (1996) from 80 normal-hearing ears at a masker spectrum level of 40 dB SPL.

To quantify the effect of condition on the ERB estimates, a grand average ERB estimate was calculated for each listener across all conditions and both runs. Then, the root-mean-squared (rms) deviation from the ERB estimates for each condition to the grand average for the ERB estimate was derived for that listener. A large rms deviation would indicate the condition's ERB estimate was deviant. Pair-wise comparisons (Bonferroni corrected) failed to indicate significant rms differences for any one condition relative to any other condition. Therefore, the manipulations of the range of  $x$  and the value of  $\beta$  in the current experiment did not appear to generate systematic changes in the ERB estimates.

In condition 5 the goal was to learn whether incorrect responses on the first three trials would cause the procedure to fail. The results indicated that this in fact occurred. Of the 12 ERB estimates, 7 were found to be unreliable ("a" under condition 5 in Table III), exceeding the expected mean ERB for a normal-hearing population by more than 2 standard deviations (Wright, 1996). The increased proportion of unreliable ERB estimates in condition 5 compared to condition 1 indicates that the qAF procedure is sensitive to early response errors.

Given the unreliability of the qAF procedure in condition 5, it seemed advantageous to modify the qAF procedure to make the procedure resilient to early errors. One potential way of modifying the qAF procedure would be to include an additional parameter reflecting lapses of attention. The logic was that errors, particularly errors early in the procedure, had the effect of shifting the assumed psychometric function too far toward large signal levels. If the psychometric function was allowed an asymptote at a percent correct values less than 100 (e.g., Wichmann and Hill, 2001; Dai and Michey, 2011; Shen and Richards, 2012), the effects of early errors, and inattention in general, might be ameliorated. Computer simulations indicated that incorporating a psychometric function with a lapse parameter into the model led to rapid convergence of the model parameters, even when the first few trials were forced to be incorrect. The effectiveness of introducing a

lapse parameter into the model was tested in experiment III using young, normal-hearing listeners.

## V. EXPERIMENT III: QAF WITH A FOUR-PARAMETER MODEL

### A. Methods

In experiment IIIa, listeners S2 and S4 repeated conditions 1 and 5 of experiment II. For consistency, these conditions will also be referred to as conditions 1 and 5 in experiment III. Recall that in condition 5 the first three responses during the qAF runs were forced to be incorrect. The stimuli, procedures, and initialization of the parameters were identical to those used in experiment II with the following exceptions. First, the qAF procedure was modified, with an additional parameter corresponding to the lapse rate,  $\lambda$ , introduced to the model in Eq. (4), so

$$PC(x, \Delta f/f_0; K, r, p) = \gamma + (1 - \gamma - \lambda) / [1 + e^{-\beta(x - 10 \log P_s)}]. \quad (5)$$

In all other aspects, the qAF procedure remained the same as previously described. With the introduction of  $\lambda$ , the model parameter space became four dimensional ( $K, r, p, \lambda$ ), while the stimulus parameter space remained two dimensional ( $x$  and  $\Delta f/f_0$ ). The potential values for the new parameter  $\lambda$  were 0, 0.05, 0.1, 0.15, and 0.2. The two listeners completed one qAF run of conditions 1 and 5 in different order, and then re-ran those conditions in the opposite order. Each qAF run contained 200 trials, broken into four 50-trial blocks.

Twenty-one listeners were recruited for experiment IIIb from the undergraduate population of the University of California, Irvine. These listener's ages were between 18 and 29 years, and all had audiometric thresholds of 20 HL or better between 250 and 4000 Hz, except for one participant whose thresholds in the left ear were 35 and 25 dB HL at 250 and 500 Hz.<sup>4</sup> The left ears were tested. None of the listeners had previous experience in psychoacoustic experiments. The experiment was completed in a single sessions of 1 h. As compensation, the listeners received course credit.

The listeners were divided into two groups. Group 1 consisted of 13 listeners (S8-S20, four male and nine female). For this group no lapse parameter was incorporated into the model. Two 200-trial qAF runs were run sequentially. The stimuli, procedures, and initialization of the parameters were identical to those for conditions 1 and 5 in experiment II. Six of the listeners in Group 1, determined in a quasi-random fashion, started condition 1 before condition 5, while the other five listeners were run in condition 5 first.

Group 2 consisted of eight listeners (S21-S28, one male and seven female). For this group the model included a lapse parameter. Two qAF runs each containing 250 trials were run using five blocks of 50 trials. Slightly larger numbers of trials were run because we were unsure of the convergence properties when four rather than three parameters were to be estimated. The stimuli, procedures, and initialization of the parameters were identical to those used for Group 1, except that the four-parameter model was used in the qAF

TABLE III. Values of ERB estimates for conditions 1–4 of experiment II.

		ERB estimates (Hz)				
		Condition 1	Condition 2	Condition 3	Condition 4	Condition 5
S1	run1	294.8	291.2	250.9	404.9 <sup>a</sup>	178.5
	run2	225.7	213.5	296.4	230.5	157.8 <sup>a</sup>
S2	run1	288.4	283.0	253.5	375.1 <sup>a</sup>	177.0
	run2	326.1	274.6	296.8	258.5	152.3 <sup>a</sup>
S4	run1	209.5	500.4 <sup>a</sup>	294.5	268.7	234.5
	run2	288.6	348.4 <sup>a</sup>	274.5	295.1	354.7 <sup>a</sup>
S5	run1	207.0	210.8	241.6	265.1	411.0 <sup>a</sup>
	run2	313.7	269.0	219.0	308.1	461.2 <sup>a</sup>
S6	run1	226.9	271.6	256.9	242.0	165.0
	run2	250.3	248.9	234.2	232.3	145.7 <sup>a</sup>
S7	run1	509.4 <sup>a</sup>	278.1	240.4	259.8	161.1 <sup>a</sup>
	run2	224.9	580.0 <sup>a</sup>	273.5	233.5	188.8

<sup>a</sup>Values are taken as outliers (see text).

procedure and the lapse parameter,  $\lambda$ , was estimated. The range for  $\lambda$  was identical to that used in experiment IIIa.

## B. Results

As for experiment II, the comparison between the two conditions focuses on the ERB estimates.

Figure 5 plots the estimated values of the ERBs for S2 (left) and S4 (right) as a function of the number of trials. The top panels are from experiment II and the bottom panels are from experiment IIIa. The black symbols are for condition 1, the gray symbols are for condition 5, and the filled and unfilled symbols indicate the data drawn from the first and second runs, respectively. Without a lapse parameter (upper panels), the results for S2 were consistent across runs, but different across conditions. For S4, the results were inconsistent across both runs and conditions, and one of the ERB estimates for condition 5 was much larger than the others. For both listeners, introducing a lapse parameter (bottom panel) resulted in ERB estimates that were similar both across conditions and across runs. More importantly, using the four-parameter model, the ERB estimates in condition 5 became more consistent with the ERB estimates obtained using the threshold-based procedure in experiment I (diamonds). This latter result was less apparent for S4 than for S2, but recall that the error in the ERB estimate from the threshold-based procedure was very large for S4 (Table I). In short, the modified qAF procedure with the inclusion of the lapse parameter reduced the variability in the ERB estimates, whether caused by forced early response errors or not. Averaging across the two runs, the estimated values of  $\lambda$  were 0 and 0.08 for S2 in conditions 1 and 5, respectively, and were 0.16 and 0.16 for S4 in conditions 1 and 5, respectively.

The left and right panels of Fig. 6 show the results for experiment IIIb. The scatter plots show the estimated ERBs in condition 5 plotted against the estimated ERBs in condition 1 for Groups 1 and 2 separately. For Group 1 (left panel), the failure associated with the ERB estimate in condition 5 was apparent in the lower estimates. This result was observed both after the first 100 trials (unfilled symbols) and after 200 trials (filled symbols). In some instances, the qAF procedure failed in condition 1. One listener (S9) had an ERB estimate of 716 Hz in condition 1 after 200 trials, which was three standard deviations away from the mean ERB estimate across all 13 listeners for the same condition. As a result, the data for this listener are far beyond the range of values plotted in Fig. 4 (the only data point that is not shown in Fig. 4). The correlation between ERB estimates for conditions 1 and 5 at the end of 200 trials did not approach significance ( $R^2 = 0.04$ ,  $p = 0.49$ ). Therefore, estimates of ERBs from these two conditions were in poor agreement.

The results for Group 2 (right panel) indicated that the inclusion of a lapse parameter provided ERB estimates that were similar across listeners and consistent across conditions 1 and 5. The correlation between the ERB estimates for the two conditions after 250 trials was statistically significant ( $R^2 = 0.68$ ,  $p = 0.01$ ). Moreover, the procedure appears to have converged reasonably well after as few as 125 trials. In

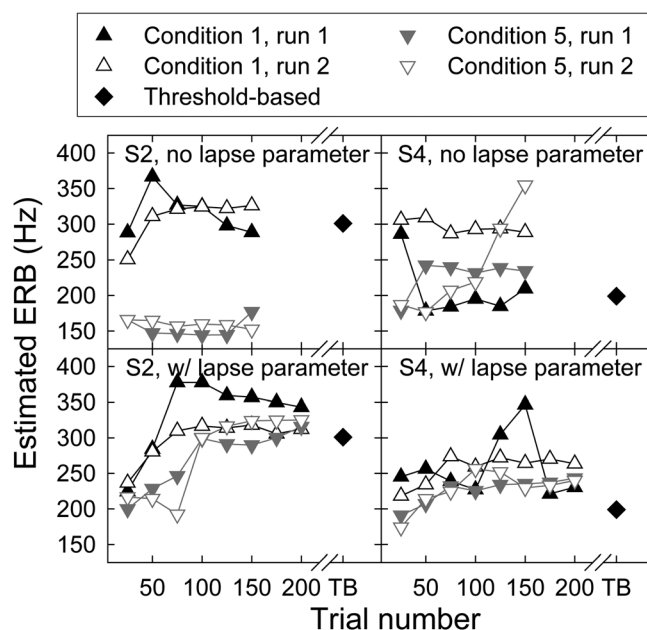


FIG. 5. Estimates of ERBs are plotted for (left) S2 and (right) S4 for experiments (top) II and (bottom) IIIa as functions of the number of trials. The black symbols are for condition 1 and the gray symbols are for condition 5. The first and second runs are indicated using filled and unfilled symbols, respectively. The diamond in each panel indicates the corresponding ERB estimate obtained using the threshold-based procedure from experiment I.

condition 1 the estimated values of  $\lambda$  ranged from 0 to 0.10 with an average of 0.06, and in condition 5 the values of  $\lambda$  ranged from 0.06 to 0.22 with an average of 0.13. The larger value of  $\lambda$  in condition 5 ( $t_{14} = 2.8$ ,  $p = 0.02$ ) was expected because the first three trials were forced to be incorrect.

In summary, experiment IIIb showed that for naive listeners the ERB estimates obtained with the qAF procedure without the lapse parameter were variable and sensitive to early response errors. In contrast, the ERB estimates from the qAF procedure with the four-parameter model were more reliable and robust against early errors.

## VI. DISCUSSION

For the estimation of the parameters of the  $roex(p,r)$  AF shape, the qAF procedure appeared to converge within 50–100 trials for approximately 85% of the runs in experiment I. The resulting parameter estimates were typically comparable with those obtained using a threshold-based procedure and 1000 trials. Where deviations were observed, the reliability of the parameter estimates tended to be poor for both methods, but somewhat superior for the threshold-based procedure. The results of experiment II indicated that the parameters estimated by the qAF procedure were relatively invariant with respect to the ranges of stimulus variables tested and that early response errors are almost assured to produce unreliable estimates of the ERB. The results of experiment III indicated that by introducing a lapse parameter into the model of the psychometric function, the parameter estimates obtained from the qAF procedure were stable, even when early errors were artificially introduced into the procedure.

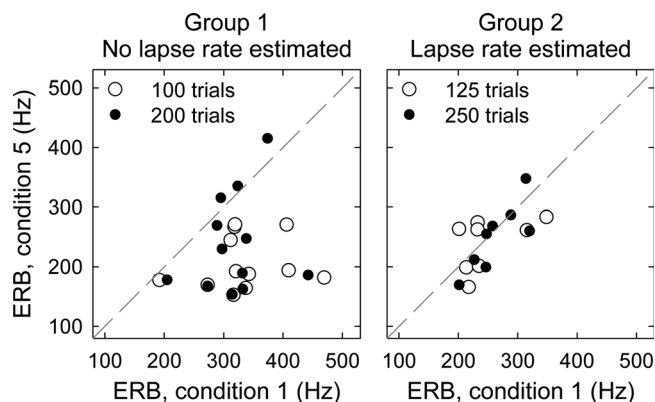


FIG. 6. Scatter plots of results for conditions 1 and 5 are plotted for (left) Group 1 and (right) Group 2 of experiment IIIb. The number of trials completed is indicated in the legends. The dashed line indicates equal estimates.

It should be noted that to simplify the experimental procedure and to reduce the number of parameters to be estimated, several assumptions were introduced. First, the AF was assumed to be symmetric about its center frequency on a linear frequency scale, which was a reasonable assumption given the moderate stimulus levels used in the current study. However, at high stimulus levels the shape of the AF becomes asymmetric with the low frequency skirt being shallower than the high frequency skirt. [Patterson and Nimmo-Smith \(1980\)](#) have suggested that the asymmetric shape of the AF at high levels could be captured by estimating the slope parameter  $p$  for the low- and high-frequency skirts of the AF separately (using parameters  $p_L$  and  $p_H$ , respectively). Future studies are needed to explore the possibility of extending the qAF procedure to asymmetric filter shapes.

The second major assumption made in the current study was the  $roex(r, p)$  formulation of the AF. This formulation had the advantage that it provided a simple analytical solution to the integration in Eq. (2). However, this does not mean that other forms of AF models cannot be considered. Widely used models, such as the *gammatone* filter (e.g., [de Boer, 1975](#); [Allerhand, et al., 1992](#); [Patterson, et al., 1995](#)) and the *gammachirp* filter (e.g., [Irinio and Patterson, 1997](#)), may also be implemented. For models that do not lead to straight forward analytical solutions of Eq. (2), the integration can be computed numerically. It should be appreciated, however, that when different models are tested, the geometry of the sample space, and of the parameter space, will change.

A third assumption made in the current study regards the form of the psychometric function. It was initially assumed that the psychometric function underlying listeners' performance was a logistic function with a fixed slope [the parameter  $\beta$  in Eq. (4)]. Other forms of the psychometric function, such as the cumulative Gaussian function, would probably serve the current purpose equally well. This is because the qAF procedure concentrates the stimuli near only one point on the psychometric function (the 75% correct point), making modest differences between the forms of psychometric functions relatively unimportant (e.g., slope change as in experiment II). Nonetheless, it is recommended that a lapse parameter be included in any formalization of

the psychometric functions to limit the vulnerability of the procedure to early response errors.

The participants of the current experiment included both naive and experienced subjects. The subjects in experiments I, II, and IIIa received financial compensation for their participation and were enrolled in the experiments across a period of a few weeks. On the other hand, the subjects in experiment IIIb participated for less than an hour to accrue extra credit for a course. It was possible that these two groups of listeners had different levels of motivation. The results, however, did not vary substantially across these groups. Both subject groups included listeners who apparently had frequent lapses in attention. Further experiments that systematically study the effect of inattention on the AF estimates should be conducted before the qAF procedure is used as a general-purpose tool to assess frequency selectivity in individuals across a wide range of ages and auditory capabilities.

Finally, it was assumed that the experimenter had little previous knowledge about the likelihood of the parameter values; therefore, weakly informative (or flat) prior distributions were used for all model parameters. To further improve the efficiency of the qAF, the experimenter should take advantage of the fact that the distributions of these model parameters are readily available in the psychoacoustic literature for both normal-hearing and hearing-impaired listener populations (e.g., [Patterson et al., 1982](#); [Moore, 1987](#); [Wright, 1996](#); [Badri et al., 2011](#)). For example, [Wright \(1996\)](#) measured the AF shape from 80 normal-hearing ears and reported the first four moments of the distribution (mean, standard deviation, skewness, and kurtosis) for each of the  $K$ ,  $p$ , and  $r$  parameters. Using these data, prior probabilities can be assigned to each grid point in the parameter space, which would, in principle, increase the rate of convergence of the parameters and reduce the potentially adverse impact of early incorrect responses on the stability of the adaptive track. When a lapse parameter ( $\lambda$ ) is used in the qAF procedure, the beta distribution might be a natural "uninformative" prior for  $\lambda$  because the lapse rate parameter is bounded by 0 and 1. Because the psychometric function [as in Eq. (5)] is only sensible if  $(1 - \gamma - \lambda > 0)$ , it seems appropriate to truncate the prior distribution for  $\lambda$  to be strictly less than  $1 - \gamma$ .

The qAF procedure introduced in the current study was developed to provide an efficient assessment of auditory spectral resolution. Alternative procedures have been proposed for this purpose that use a Békésy tracking procedure (e.g., [Şek et al., 2005](#); [Malicka et al., 2009](#); [Charaziak et al., 2012](#)). Compared to those methods, the qAF procedure takes a very different approach. It takes advantage of the fact that the form of the underlying model that governs auditory spectral resolution has been well studied; hence, the stimuli in a qAF run can be adjusted adaptively to improve the estimation of the model parameters directly. Because the qAF procedure does not require any modification to the original task design used in the traditional experiments, it might be preferred over the procedures based on the Békésy method in some situations.

It is worth pointing out that although our experiments demonstrated that the qAF procedure could be very efficient, completing the assessment of frequency resolution at one

signal frequency under 10 min, the current study only tested the procedure for young, normal-hearing listeners at a moderate stimulus level. It is not yet clear whether the stability and efficiency of the qAF procedure will remain when applied to other populations (e.g., hearing impaired listeners) or at high stimulus levels. Detailed investigation of these issues in future studies is required to enable the application of the qAF procedure for clinical use.

## ACKNOWLEDGMENTS

This work was supported by NIH NIDCD grant R21 DC010058 awarded to the second author. We thank Theodore S. Lin and Andrew S. Silva for assistance with data collection and analysis. Dr. Gerald Kidd, Jr. provided many helpful suggestions on a previous version of this manuscript.

<sup>1</sup>The current application involves a 3-D parameter space. There is no reason to expect that parameter spaces with relatively modest numbers of parameters would cause difficulty for this approach. For a large-dimensional parameter space, a sampling procedure that does not incorporate a grid search might be considered.

<sup>2</sup>Prior to introducing the spectral notch, the rms value of the masker was scaled to a constant value fixed across all conditions, meaning that relative to a true Gaussian noise there was less variance than expected.

<sup>3</sup>The ERB estimates reported in the current study were based on an approximation that  $ERB = 4f_0/p$ . This approximation has been shown to be accurate when the value of  $r$  is small. When the value of  $r$  exceeds -20 dB, the derivation of the exact ERB value should take  $r$  into account (Patterson *et al.*, 1982; Moore and Glasberg, 1983).

<sup>4</sup>For this listener, the mild hearing loss at the two lowest frequencies tested might suggest a conductive hearing loss. However, due to the limited availability (1 h) of this listener, no procedure was implemented to confirm the presence of a conductive loss.

Allerhand, M., Butterfield, S., Cutler, R., and Patterson, R. (1992). "Assessing syllable strength via an auditory model." *Proc. Inst. Acoust.* **14**, 297–304.

Badri, R., Siegel, J. H., and Wright, B. A. (2011). "Auditory filter shapes and high-frequency hearing in adults who have impaired speech in noise performance despite clinically normal audiograms." *J. Acoust. Soc. Am.* **129**, 852–863.

Charaziak, K. K., Souza, P., and Siegel, J. H. (2012). "Time-efficient measures of auditory frequency selectivity." *Int. J. Audiol.* **51**, 317–325.

Dai, H., and Micheyl, C. (2011). "Psychometric functions for pure-tone frequency discrimination." *J. Acoust. Soc. Am.* **130**, 263–272.

de Boer, E. (1975). "Synthetic whole-nerve action potentials for the cat." *J. Acoust. Soc. Am.* **58**, 1030–1045.

Fletcher, H. (1940). "Auditory patterns." *Rev. Mod. Phys.* **12**, 47–65.

Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data." *Hear. Res.* **47**, 103–138.

Green, D. M. (1990). "Stimulus selection in adaptive psychophysical procedures." *J. Acoust. Soc. Am.* **87**, 2662–2674.

Hopkins, K., and Moore, B. C. J. (2011). "The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise." *J. Acoust. Soc. Am.* **130**, 334–349.

Irino, T., and Patterson, R. D. (1997). "A time-domain, level-dependent auditory filter: The gammachirp." *J. Acoust. Soc. Am.* **101**, 412–419.

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

Kontsevich, L. L., and Tyler, C. W. (1999). "Bayesian adaptive estimation of psychometric slope and threshold." *Vis. Res.* **39**, 2729–2737.

Leek, M. R., Dubno, J. R., He, N., Ahlstrom, J. B. (2000). "Experience with a yes-no single-interval maximum-likelihood procedure." *J. Acoust. Soc. Am.* **107**, 2674–2684.

Leeuw, A. R., and Dreschler, W. A. (1994). "Frequency-resolution measurements with notched noise for clinical purposes." *Ear. Hear.* **15**, 240–255.

Lesmes, L. A., Jeon, S.-T., Lu, Z.-L., Doshier, B. A. (2006). "Bayesian adaptive estimation of threshold versus contrast external noise functions: The quick TvC method." *Vis. Res.* **46**, 3160–3176.

Lesmes, L. A., Lu, Z.-L., Baek, J., and Albright, T. D. (2010). "Bayesian adaptive estimation of threshold of contrast sensitivity function: The quick CSF method." *J. Vis.* **10**, 1–21.

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics." *J. Acoust. Soc. Am.* **49**, 466–477.

Malicka, A. N., Munro, K. J., and Baker, R. J. (2009). "Fast method for psychophysical tuning curve measurement in school-age children." *Int. J. Audiol.* **48**, 546–553.

Moore, B. C. J. (1987). "Distribution of auditory-filter bandwidths at 2 kHz in young normal listeners." *J. Acoust. Soc. Am.* **81**, 1633–1635.

Moore, B. C. J. (1995). "Frequency analysis and masking," in *Hearing*, edited by Brian C. J. Moore (Academic Press, San Diego, CA), Chap. 5, pp. 161–205.

Moore, B. C. J., and Glasberg, B. R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns." *J. Acoust. Soc. Am.* **74**, 750–753.

Nakaichi, T., Watanuki, K., and Sakamoto, S. (2003). "A simplified measurement method of auditory filters for hearing-impaired listeners." *Acoust. Sci. Technol.* **24**, 365–375.

Patterson, R. D. (1976). "Auditory filter shapes derived with noise stimuli." *J. Acoust. Soc. Am.* **59**, 640–654.

Patterson, R. D., Allerhand, M. H., and Giguère, C. (1995). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform." *J. Acoust. Soc. Am.* **98**, 1890–1894.

Patterson, R. D., and Nimmo-Smith, I. (1980). "Off-frequency listening and auditory-filter asymmetry." *J. Acoust. Soc. Am.* **67**, 229–245.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (1982). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold." *J. Acoust. Soc. Am.* **72**, 1788–1803.

Rosen, S., and Baker, R. J. (1984). "Characterising auditory filter non-linearity." *Hear. Res.* **73**, 231–243.

Sek, A., Alcántara, J., Moore, B. C. J., Kluk, K., and Wicher, A. (2005). "Development of a fast method for determining psychophysical tuning curves." *Int. J. Audiol.* **50**, 237–242.

Shen, Y. (2013). "Comparing adaptive procedures for estimating the psychometric function for an auditory gap detection task." *Atten. Percept. Psychophys.* **75**, 771–780.

Shen, Y., and Richards, V. M. (2012). "A maximum-likelihood procedure for estimating psychometric functions: Thresholds, slopes, and lapses of attention." *J. Acoust. Soc. Am.* **132**, 957–967.

Shen, Y., and Richards, V. M. (2013). "Temporal modulation transfer function for efficient assessment of auditory temporal resolution." *J. Acoust. Soc. Am.* **133**, 1031–1042.

Stone, M. A., Glasberg, B. R., and Moore, B. C. J. (1992). "Simplified measurement of impaired auditory filter shapes using the notched-noise method." *Br. J. Audiol.* **26**, 329–334.

Wegel, R. L., and Lane, C. E. (1924). "The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear." *Phys. Rev.* **23**, 266–285.

Wichmann, F. A., and Hill, N. J. (2001). "The psychometric function. I. Fitting, sampling, and goodness of fit." *Percept. Psychophys.* **63**, 1293–1313.

Wright, B. A. (1996). "Auditory filter asymmetry at 2000 Hz in 80 normal-hearing ears." *J. Acoust. Soc. Am.* **100**, 1717–1721.