Reduction of internal noise in auditory perceptual learning

Pete R. Jones,[a] David R. Moore, and Sygal Amitay
MRC Insitute of Hearing Research, University Park, Nottingham NG7 2RD, United Kingdom

Daniel E. Shub
School of Psychology, Nottingham University, University Park, Nottingham NG7 2RD, United Kingdom

This paper examines what mechanisms underlie auditory perceptual learning. Fifteen normal hearing adults performed two-alternative, forced choice, pure tone frequency discrimination for four sessions. External variability was introduced by adding a zero-mean Gaussian random variable to the frequency of each tone. Measures of internal noise, encoding efficiency, bias, and inattentiveness were derived using four methods (model fit, classification boundary, psychometric function, and double-pass consistency). The four methods gave convergent estimates of internal noise, which was found to decrease from 4.52 to 2.93 Hz with practice. No group-mean changes in encoding efficiency, bias, or inattentiveness were observed. It is concluded that learned improvements in frequency discrimination primarily reflect a reduction in internal noise. Data from highly experienced listeners and neural networks performing the same task are also reported. These results also indicated that auditory learning represents internal noise reduction, potentially through the re-weighting of frequency-specific channels. © 2013 Acoustical Society of America.
[http://dx.doi.org/10.1121/1.4773864]

I. INTRODUCTION

Perceptual learning is improved performance on a sensory judgment task as a result of practice. While the phenomenon is well established, little is known about the mechanisms underlying such improvements. In the visual literature it has been variously suggested that reductions in internal noise (Dosher and Lu, 1998), or improvements in encoding efficiency (Gold *et al.*, 1999) may underlie learning. In this paper we examine whether either of these factors change during auditory (frequency discrimination) learning. We also examine two further potential limiting factors that have not previously been considered: response bias and attentiveness.

Internal noise is uncertainty in the internal response to a sensory input which, in contrast with external noise, is generated by sources intrinsic to the observer. Internal noise is therefore synonymous with intrinsic variability, and the two terms are often used interchangeably. Internal noise is fundamental to signal detection theory (SDT) (Green and Swets, 1974; Macmillan and Creelman, 2005). It is also a prominent concept in psychophysics (Gescheider, 1997; Klein, 2001), where the ogival psychometric function is theoretically justified as the cumulative form of a random variable with a bell-shaped distribution. Potential sources of internal noise include non-deterministic transduction (e.g., due to Brownian motion of hair cells) (Denk *et al.*, 1989), stochastic neural encoding and transmission both in the auditory periphery (Javel and Viemeister, 2000) and more centrally (e.g., Vogels *et al.*, 1989), and physiological maskers such as heartbeats and blood flow (Soderquist and Lindsey, 1971).

Over the last 50 years a number of measures of internal noise have been developed. These include external noise titration (Lu and Dosher, 2008), model-fitting (Jesteadt *et al.*, 2003), n-pass consistency (Green, 1964), multiple-looks (Swets, 1959), and direct variability estimates derived from distributions of errors (e.g., Buss *et al.*, 2009). Following related work in the visual literature (e.g., Gold *et al.*, 1999), we here utilized the model-fitting and double-pass consistency techniques. In addition, we also considered two direct variability estimates which were derived using the same data.

In contrast with internal noise, encoding efficiency constitutes a systematic rather than random limitation on performance (cf. Berg, 2004; Berg and Green, 1990). In sensory tasks, encoding efficiency primarily describes how well the listener is able to selectively integrate information across channels. How these channels are conceived depends on the task. For example, in spectral profile analysis, listeners must detect when the levels of one or more components of a multitone stimulus are changed. In such a task, if the frequency components are widely spaced then every frequency component in the complex can be considered a channel, and a good strategy would be to attend predominantly to those components where the level difference is greatest relative to the internal noise. In the present study, each interval in a two-interval, forced-choice paradigm is considered to be a channel, with similar quantities of internal noise in both channels. In this case a good strategy would be to attend equally to both intervals. Encoding efficiency can either be inferred by comparing observed sensitivity to the ideal (e.g., Berg and Green, 1990; Tanner and Birdsall, 1958), or by comparing a listener's estimated strategy to the ideal (e.g., Dai and Berg, 1992; Alexander and Lutfi, 2004). Here we used variations on both these approaches. Signal detection theory was used to derive a model containing an encoding efficiency

---

[a]Author to whom correspondence should be addressed. Electronic mail: p.r.jones@ucl.ac.uk

parameter which was fitted to observed performance, while a novel classification boundary approach was used to estimate listeners' encoding strategies.

Response bias (hereafter, bias) is the tendency to favor one response over another, irrespective of the stimulus features. Thus, a listener who is biased towards one alternative may select it even when the sensory evidence makes it more likely that the other is true. Psychometric thresholds are liable to be negatively affected by bias, unless either explicit corrections are made, or metrics such as d' used that are designed to partial out these effects. Indices of response bias can be derived from lateral shifts in psychometric functions (Gescheider, 1997), or by using SDT to calculate the distance of the listener's criterion from the ideal (Macmillan and Creelman, 2005).

Inattentiveness is the complement of sustained attention. It expresses the fact that in a proportion of trials listeners appear to respond independently of the sensory information, possibly reflecting a lapse in concentration. For simplicity, it is common to assume that inattention is a binary process that occurs independently of the stimulus level or trial number (cf. Viemeister and Schlauch, 1992). Historically, inattentiveness has been little studied relative to the other limitations described here. This may in part be because inattentiveness is specifically selected against in many psychophysical experiments (which tend to be populated by highly experienced, reliable and well-motivated observers). Nonetheless, a number of behaviors have been identified from which metrics of inattention may be derived, such as the amount and/or profile of excursions from threshold in an adaptive track (Moore et al., 2008), or asymptotic performance on the psychometric function (Green, 1995).

In this study, we investigated the extent to which each of these mechanisms (internal noise, encoding efficiency, response bias, inattentiveness) contributes to auditory perceptual learning. The task was two-interval, two-alternative, forced-choice (2I2AFC) frequency discrimination in which the frequency of both tones was jittered by adding Gaussian noise. Frequency discrimination was selected due to both its prevalence in the learning literature (e.g., Hawkey et al., 2004; Demany, 1985) and its robust tendency to improve with practice relative to other psychoacoustic tasks (cf. Wright and Zhang, 2009). Jitter was used to introduce an external noise component as a reference for internal noise

magnitude. On simple auditory tasks requiring judgments based on pure tone stimuli, the limiting factor in performance is often suggested to be internal noise (e.g., Houtsma, 1995; Durlach and Braida, 1969). If this is the case during learning, then the magnitude of internal noise should decrease as a function of practice, concomitant with improved discriminability. Conversely though, there has been a tendency in the visual literature to conclude that changes in encoding efficiency underlie learning (e.g., Gold et al., 1999, 2004; Chung et al., 2005, though see Lu and Dosher, 2009). If auditory perceptual learning is analogous to visual perceptual learning then we might expect predominant changes in encoding efficiency. There has been comparatively little research into response bias and inattentiveness during learning. We therefore made no predictions as to their prevalence or whether they would change with practice.

II. GENERAL METHODS

A. Stimuli and apparatus

The stimuli in all conditions were 300 ms sinusoids, gated on/off by 10 ms $\cos^2$ ramps and presented at 70 dB SPL. Stimuli were digitally synthesized in MATLAB v7.4 (2007a, The MathWorks, Natick, MA) using a sampling rate of 44.1 kHz and 24-bit quantization. Digital-to-analog conversion was carried out by a PCI sound card (Darla Echo; Echo Digital Audio Corporation, Carpinteria, CA), interfaced via the Psychophysics Toolbox v3 (Brainard, 1997) ASIO wrapper (Steinberg Media Technologies, Hamburg). Stimuli were presented diotically via Sennheiser HD 25-I headphones. Participants were tested individually in a double-walled sound-attenuating booth. They responded by pressing one of two buttons on a button box. Visual fixation cues and feedback were presented on an LCD monitor.

B. Procedure

The task was 2I2AFC frequency discrimination, for which participants were asked to "pick the higher-pitched tone." Each trial commenced with a 400 ms warning interval during which a visual fixation cross was displayed, followed by two 300 ms tones separated by a 400 ms interstimulus interval. On each trial a pair of tones was sampled in random order from a pair of Gaussian distributions[1] with a common
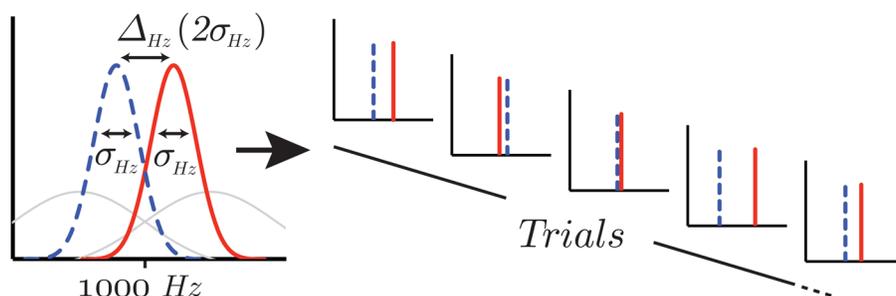
FIG. 1. (Color online) Stimulus schema for a single external noise condition. The dashed and solid distributions are the jittered "low" and "high" tone distributions, respectively. On each trial a tone was independently drawn from each distribution in random order. Randomly drawn frequency values for the first five trials are shown on the right. The difference in Hz between the means of the two normal distributions, Δ_{Hz}, was determined by the condition. The common standard deviation of the two distributions, σ_{Hz}, was set so as: $\sigma_{Hz} = \Delta_{Hz}/2$. Conditions were fixed within a block, and in every condition the two distributions were symmetric about 1 kHz. An example pair of distributions corresponding to a greater Δ_{Hz} condition is shown in light gray hairlines.

standard deviation of σ_{Hz} and means of $1000 \pm \sigma_{Hz}$ (Fig. 1). Participants were given an unlimited time to respond, after which visual feedback was presented for 400 ms prior to the next trial onset.

The standard deviation of the jitter, σ_{Hz}, took on the values 0.5, 1.5, 2.5, 3.5, 4.5, and 5.5 Hz. This range of values was chosen to accommodate the most likely magnitude of internal noise based on pilot data. In keeping with Jesteadt *et al.* (2003), the separation between distributions, Δ_{Hz}, was co-varied along with the amount of jitter, σ_{Hz}, such that $\Delta_{Hz} = 2\sigma_{Hz}$. The overlap between distributions was therefore constant across all six conditions and resulted in an invariant d'_{ideal} of 2.0 (i.e., the ideal listener would be expected to score ~92% correct in all conditions).

Feedback was determined by the response of the subject relative to the actual frequencies presented, and consisted of a "happy" or a "sad" smiley face. It was designed to reinforce the optimal response behavior of responding to the higher frequency tone, and to discourage the use of non-stimulus driven strategies. Additional feedback was presented at the end of each block in the form of a percentage score, again based on the frequencies of sounds presented (tones + noise) rather than on their values prior to jittering.

Each test block consisted of fifty trials drawn from one of the six frequency differences, Δ_{Hz}. Each session consisted of thirty-two test blocks. The number of trials per session (1600) was large given typical frequency-discrimination learning rates (e.g., Molloy *et al.*, 2012), but is consistent with the slower learning observed when the training stimuli are randomly varied (Amitay *et al.*, 2005).

The test blocks in the first session were preceded by two short practice blocks consisting of 10 "easy" (150 Hz difference) and 10 "difficult" (8 Hz difference) trials, intended to familiarize participants with the procedure. In blocks 1 to 24, each frequency difference was tested four times in pseudorandom order. These 1200 trials were used in the model fit analysis (see below). In the final eight blocks, all the previous blocks from the narrowest ($\sigma_{Hz} = 0.5; \Delta_{Hz} = 1$) and broadest ($\sigma_{Hz} = 5.5; \Delta_{Hz} = 11$) frequency differences were repeated in pseudorandom order. These last 400 trials were used in the double-pass consistency analysis. They were identical to the trials heard earlier in the experiment, although the order of the trials within each block was randomized in order to avoid the potential confound of response dependencies on consistency (for discussion see Levi *et al.*, 2005; Spiegel and Green, 1981). None of the listeners reported, when questioned, being aware of the fact that the last eight blocks consisted of repetitions of earlier trials. All 1600 trials were used to carry out the psychometric function and classification boundary analyses. Sessions lasted approximately 80–90 min in total, including two rest breaks. All listeners took part in one session per day for four consecutive days.

C. Analyzing learning

Learning was assessed by examining sensitivity as a function of session. For each stimulus condition, successive pairs of test blocks were concatenated to yield blocks of 100 trials. Each analysis block was then used independently to derive estimates of sensitivity, d', and response criterion, λ, as per Wickens (2002). In two blocks participants responded 100% correctly to one interval. In these two cases, the number of correct responses was adjusted by 0.5 to yield a defined d' value (Macmillan and Creelman, 2005).

D. Modeling behavior

Measures of internal noise, encoding efficiency, bias and inattentiveness were derived using four methods of analysis: model fit, classification boundary, psychometric function, double-pass consistency. Although all related, each method differs in terms of its precise derivations, assumptions, and how it partitions performance into various limiting parameters. The use of multiple methods allowed for constructs common across methods (e.g., internal noise) to be cross-validated, and for a greater range of constructs to be examined. Example individual data for a single listener derived using each method are shown in Fig. 2 (n.b. there is no graphical analog to the double-pass method). Each panel is discussed in the context of its associated methodology.

1. Model fit

Encoding efficiency, η (cf. Berg, 2004), and the standard deviation of a zero-mean Gaussian internal noise, σ_{Int}, were calculated by fitting observed sensitivities to the model:

$$d' = \frac{\eta \cdot \Delta_{Hz}}{\sqrt{\sigma_{Int}^2 + \sigma_{Hz}^2}}, \qquad (1)$$

where Δ_{Hz} and σ_{Hz} represent the mean separation and the common deviations of the stimulus distributions, respectively. This model represents a version of that described previously by Jesteadt *et al.* (2003), extended to include an encoding efficiency parameter that reflects any deterministic limitations on performance arising from the listener's encoding strategy. The derivation of Eq. (1) is given in Appendix A.

As shown in Fig. 2 (top-left), least-squares fits to Eq. (1) were made to observed sensitivities. These fits were constrained by transformation to yield finite and positive parameter values. Fits were made independently to each set of 600 trials (two blocks from each condition), yielding two estimates of internal noise and encoding efficiency per listener, per session. These estimates were averaged to provide a single value for comparison with the other three measures.

2. Classification boundary

The listener's task in 2I2AFC frequency discrimination can be conceptualized as a binary classification problem. As shown in Fig. 2 (top-right), the decision space is two-dimensional, with each axis corresponding to the frequency in a given interval. The target variable is the interval containing the higher tone (either "interval 1" or "interval 2"). When interval 1 is plotted on the abscissa, the data points
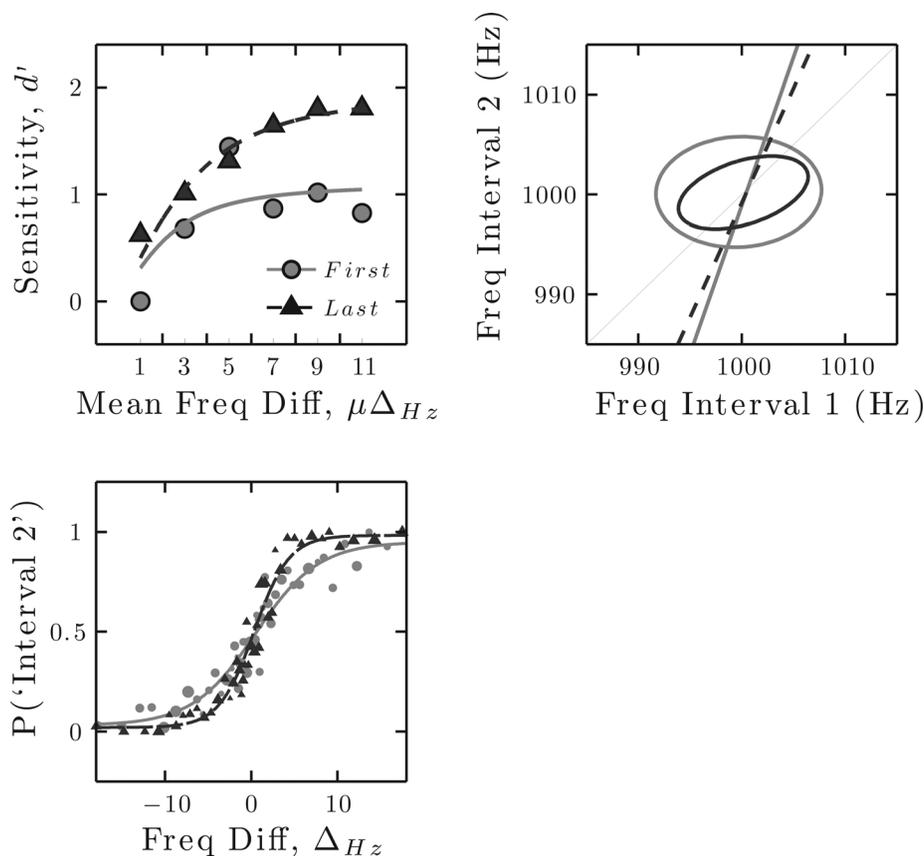
FIG. 2. Individual model fits for a single listener; first and last session only. (Top-left) Model-fits to observed sensitivities. Curves represent least-square fits to Eq. (1), from which internal noise and encoding efficiency parameters are derived. (Top-right) Estimated classification boundaries (solid lines) and standard deviations of errors with respect to their boundaries (ellipses). Smaller ellipses indicate less internal noise, while a classification boundary closer to the identity function indicates a more efficient encoding strategy. (Bottom-left) Cumulative Gaussian psychometric fits to Eq. (2). The proportion of interval 2 responses are given as a function of frequency difference (freq$_2$ − freq$_1$), post-jittering. A steeper slope indicates less internal noise, while asymptotic performance closer to the upper/lower bounds (0 and 1) indicates more attentiveness. (See body text for further details.)

belonging to class "interval 1" will be below the identity function, while class "interval 2" points will be above the identity function. Since the stimulus distributions are arranged symmetrically around 1 kHz, the ideal classification boundary will have a slope of one and pass through the origin. Alternatively, less optimal strategies may be employed. For example, the listener shown in Fig. 2 gives disproportionate weight to interval 1 in both session 1 and (to a lesser extent) in session 4.

Each listener's classification boundary was estimated by finding the linear function that best predicts their responses given the presented frequencies (i.e., after the addition of external noise). The angle from the observed slope to the ideal was taken as an index of encoding efficiency, η. The spread of misclassifications given this boundary was interpreted as an index of internal noise magnitude, σ_{Int}. Spread was computed as the standard deviations of 2-D Gaussians fitted to errors (shown by the ellipses in Fig. 2). The Euclidean distance of the classification boundary from the point of physical equality {1000,1000} was interpreted as interval response bias, CE.

Linear discriminant analysis was used to fit classification boundaries to the data from each session[2] (1600 trials per fit). This yielded one estimate of internal noise, encoding efficiency and bias per listener, per session.

3. Psychometric function

Psychometric functions were estimated by maximum likelihood fits to the function

$$P(\text{Int}\,2) = \gamma_{\text{lo}} + (\gamma_{\text{up}} - \gamma_{\text{lo}})\Phi(x;\mu,\sigma), \qquad (2)$$

where $P(\text{int}\,2)$ is the proportion of interval 2 responses, γ_{lo} and γ_{up} are lower and upper asymptotes, and $\Phi(x;\mu,\sigma)$ is the Gaussian cumulative distribution function with mean μ and standard deviation σ, evaluated at the values x. In our task, x is the linear difference in frequency between the two intervals, with a positive value representing a higher frequency in the second interval. When fitting psychometric functions, some authors additionally include a variable exponent term, which introduces a potential non-linearity to the slope of the sigmoid (e.g., Dai and Micheyl, 2011; Dai and Richards, 2011). Such a term did not substantively effect the present findings, and so was omitted (see Appendix B).

The fitted value of σ was taken as a measure of internal noise. The psychometric function was also used to derive two additional measures: response bias and inattentiveness. Response bias was indexed by constant error (CE): the estimated point of subjective equality, $\hat{\mu}$, minus the point of physical equality on the psychometric function. Inattention was modeled as a stationary, stochastic process by which listeners, on some proportion of trials K, respond independently of the sensory evidence. Following (Green, 1995, see also Wightman and Allen, 1992), K was derived from the estimated asymptote values, thus

$$K = 1 - \gamma_{\text{up}} + \gamma_{\text{lo}}. \qquad (3)$$

The main caveat with this approach as a measure of internal noise is that the psychometric function confounds random and deterministic limitations on performance, the latter of which are inconsistent with the notion of noise as random variability (Green, 1964). In the limit, a listener who attends only to uninformative channels will have a slope of zero.

Changes in the gradient of the psychophysical slope are therefore ambiguous; they may reflect either more variability in the decision variable, or a less efficient strategy, or a mixture of both. This ambiguity can be resolved either by assuming (often implicitly) that the encoding strategy is ideal (e.g., Glasberg *et al.*, 2001; Tanner, 1958), or by estimating the listener's encoding strategy and making fits to the actual, trial-by-trial decision variable, thereby partialing out any systematic performance limitations (e.g., Berg, 2004). In the present work we assumed that the encoding strategy is ideal. However, in doing so we acknowledge that the resultant value will be an upperbound on internal noise magnitude. The extent that this value approximates the true value will depend on the efficiency of the encoding strategy. This will be indicated both by the model-fit analysis and the classification boundary analysis.

Psychometric functions were fitted using the 'psignifit' Matlab toolbox (v2.5.6), which implements the maximum-likelihood method described by Wichmann and Hill (2001). As shown in Fig. 2 (bottom-left), fits were made independently for each session, using all 1600 trials. This yielded one estimate of internal noise, inattentiveness and bias per listener, per session.

4. Double pass consistency

The central tenet of the *n*-pass consistency technique (Green, 1964; Spiegel and Green, 1981) is that when the same stimulus is presented multiple times, the probability of agreement between each of the listener's responses is determined by the ratio of internal-to-external noise. The mathematics of this is expounded by (Lu and Dosher, 2008, see also Burgess and Colborne, 1988), who show that, assuming a normally distributed internal noise drawn independently on each observation, the probability of two answers agreeing, P_A, is determined solely by the ratio of internal-to-external noise, α, together with the stimulus-determined parameters $(\Delta_{Hz}, \sigma_{Hz})$:

$$P_A = \int \phi(x - \Delta_{Hz}; 0, \sqrt{2}\sigma_{Hz})\{\Phi^2(x; 0, \sqrt{2}\alpha\sigma_{Hz})$$
$$\times [1 - \Phi(x; 0, \sqrt{2}\alpha\sigma_{Hz})]^2\}dx, \qquad (4)$$

where $\phi(x; 0, \sigma)$ is a Gaussian random variable with mean 0 and standard deviation σ, and $\Phi(x; 0, \sigma)$ is its cumulative distribution function. In short, this equation states that the probability of agreement can be computed from the probability of the same response occurring twice for a given signal, weighted by the probability of that signal occurring. In turn, the probability of the same internal response occurring twice is the probability of a greater interval 1 internal response occurring on the first pass, multiplied by the probability of a greater interval 1 internal response occurring on the second pass (which, assuming independent, identically distributed noise, is the square of either probability considered singularly), additively combined with the analogous product of the corresponding interval 2 probabilities.

Consistency was examined independently for each session, and separately for the low and high external noise con-

ditions. Specifically, a subset of the trials were presented in a two-pass manner to allow for double pass consistency (DPC) to be estimated. Response consistency was calculated as the proportion of trials where the listener responded the same way across both presentations, irrespective of whether the response was correct. The consistency score was then used to derive estimates of internal noise by numerically solving Eq. (4). This yielded two estimates of internal noise and encoding efficiency per listener, per session (i.e., one each for the lowest and highest external noise conditions). However, performance was so low in the hardest condition ($\sigma_{Hz} = 0.5$; $\Delta_{Hz} = 1$) that it appeared that some listeners were not able to maintain a stable criterion. Thus, only the internal noise estimates from the low external noise condition ($\sigma_{Hz} = 5.5$; $\Delta_{Hz} = 11$) are reported here.

III. EXPERIMENT I: LEARNING IN NAIVE LISTENERS

A. Listeners

Sixteen listeners participated, none of whom had any prior experience of auditory psychophysics. Eleven were female (mean age 22.3), five were male (mean age 25.3). All had normal hearing, as assessed by audiometric screening administered in accordance with the BSA standard procedure (≤ 20 dB HL or less bilaterally at 0.5–4 kHz octaves; British Society of Audiology, 2004). Listeners were not screened based on initial task performance, were recruited through advertisements placed around Nottingham University campus, and received an inconvenience allowance for their time. The study was conducted in accordance with Nottingham University Hospitals Research Ethics Committee approval and informed written consent was obtained from all participants.

One listener was excluded from all analyses due to performing at chance in all conditions throughout all four sessions. Two additional listeners were not included in the double-pass analysis due to a technical error.

B. Results

1. Learning

Group mean sensitivity (d') for listeners across sessions is shown for each stimulus condition in Fig. 3. Sensitivity increased as a function of session $[F(3, 42) = 16.7, p < 0.001, \eta_p^2 = 0.54]$, indicating improvement with practice. There was no significant interaction between session and condition $[F(15, 210) = 1.3, p = 0.21]$, indicating that learning occurred irrespective of external noise condition. Response criterion (λ) did not change across sessions $[F(3, 42) = 1.3, p = 0.30]$. There was substantial variability in performance between listeners, with d' ranging by approximately one unit within each session. There was also a large degree of variability in learning, with changes in mean sensitivity, $\Delta d'$, varying from -0.04 to 0.92 across listeners.

2. Model fit

Least-square fits were made to the model given in Eq. (1). Figure 4 shows the group mean values of internal noise
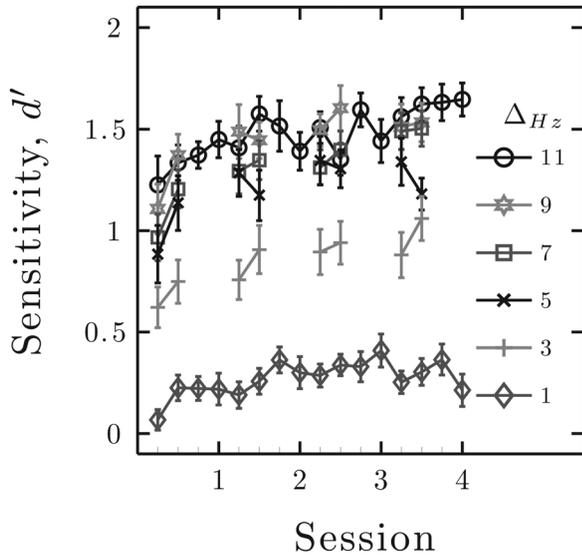
FIG. 3. Frequency discrimination learning. Each point shows group-mean sensitivity, d', as a function of session, averaged over all 15 listeners. Error bars represent $\pm$ 1 standard error of the mean, both here and in all subsequent figures. Each stimulus condition is shown separately. The breaks between data points in conditions $\Delta_{Hz} = \{3 - 9\}$ are due to the fact that blocks from these conditions were not repeated at the end of each session (i.e., when assessing consistency).

(σ_{Int}) and encoding efficiency (η). Internal noise estimates decreased significantly across sessions [$F(3, 42) = 4.7$, $p = 0.007, \eta_p^2 = 0.25$]. There was a non-significant trend towards an improvement in encoding efficiency, with improvements observed in 11 of 15 listeners. [$F(3, 42) = 2.4$,
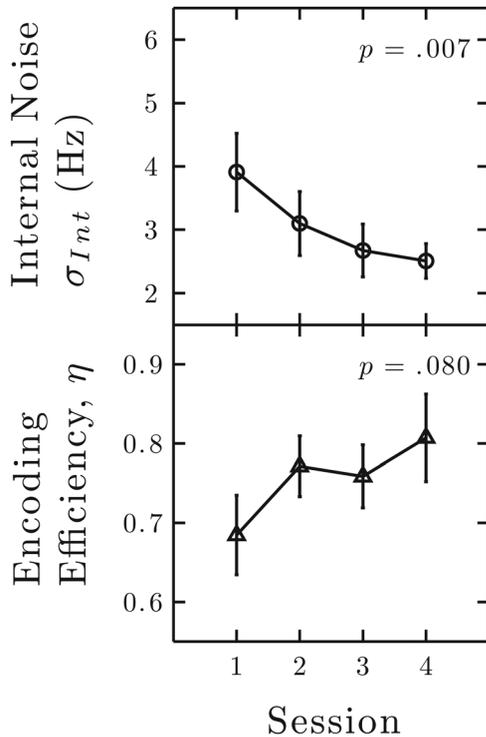

FIG. 4. Changes in model fit parameter estimates with practice. (Top) Group mean internal noise, σ_{Int}, as a function of session. (Bottom) Group mean encoding efficiency, η, as a function of session. In each panel, the main effect p value from the associated repeated measures analysis of variances are shown top-right; see body text for details.
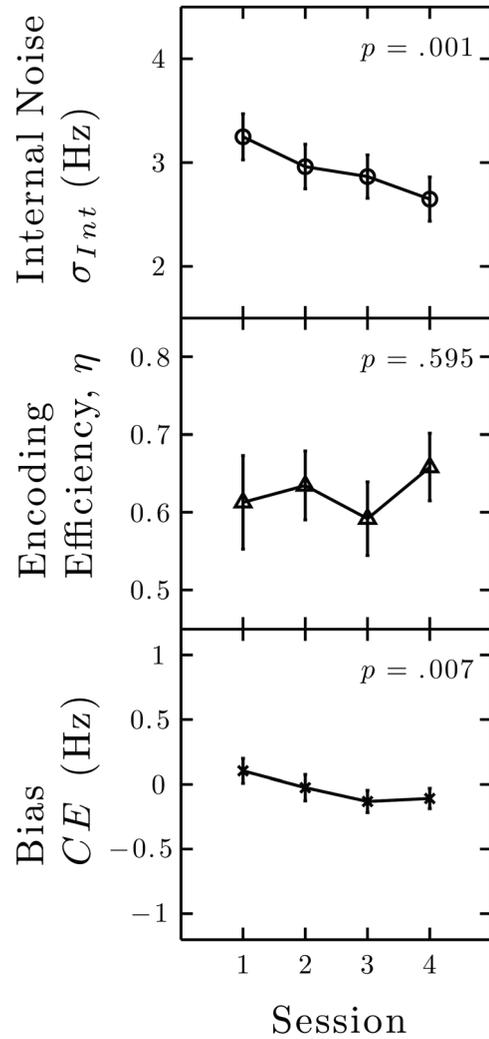

FIG. 5. Changes in classification-boundary parameter estimates with practice. Panels show the following group mean values as a function of session: (top) standard deviation of errors (given an estimated classification boundary) as a measure of internal noise; (middle) distance of the boundary slope from the ideal, as a measure of encoding efficiency; and (bottom) CE as a measure of bias (a negative CE value indicates an interval 1 response preference). This figure follows the same format as Fig. 4, with which the internal noise estimates are directly comparable.

$p = 0.08$]. Goodness-of-fit improved throughout the study, with median $r^2 = 0.53$ in session 1 increasing to $r^2 = 0.63$ in session 4.

3. Classification boundary

Group mean values of internal noise (σ_{Int}), encoding efficiency (η) and bias (CE), as derived using the classification boundary technique, are given as a function of session in Fig. 5. Internal noise estimates decreased significantly across sessions [$F(3, 42) = 6.9, p < 0.001, \eta_p^2 = 0.33$]. No change in encoding efficiency was observed [$F(3,42) = 0.6$, $p = 0.60$]. Bias did significantly change over sessions [$F(3, 42) = 4.6, p = 0.007, \eta_p^2 = 0.25$], with listeners tending to favor interval 2 in session 1 (CE = 0.10), and interval 1 in session 4 (CE = -0.11), though none of the session means significantly differed from 0 (no bias) [Hotelling's T^2; $T^2(4, 11) = 13.2, p = 0.10$].

4. Psychometric function

Psychometric function fits were made to Eq. (2). [Mean goodness-of-fit: $r^2 = 0.87$.] The slope of the function (internal noise) became steeper in 87% of listeners. There was little change in lower or upper asymptote (inattention) or in constant error (bias). Group mean values of internal noise (σ_{Int}), inattention (K) and bias (CE) are given as a function of session in Fig. 6. Internal noise estimates decreased significantly across sessions [$F(3,42) = 8.2$, $p < 0.001, \eta_p^2 = 0.37$]. No changes in inattention [$F(3,42) = 0.60$, $p = 0.62$] or bias [$F(3,42) = 0.68$, $p = 0.57$] were observed, with mean bias remaining indistinguishable from 0 throughout [$T^2(4,11) = 2.9, p = 0.69$].

5. Double pass consistency

Group mean values of internal noise (σ_{Int}) as derived using the DPC technique are given as a function of session in Fig. 7. Internal noise estimates decreased significantly across sessions [$F(3,36) = 9.9, p < 0.001, \eta_p^2 = 0.45$].

FIG. 7. Changes in double-pass internal noise estimates with practice. Each point shows group mean internal noise, σ_{Int}, as a function of session, estimated using the double-pass consistency method.

6. Comparison of metrics

As shown in Table I, correlations between the four sets of internal noise estimates were strong [$r \geq 0.69$; all $p < 0.001$]. Positive correlations were also observed between the bias estimates from the classification boundary and psychometric fit approaches [$r = 0.63; p < 0.001$], and between the encoding efficiency estimates from the model fit and classification boundary measures [$r = 0.37; p = 0.004$]. Individual internal noise estimates for the first and last sessions are given for each test in Table II. The double-pass consistency method tended to produce the somewhat larger estimates, being the greatest of the four in 88% of cases. Conversely, the model fit and classification boundary methods tended to produce the smallest noise estimates.

C. Discussion

Frequency discrimination sensitivity improved significantly with practice, although there was substantial individual variability in both performance and learning. Improvements in sensitivity were accompanied by a significant decrease in internal noise with little change in encoding efficiency, bias, and inattentiveness. The results show that practice-induced improvements in frequency discrimination sensitivity primarily represent a reduction in internal noise. Averaged over the four methods, mean internal noise values ranged from 3.2 to 6.0 Hz in session 1, and 2.5 to 2.9 Hz in session 4.
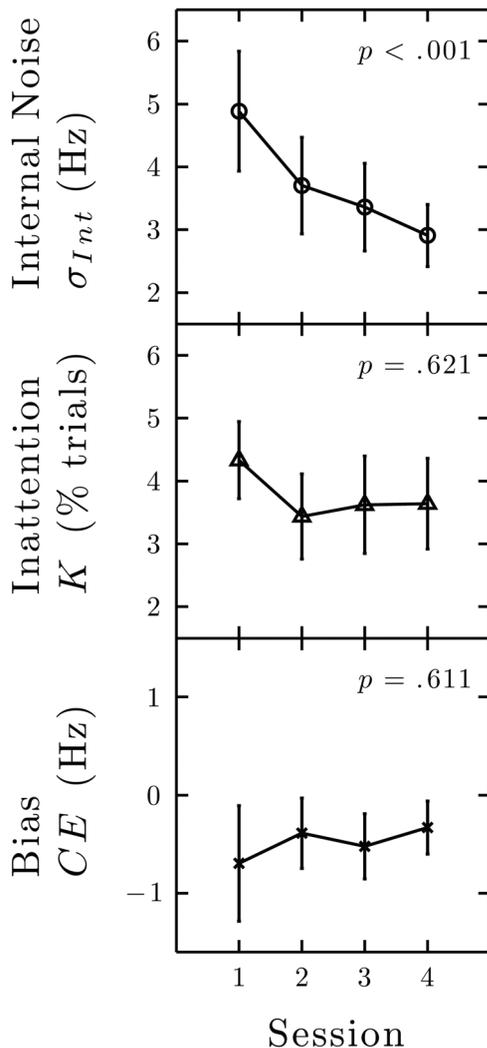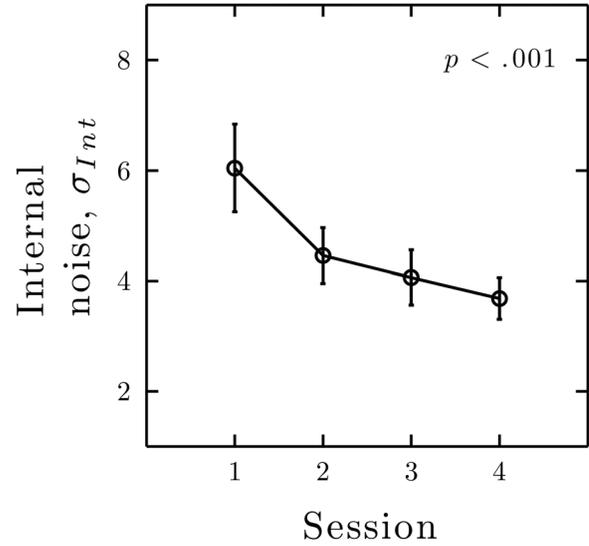

FIG. 6. Changes in psychometric function parameter estimates with practice. Panels show the following group mean values as a function of session: (top) fitted Gaussian standard deviation as a measure of internal noise, σ_{Int}; (middle) inattentiveness (derived from asymptotic performance), K, as a measure of sustained attention; (bottom) CE as a measure of bias.

TABLE I. Correlation coefficients, r, between internal noise estimates, σ_{Int}, from the model fit (MF), classification boundary (CB), psychometric function (PF), and double-pass consistency (DPC) methods.

	MF	CB	PF
DPC	0.68	0.81	0.82
PF	0.80	0.82	–
CB	0.62	–	–

TABLE II. Summary of internal noise results, σ_{Int}, for individual listeners during the first and last session. Initialisms follow the same format as Table I.

Listener	Session 1				Session 4			
	MF	CB	PF	DPC	MF	CB	PF	DPC
L1	3.0	3.1	3.2	5.3	1.3	2.0	2.0	2.9
L2	4.6	3.2	4.5	6.1	4.8	2.9	4.0	4.5
L3	2.3	2.5	3.3	5.2	2.5	2.2	2.4	3.3
L4	1.8	2.2	2.2	3.2	2.0	2.1	1.9	3.2
L5	1.4	3.4	3.7	6.1	2.1	2.3	2.0	3.2
L6	5.0	4.1	6.1	11.6	4.5	3.7	3.9	7.4
L7	1.8	2.1	2.0	2.9	1.4	1.1	1.5	1.8
L8	2.7	2.3	2.0	–	2.7	3.5	2.6	–
L9	10.3	5.1	15.8	–	2.0	4.8	9.4	–
L10	5.1	2.9	4.1	5.0	3.3	2.0	2.7	3.0
L11	2.3	2.3	2.6	3.6	1.5	2.1	2.0	2.6
L12	5.3	4.8	10.2	11.9	1.6	3.0	1.9	4.5
L13	2.6	2.7	3.0	4.0	2.8	2.0	2.6	4.4
L14	6.8	3.5	5.5	6.9	3.1	2.8	2.6	3.2
L15	3.5	3.2	5.1	6.7	2.0	1.5	2.1	4.0

The four methods yielded highly correlated estimates of internal noise. Notably, since encoding efficiency was less than ideal, the internal noise estimates from psychometric functions tended to be consistently greater than with the model-fit and classification boundary methods. However, encoding efficiency remained largely invariant throughout. The changes in internal noise observed using psychometric functions therefore remained robust.

IV. EXPERIMENT II: EXPERIENCED LISTENERS

Group mean performance in our naive listeners (experiment I) failed to asymptote after four sessions. It may therefore be that sensitivity could be further improved with additional training. It may also be that any such additional learning is limited by factors other than internal noise. To assess these possibilities, two listeners with extensive prior task experience (one of whom was the first author) were tested using the same stimuli.

Furthermore, a potential concern with the methodology of experiment I is that the external noise (introduced via jittering) may not have been independent of listeners' internal noise, and thus may have introduced additional variability into listeners' decisions, not normally present during frequency discrimination. The two experienced listeners were therefore also tested using unjittered stimuli. Psychometric functions fitted to "zero noise" data were compared to those derived under jittering. Greater internal noise would be indicated by systematically shallower slopes in the jittered condition.

A. Methods

The stimuli followed those described in Experiment I, except that all stimulus parameter values (σ_{Hz}, Δ_{Hz}) were halved. This adjustment was necessary since these listeners performed at ceiling when $\Delta_{\text{Hz}} > 5$ Hz. Both listeners performed three practice sessions, followed by nine test sessions.

over two weeks. Each session consisted of 12 blocks, equivalent to the first phase of the session in the main experiment. Listeners then performed 3 additional test sessions in which no external noise was added ($\sigma_{\text{Hz}} = 0$).

B. Results and discussion

1. Performance and model estimates

The results of two experienced listeners are summarized in Table III, along with the group-mean data from the final training session of experiment I for comparison. Given the amount of prior task experience no improvement in sensitivity was expected across test sessions, and none was observed [$F(7) \leq 2.3, p \geq 0.176$]. Because of the different stimulus conditions, d' values were not comparable between experiments. As such, performance was quantified as the mean of listeners' discrimination limens at the 75% and 25% correct levels, FDL_{Hz}.

Both listeners' frequency discrimination limens were significantly lower than in the post-training naive listeners [$t(14) \geq 4.5, p < 0.001$], indicating that further learning beyond that observed in experiment I is possible. As per experiment I, the model fit and psychometric fit techniques were used to estimate internal noise, encoding efficiency, inattention and bias. The pattern of results continued the learning trend observed in experiment I. Relative to the less experienced listeners of experiment I, internal noise magnitude was further decreased [$t(14) \geq 4.1, p \leq 0.001$], with no differences in encoding efficiency [$t(14) \leq 0.2, p \geq 0.828$] or bias [$t(14) \leq 0.5, p \geq 0.632$]. This finding corroborates our conclusion that changes in internal noise underlie frequency discrimination learning. Inattentiveness was also lower than the naive group-mean [$t(14) \geq 4.9, p < 0.001$], suggesting that very highly trained listeners may also benefit from improved sustained attention.

2. Internal noise with and without external noise

Figure 8 shows psychometric functions with and without external noise. Performance in the two cases was virtually indistinguishable. In one listener (PJ) estimated internal noise was marginally (0.1 Hz) smaller, while in KM estimated internal noise was marginally (0.2 Hz) greater. These results indicate that the use of jittering did not affect the internal noise estimates, either here or in experiment I. These results are consistent with Jesteadt

TABLE III. Summary of frequency difference limens (FDLs) in Hz, and fitted behavioral parameters for group-mean naive listeners (final session) and the experienced listeners KM and PJ. Fitted parameters were internal noise (σ_{Int}) and encoding efficiency (η), estimated using the model fit; and internal noise (σ_{Int}), inattentiveness (K), and bias (CE) estimated from psychometric functions.

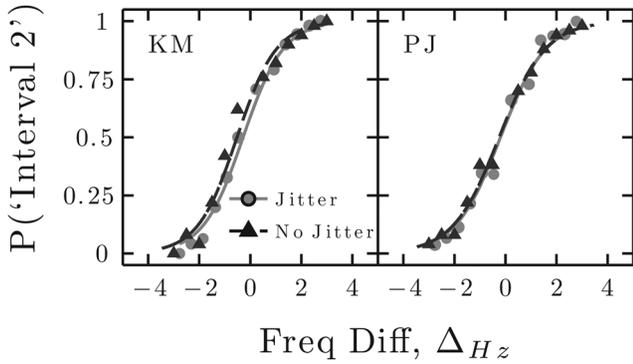| Listener	FDL_{Hz}	Model fit		Psychometric fit		
		σ_{Int}	η	σ_{Int}	K	CE
μNaïve	2.8	3.0	0.8	2.9	0.0	−0.3
KM	0.8	0.7	0.8	0.8	0.0	−0.0
PJ	1.3	1.1	0.8	1.2	0.0	0.0

FIG. 8. Psychometric functions for experiment II. Black triangles and dashed-lines indicate raw data and psychometric fits (respectively) given non-jittered stimuli. Gray circles and lines indicate analogous binned raw data and psychometric fits given jittered stimuli. In both cases fits were made to Eq. (2).

et al. (2003), who also observed good agreement between estimates of internal noise derived under jittering, and the slope of a psychometric function fitted to data without external variability.

V. EXPERIMENT III: SIMULATIONS

It has been suggested in the visual literature that perceptual learning represents "re-weighting of stable early sensory representations" (Lu and Dosher, 2009; Mollon and Danilova, 1996). Although we found no evidence of channel re-weighting at the behavioral level (where each stimulus presentation interval was modeled as a channel), our data are consistent with a process of iterative re-weighting of channels at a neural level of description. Such channel re-weighting is a plausible explanation for learning on a frequency discrimination task, since psychophysical thresholds are substantially poorer than would be predicted from the precision of information encoded at the periphery (e.g., Siebert, 1970; Heinz *et al.*, 2001). To investigate whether a process of early sensory re-weighting can produce the observed pattern of learning, a simple neural network model was trained and analyzed using the same methods as the human listeners.

A. Methods

The neural network consisted of a single-layer perceptron (Dayan and Abbott, 2001), with 60 input units innervating a single output unit. The input layer simulated a population of human auditory nerve fibres, with 60 gammatone filters ERB-spaced between 100 and 10 000 Hz (Glasberg and Moore, 1990). This array was constructed using the same model and parameters as described in Heinz *et al.* (2001). The mean firing rate of each node (i.e., rate-place encoding) was combined in a linear weighted sum by the output node. The decision rule was to select the interval that maximized the output, thus

$$ \text{out} = \begin{cases} \text{Int 1,} & \text{if } \left(\left[\sum_{i=1}^{n} \omega_i a_i \right] - \left[\sum_{i=1}^{n} \omega_i b_i \right] \right) > 0, \\ \text{Int 2,} & \text{otherwise,} \end{cases} \quad (5) $$

where out is the system output, a_i and b_i represent the ith input unit's response to the first and second stimulus, respectively, and where ω_i represents the strength of the connection between the ith input unit and the output unit (which may be negative). All learning occurred via changes in the connection strengths between the input nodes and output node. The simulations were presented with the same stimuli/protocol as the human listeners. Weight adjustments were made online (i.e., after every trial) via the delta rule (Dayan and Abbott, 2001). The range of learning and starting rates were selected based on a brief period of trial-and-error using a validation dataset, but the precise values were randomly generated at the point of testing.

B. Results and discussion

Fifteen independent simulations were run and were analyzed in the exact same manner as the human listeners. The key results are summarized in Fig. 9. The upper panel expresses how frequency discrimination sensitivity increased as a function of session [$p < 0.001$]. The lower panel shows the concomitant decrease in internal noise as estimated with the same four methods as described previously [all $p < 0.001$]. In short, through the selective re-weighting of simulated auditory nerve responses, the model exhibited a qualitatively similar pattern of learning to human listeners in terms of increased performance and reduced internal noise. This indicates that the observations of reduced internal noise in human listeners are consistent with the hypothesis of
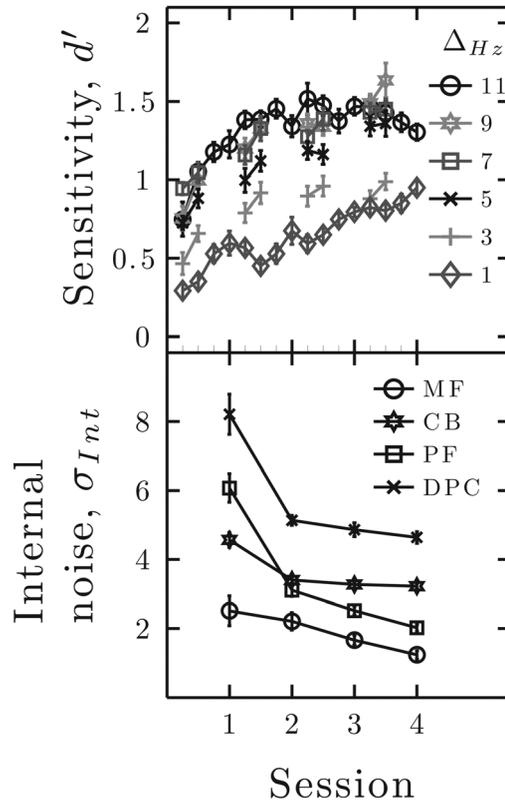

FIG. 9. Simulated frequency discrimination learning. The top panel shows changes in d' as a function of block/session for each stimulus condition, in the same format as the human listener data given in Fig. 3. The bottom panel shows internal noise estimates as a function of session using each of the following measures: model fit (MF), classification boundary (CB), psychometric function (PF), and double-pass consistency (DPC).

Lu and Dosher (2009) that perceptual learning reflects are re-weighting of early sensory representations.

VI. GENERAL DISCUSSION

The purpose of the experiments reported here was to determine the mechanisms underlying auditory perceptual learning. With each of four separate techniques, significant improvements on a frequency discrimination task was best modeled as a decrease in internal noise magnitude. No significant changes in encoding efficiency, bias or inattentiveness were observed. This pattern of results was continued in very highly trained listeners (though these listeners also exhibited less inattentiveness in addition to decreased internal variability and improved frequency discrimination).

The finding that internal noise underlies learning is consistent with recent work in auditory development, where differences in internal noise have also been effective in explaining age-related changes in pure tone discrimination performance. For example, a recent paper by (Buss et al., 2009, see also Buss et al., 2006) concluded, based on the slopes of psychometric fits, that children's poorer intensity discrimination limens were due to elevated levels of internal noise.

However, our finding conflicts with a prominent claim in the visual perceptual learning literature that "signal [enhancement] but not noise changes with perceptual learning" (Gold et al., 1999, see also Gold et al., 2004). In such papers, signal enhancement is conceived as occurring through the appropriate, relative weighting of spatially distributed channels (e.g., by concentrating on those parts of an image that contain the greatest signal-to-external-noise ratios). Such signal enhancement corresponds to our "encoding efficiency" concept. The claim of "signal not noise" is therefore diametrically opposed to our finding that internal noise underlies learning. This may indicate qualitative differences between auditory and visual learning. However, the claim by Gold et al. (1999) lacks coherence. In Gold et al. (1999) observers attempted to identify images corrupted by simultaneous Gaussian masker. Using a model equivalent to the SDT model presented in Eq. (1) an increase in signal enhancement was reported, with no change in internal noise magnitude. However, using a double-pass consistency analysis, a constant ratio of internal-to-external noise was reported. Given the nature of the noise, an optimization of spatial channel weights implies a reduction in effective external noise. Thus, a constant ratio of internal-to-external noise therefore implies a concomitant *reduction* in internal noise (see Lu and Dosher, 2009 for further discussion).[3]

A more cohesive account of visual perceptual learning is given by Lu and Dosher (e.g., Dosher and Lu, 1999), who argue that learning consists of both internal noise reduction and external noise exclusion. Given that our task precluded external noise exclusion (cf. Lu and Dosher, 2008, for discussion), our finding that internal noise reduction was the primary mechanism of learning is consistent with the theory of visual perceptual learning of Lu and Dosher. We predict that our finding would generalize to other pure tone auditory tasks (e.g., see Wright and Fitzgerald, 2005), which, together with frequency discrimination, constitute the substantial majority of the auditory perceptual learning literature. However, it remains an important and open question as to whether external noise reduction also occurs in auditory learning. For example, everyday listening situations often involve a substantial masking noise component. The filtering out of such noise may constitute a distinct and important perceptual learning process. Given the results from visual tasks, we predict that learning in such situations will be subserved by both additive internal noise reduction and an external noise exclusion mechanism.

VII. CONCLUSIONS

(1) Learning on a pure tone frequency discrimination task is subserved by a reduction in internal noise, potentially through re-weighting of early sensory information. Changes in encoding efficiency, bias or attentiveness do not contribute to learning.
(2) Estimates of internal noise derived from four methods (model fit, classification boundary, psychometric function, double pass consistency) yield values in close agreement.

APPENDIX A: MODEL DERIVATION

We assume that listeners perform the 2I2AFC task by linearly summing weighted activities across multiple channels. Here we shall treat each stimulus presentation interval as a channel. We further assume that (a) a given set of stimuli, $\langle S_1, S_2 \rangle$, generates fixed responses S_1 in channel 1, and S_2 in channel 2; (b) the external noise is a zero-mean Gaussian variable with standard deviation σ_{Hz} [$\phi(0, \sigma_{Ext}^2)$], which is independently and identically distributed across both channels; (c) the internal noise is zero-mean Gaussian variable with standard deviation, σ_{Int} [$\phi(0, \sigma_{Int}^2)$], which is independently and identically distributed across both channels; (d) the total activity in each channel is the difference between the signal stimuli and some fixed criterion value $|\lambda - S|$, additively combined with observations from each of the noise distributions; (e) the relative weight given to channel 1 and 2 are denoted by the scalars ω_1 and ω_2, respectively, the squared values of which sum to 1; (f) the observer chooses interval 1 if $([\lambda - S_2 + \phi(0, \sigma_{Int}^2) + \phi(0, \sigma_{Ext}^2)] \cdot \omega_1 + [\lambda - S_1 + \phi(0, \sigma_{Int}^2) + \phi(0, \sigma_{Ext}^2)] \cdot \omega_2) < 0$ (and interval 2 otherwise); (g) the ideal weights are given by the values $\langle \alpha_1, \alpha_2 \rangle$, which, when both intervals are equally informative will take the values $[-\sqrt{2}/2(+\sqrt{2}/2)]$. Given these assumptions, observed sensitivity, d', in the 2AFC case is

$$d'_{obs} = \frac{\sum |\omega \Delta_{Hz}|}{\sqrt{\sigma_{Int}^2 + \sigma_{Hz}^2}}, \tag{A1}$$

where ω is an array of relative channel weights, and Δ_{Hz} is an array of mean differences between criterion and signal values, $|\lambda - S|$. The performance of an observer limited only by their adopted relative weights is

$$d'_{\mathrm{weight}} = \frac{\sum |\omega \Delta_{\mathrm{Hz}}|}{\sigma_{\mathrm{Hz}}}. \qquad (A2)$$

While ideal performance is

$$d'_{\mathrm{ideal}} = \frac{\sum |\alpha \Delta_{\mathrm{Hz}}|}{\sigma_{\mathrm{Hz}}} = \frac{\Delta_{\mathrm{Hz}}}{\sigma_{\mathrm{Hz}}}, \qquad (A3)$$

where Δ_{Hz} is the difference in mean frequency of the two stimulus classes. Following the concept of efficiency of Berg (2004), we can partition overall observed efficiency, η_{total}, into the loss of efficiency due to non-optimal weights, η_{weight}, and due to internal noise, η_{noise}, thus:

$$\eta_{\mathrm{total}} = \frac{(d'_{\mathrm{obs}})^2}{(d'_{\mathrm{ideal}})^2} = \frac{(d'_{\mathrm{obs}})^2}{(d'_{\mathrm{weight}})^2} \cdot \frac{(d'_{\mathrm{weight}})^2}{(d'_{\mathrm{ideal}})^2}$$
$$= \eta_{\mathrm{noise}} \eta_{\mathrm{weight}}, \qquad (A4)$$

where

$$\eta_{\mathrm{weight}} = \left(\frac{d'_{\mathrm{weight}}}{d'_{\mathrm{ideal}}} \right)^2 = \left(\frac{\sum |\omega \Delta_{\mathrm{Hz}}|}{\sum |\alpha \Delta_{\mathrm{Hz}}|} \right)^2 \qquad (A5)$$

and

$$\eta_{\mathrm{noise}} = \left(\frac{d'_{\mathrm{obs}}}{d'_{\mathrm{weight}}} \right)^2 = \left(\frac{\sigma_{\mathrm{Hz}}}{\sqrt{\sigma_{\mathrm{Int}}^2 + \sigma_{\mathrm{Hz}}^2}} \right)^2. \qquad (A6)$$

Note that by definition $0 \leq \sqrt{\eta_{\mathrm{weight}}} \leq 1$. Applying this partitioning of efficiency (A4)–(A6) to the d' equations (A1)–(A3):

$$d'_{\mathrm{obs}} = d'_{\mathrm{ideal}} \sqrt{\eta_{\mathrm{total}}} \qquad (A7a)$$

$$= d'_{\mathrm{ideal}} \left| \frac{d'_{\mathrm{obs}}}{d'_{\mathrm{weight}}} \right| \sqrt{\eta_{\mathrm{weight}}} \qquad (A7b)$$

$$= d'_{\mathrm{ideal}} \left| \frac{\frac{\sum |\omega \Delta_{\mathrm{Hz}}|}{\sqrt{\sigma_{\mathrm{Int}}^2 + \sigma_{\mathrm{Hz}}^2}}}{\frac{\sum |\omega \Delta_{\mathrm{Hz}}|}{\sigma_{\mathrm{Hz}}}} \right| \sqrt{\eta_{\mathrm{weight}}} \qquad (A7c)$$

$$= d'_{\mathrm{ideal}} \frac{\sigma_{\mathrm{Hz}}}{\sqrt{\sigma_{\mathrm{Int}}^2 + \sigma_{\mathrm{Hz}}^2}} \sqrt{\eta_{\mathrm{weight}}} \qquad (A7d)$$

$$= \frac{\Delta_{\mathrm{Hz}}}{\sigma_{\mathrm{Hz}}} \frac{\sigma_{\mathrm{Hz}}}{\sqrt{\sigma_{\mathrm{Int}}^2 + \sigma_{\mathrm{Hz}}^2}} \sqrt{\eta_{\mathrm{weight}}} \qquad (A7e)$$

$$= \frac{\sqrt{\eta_{\mathrm{weight}}} \Delta_{\mathrm{Hz}}}{\sqrt{\sigma_{\mathrm{Int}}^2 + \sigma_{\mathrm{Hz}}^2}}. \qquad (A7f)$$

For simplicity, d'_{obs} and $\sqrt{\eta_{\mathrm{weight}}}$ are henceforth referred to as d' and η, thus:

$$d' = \frac{\eta \cdot \Delta_{\mathrm{Hz}}}{\sqrt{\sigma_{\mathrm{Int}}^2 + \sigma_{\mathrm{Hz}}^2}}. \qquad (A8)$$

APPENDIX B: NON-LINEAR SLOPES IN PSYCHOMETRIC FITS

Several studies concerning 2I2AFC pure tone discrimination tasks (e.g., Dai and Micheyl, 2011; Dai and Richards, 2011) have fitted psychometric functions in which sensitivity is related to signal strength, x, as follows: $d' = (|x|/\alpha)^\beta$. The β term in such models serves to vary the linearity of the psychometric slope (see Fig. 1 of Dai and Richards, 2011). Such non-linearity can be incorporated into the cumulative Gaussian fits described in Eq. (2), thus

$$P(\mathrm{Int}\,2) = \gamma_{\mathrm{lo}} + (\gamma_{\mathrm{up}} - \gamma_{\mathrm{lo}}) \Phi(\mathrm{sign}(x)|x|^\beta; \mu, \sigma^\beta). \qquad (B1)$$

The psychometric functions reported in the present study can thus be considered a special case of Eq. (B1), in which $\beta = 1$. By force-fitting linear ($\beta = 1$) slopes, an alternative explanation of learning may have been occluded. Moreover, since the value of β is liable to affect the other parameter estimates, the values of $\hat{\mu}$, $\hat{\sigma}$, $\widehat{\gamma_{\mathrm{lo}}}$, and $\widehat{\gamma_{\mathrm{up}}}$ may have been biased. To assess these possibilities, Eq. (B1) was fitted to each listener's session-by-session data, both when $\beta = 1$, and when β was a free parameter, constrained to be >0.

Consistent with Dai and Micheyl (2011), estimated values of β did not deviate from unity in any of the four sessions [Hotelling's $T^2; T^2(4, 11) = 12.2, p = 0.11$]. Accordingly, unconstraining β had a minimal effect on the estimates of the other four parameters. In each case, no significant differences were observed when β was allowed to vary [Hotelling's $T^2; T^2(4, 11) = 3.0 - 9.7, p = 0.18 - 0.67$], although, consistent with Dai and Micheyl (2011), there was a general trend towards lower lapse rates (e.g., grand-mean $\widehat{\gamma_{\mathrm{lo}}}$ decreased by 0.5%, while $\widehat{\gamma_{\mathrm{up}}}$ increased by 0.7%); this difference was not significant, however.

These results indicate that the assumption of linearity is acceptable for pure tone frequency discrimination, and that the use of a non-linear term, β, would not have substantively effected the findings reported in the present study.

[1]Jitter was normally distributed on a linear frequency scale. This was intended to introduce Gaussian variance on the underlying decision dimension. For frequency discrimination the decision dimension is likely to correspond most directly to logarithmic frequency (e.g., Wier et al., 1976). Given the very narrow range of frequencies employed in this experiment, we do not believe that this discrepancy has any significant effect on the results. For example, even in the greatest frequency difference condition, the Hellinger distance (Nikulin, 2001) between the linear and logarithmic distributions was slight [$H < 0.003$; where $0 \leq H \leq 1$].

[2]Classification boundary fits were also made using a support vector machine (Cortes and Vapnik, 1995), but this procedure yielded virtually identical results and as such is not reported.

[3]In contrast, see Appendix A for a description of how multiple information channels can "enhance the signal," independent of external noise level.

Alexander, J., and Lutfi, R. (2004). "Informational masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights," J. Acoust. Soc. Am. 116, 2234–2247.

Amitay, S., Hawkey, D., and Moore, D. (2005). "Auditory frequency discrimination learning is affected by stimulus variability," Atten. Percept. Psychophys. 67, 691–698.

Berg, B. (**2004**). "A molecular description of profile analysis: Decision weights and internal noise," J. Acoust. Soc. Am. **115**, 822–829.

Berg, B., and Green, D. (**1990**). "Spectral weights in profile listening," J. Acoust. Soc. Am. **88**, 758–766.

Brainard, D. (**1997**). "The psychophysics toolbox," Spat. Vis. **10**, 433–436.

British Society of Audiology (**2004**). "Recommended procedure for pure tone air and bone conduction threshold audiometry with and without masking and determination of uncomfortable loudness levels," technical report, British Society of Audiology.

Burgess, A., and Colborne, B. (**1988**). "Visual signal detection. IV. observer inconsistency," J. Opt. Soc. Am. A **5**, 617–627.

Buss, E., Hall, J., III, and Grose, J. (**2006**). "Development and the role of internal noise in detection and discrimination thresholds with narrow band stimuli," J. Acoust. Soc. Am. **120**, 2777–2788.

Buss, E., Hall, J., III, and Grose, J. (**2009**). "Psychometric functions for pure tone intensity discrimination: Slope differences in school-aged children and adults," J. Acoust. Soc. Am. **125**, 1050–1058.

Chung, S., Levi, D., and Tjan, B. (**2005**). "Learning letter identification in peripheral vision," Vis. Res. **45**, 1399–1412.

Cortes, C., and Vapnik, V. N. (**1995**). "Support vector machines," Mach. Learn. **20**, 273–297.

Dai, H., and Berg, B. (**1992**). "Spectral and temporal weights in spectral-shape discrimination," J. Acoust. Soc. Am. **92**, 1346–1355.

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

Dai, H., and Richards, V. (**2011**). "On the theoretical error bound for estimating psychometric functions," Atten. Percept. Psychophys. **73**, 919–926.

Dayan, P., and Abbott, L. (**2001**). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (MIT Press, Cambridge, Massachusetts), pp. 313–322.

Demany, L. (**1985**). "Perceptual learning in frequency discrimination," J. Acoust. Soc. Am. **78**, 1118–1120.

Denk, W., Webb, W., and Hudspeth, A. (**1989**). "Mechanical properties of sensory hair bundles are reflected in their Brownian motion measured with a laser differential interferometer," Proc. Natl. Acad. Sci. U.S.A. **86**, 5371–5375.

Dosher, B., and Lu, Z. (**1998**). "Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting," Proc. Natl. Acad. Sci. U.S.A. **95**, 13988–13993.

Dosher, B., and Lu, Z. (**1999**). "Mechanisms of perceptual learning," Vis. Res. **39**, 3197–3221.

Durlach, N., and Braida, L. (**1969**). "Intensity perception. I. preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372–383.

Gescheider, G. (**1997**). *Psychophysics: The Fundamentals* (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 73–124.

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

Glasberg, B., Moore, B., and Peters, R. (**2001**). "The influence of external and internal noise on the detection of increments and decrements in the level of sinusoids," Hear. Res. **155**, 41–53.

Gold, J., Bennett, P., and Sekuler, A. (**1999**). "Signal but not noise changes with perceptual learning," Nature **402**, 176–178.

Gold, J., Sekuler, A., and Bennett, P. (**2004**). "Characterizing perceptual learning with external noise," Cog. Sci. **28**, 167–207.

Green, D. (**1964**). "Consistency of auditory detection judgments," Psychol. Rev. **71**, 392–407.

Green, D. (**1995**). "Maximum-likelihood procedures and the inattentive observer," J. Acoust. Soc. Am. **97**, 3749–3760.

Green, D., and Swets, J. (**1974**). *Signal Detection Theory and Psychophysics* (Krieger, Melbourne, FL), pp. 1–455.

Hawkey, D., Amitay, S., and Moore, D. (**2004**). "Early and rapid perceptual learning," Nat. Neurosci. **7**, 1055–1056.

Heinz, M., Colburn, H., and Carney, L. (**2001**). "Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve," Neural Comput. **13**, 2273–2316.

Houtsma, A. J. M. (**1995**). "Pitch perception," in *Hearing*, edited by B. Moore (Academic, San Diego, CA), pp. 267–296.

Javel, E., and Viemeister, N. (**2000**). "Stochastic properties of cat auditory nerve responses to electric and acoustic stimuli and application to intensity discrimination," J. Acoust. Soc. Am. **107**, 908–921.

Jesteadt, W., Nizami, L., and Schairer, K. (**2003**). "A measure of internal noise based on sample discrimination," J. Acoust. Soc. Am. **114**, 2147–2157.

Klein, S. (**2001**). "Measuring, estimating, and understanding the psychometric function: A commentary," Atten. Percept. Psychophys. **63**, 1421–1455.

Levi, D., Klein, S., and Chen, I. (**2005**). "What is the signal in noise?," Vis. Res. **45**, 1835–1846.

Lu, Z., and Dosher, B. (**2008**). "Characterizing observers using external noise and observer models: Assessing internal representations with external noise," Psychol. Rev. **115**, 44–82.

Lu, Z., and Dosher, B. (**2009**). "Mechanisms of perceptual learning," Learn. Per. **1**, 19–36.

Macmillan, N., and Creelman, C. (**2005**). *Detection Theory: A User's Guide* (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 1–495.

Mollon, J., and Danilova, M. (**1996**). "Three remarks on perceptual learning," Spat. Vis. **10**, 51–58.

Molloy, K., Moore, D., Sohoglu, E., and Amitay, S. (**2012**). "Less is more: Latent learning is maximized by shorter training sessions in auditory perceptual learning," PLoS One **7**, e36929.

Moore, D., Ferguson, M., Halliday, L., and Riley, A. (**2008**). "Frequency discrimination in children: Perception, learning and attention," Hear. Res. **238**, 147–154.

Nikulin, M. (**2001**). "Hellinger distance," in *Encyclopaedia of Mathematics*, edited by M. Hazewinkel (Kluwer Academic Publishers, Norwell, MA).

Siebert, W. (**1970**). "Frequency discrimination in the auditory system: Place or periodicity mechanisms?," Proc. IEEE **58**, 723–730.

Soderquist, D., and Lindsey, J. (**1971**). "Physiological noise as a low-frequency masker: The cardiac cycle," J. Acoust. Soc. Am. **50**, 143.

Spiegel, M., and Green, D. (**1981**). "Two procedures for estimating internal noise," J. Acoust. Soc. Am. **70**, 69–73.

Swets, J. (**1959**). "Multiple observations of signals in noise," J. Acoust. Soc. Am. **31**, 514–521.

Tanner, W., Jr. (**1958**). "What is masking?," J. Acoust. Soc. Am. **30**, 919–921.

Tanner, W., Jr., and Birdsall, T. (**1958**). "Definitions of d' and n as psychophysical measures," J. Acoust. Soc. Am. **30**, 922–928.

Viemeister, N. F., and Schlauch, R. S. (**1992**). "Issues in infant psychoacoustics," in *Developmental Psychoacoustics*, edited by L. Werner and E. Rubel (American Psychological Association, Washington, DC), pp. 191–210.

Vogels, R., Spileers, W., and Orban, G. (**1989**). "The response variability of striate cortical neurons in the behaving monkey," Exp. Brain Res. **77**, 432–436.

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

Wickens, T. (**2002**). *Elementary Signal Detection Theory* (Oxford University Press, New York), pp. 114–118.

Wier, C., Jesteadt, W., and Green, D. (**1976**). "A comparison of method-of-adjustment and forced-choice procedures in frequency discrimination," Atten. Percept. Psychophys. **19**, 75–79.

Wightman, F., and Allen, P. (**1992**). "Individual differences in auditory capability among preschool children," in *Developmental Psychoacoustics*, edited by L. Werner and E. Rubel (American Psychological Association, Washington, DC), pp. 113–134.

Wright, B., and Fitzgerald, M. (**2005**). "Learning and generalization of five auditory discrimination tasks as assessed by threshold changes," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigne, S. McAdams, and L. Collet (Springer Science+Business Media, New York), pp. 509–515.

Wright, B., and Zhang, Y. (**2009**). "A review of the generalization of auditory learning," Philos. Trans. R. Soc. London, Ser. B **364**, 301–311.