Published in final edited form as:

J Acoust Soc Am. 2009 April ; 125(4): 2221–2232. doi:10.1121/1.3082102.

Specificity of Binaural Perceptual Learning for Amplitude Modulated Tones: A Comparison of two Training Methods

Daniel Kumpik, Jeremy Ting, Robert A. A. Campbell, Jan W. H. Schnupp, and Andrew J. King

Department of Physiology, Anatomy and Genetics, Sherrington Building, University of Oxford, Parks Road, Oxford OX1 3PT, UK

Abstract

The specificity of auditory perceptual learning has been taken as an indicator of the likely locus within the brain at which underlying neuronal changes occur. This study examined interaural level difference (ILD) discrimination learning with sinusoidally amplitude modulated (SAM) tones and whether training-induced threshold improvements generalize from one side of auditory space to the other and to an untrained carrier frequency. A novel, dual-staircase adaptive method was adopted that was designed to prevent participants from identifying the nature of the adaptive track. ILD thresholds obtained with this method were compared with a constant-stimulus technique using otherwise identical stimuli. Adaptive thresholds derived from psychometric functions were found to be biased compared to those obtained from reversals. Although adaptive and constant-stimulus procedures appeared to yield different temporal patterns of learning, no global differences were found between them in terms of training outcomes. These data show that ILD discrimination learning with SAM tones does generalize to an untrained carrier frequency, but does not generalize across the midline. This implies that the neural substrate for binaural plasticity is found at a relatively high level of the auditory pathway where information is combined across frequency and where each side of auditory space is represented separately.

I. INTRODUCTION

The finding that sound localization performance can improve after perceptual training (for a review, see Wright and Zhang, 2006) is highly relevant for hearing impaired populations, for whom binaural cue values may have been altered as a result of either auditory deprivation or stimulation (Boothroyd, 1993; Philibert *et al.*, 2005; Silman *et al.*, 1984). An understanding of the potential extent and generality of perceptual learning is essential for developing optimal assessment and training strategies for these populations and others, including children with linguistic deficits (Constantinides *et al.*, 2003; Kujala *et al.*, 2001; McArthur and Bishop, 2001). To this end, two aspects of perceptual learning – the time-course over which it occurs and the potential for learning to transfer to other stimuli or tasks – are of crucial importance in determining the clinical utility of auditory plasticity.

The degree of transfer of perceptual learning to untrained tasks and stimuli can also provide valuable insights into the possible locus within the brain at which changes in neural circuitry take place and therefore guide subsequent physiological studies. For example, it has frequently been reported that training adult humans to perform visual discrimination tasks

Correspondence: Dr Andrew King Department of Physiology, Anatomy and Genetics, Sherrington Building, University of Oxford, Parks Road, Oxford OX1 3PT, UK Telephone: +44(0)1865 272523 Fax: +44(0)1865 272469 andrew.king@dpag.ox.ac.uk.

ASA-PACS numbers: 43.66.Pn Binaural hearing 43.66; Qp Localization of sound sources 43.66; Rq Dichotic listening; 43.66.Lj Perceptual effects of sound

leads to improvements that are specific to the location on the visual receptor surface and for the orientation of the trained stimulus (reviewed by Gilbert *et al.*, 2001). This suggests the involvement of early stages in cortical visual processing, where neurons are orientation selective and have small receptive fields that are arranged in a precise, topographic order.

Some of the most dramatic demonstrations of the adaptive capabilities of the adult auditory system have come from studies in which auditory localization cues have been altered (Hofman *et al.*, 1998; Kacelnik *et al.*, 2006; Van Wanrooij and Van Opstal, 2005). However, training can also alter the discrimination of interaural level differences (ILDs) and interaural time differences (ITDs) by normal hearing listeners (Rowan and Lutman, 2006; Spierer *et al.*, 2007; Wright and Fitzgerald, 2001; Zhang and Wright, 2007). Wright and Fitzgerald (2001) found that training-induced improvements in ILD discrimination with pure tones generalized to an untrained standard stimulus location (ILD value), but only at the trained frequency, suggesting that the neural processing of localization cues was altered within frequency-specific channels. On the other hand, generalization across different types of stimuli was reported by Rowan and Lutman (2007), who found that ITD discrimination training with either pure tones or higher-frequency transposed tones led to reduced thresholds for both stimuli as well as for sinusoidally amplitude modulated (SAM) tones with the same modulation rate and carrier frequency as the transposed tones. This suggests that generalization of ILD learning to other stimulus frequencies might also be found if more complex stimuli, such as SAM tones, are used for training. We examined this question in the present study, as well as the extent to which perceptual learning in a binaural discrimination task is specific for spatial location.

There is a growing body of evidence that intensive training can facilitate improvements in perceptual abilities in humans and animals with central auditory processing or peripheral deficits (Kacelnik *et al.*, 2006; Merzenich *et al.*, 1996; Tallal *et al.*, 1996; Zhou and Merzenich, 2008). Various psychophysical methods have been adopted for the study of perceptual learning. In order to maximize the benefits of training in clinical populations, it is important to compare the degree of learning provided by different methods. This issue seems to have been neglected with respect to transfer of auditory perceptual learning. Since the seminal paper of Levitt (1971), adaptive staircase methods (such as the transformed up-down procedure) have become increasingly widespread in the literature (reviewed by Leek, 2001). Compared with constant-stimulus methods, where the range of stimulus values covered by the underlying psychometric function must be chosen and fixed beforehand, staircase methods require fewer restrictive assumptions and target stimulus values efficiently around a specific threshold on the function based on the listener's previous performance.

The relative flexibility and efficiency of adaptive methods are offset by a number of disadvantages, not least the fact that attentive listeners may become aware of the algorithm being used and try to "second-guess" the strategy, which can lead to distraction from the task at hand, possibly affecting the observed threshold. One remedy that has been suggested to address this particular issue is to use one or more simultaneous staircases, each of which governs the stimulus selection in turn from one trial to the next (e.g. Rammsayer, 1992). Another disadvantage is that adaptive tracking techniques tend to focus rapidly on a predetermined threshold value, which can limit their ability to provide accurate information regarding the slope and shape of the underlying psychometric function. A related problem is the fact that adaptive procedures may not allow an observer sufficient exposure to the range of binaural cue values required for learning to occur (Trahiotis *et al.*, 1990), which could influence the observed time-course of learning or the degree of generalization seen. Finally, several simulation studies have shown that psychometric functions obtained from adaptive data can exhibit an inherent bias that causes the slope to be steeper than that of the underlying psychometric function, again potentially affecting threshold estimates (e.g. Leek

et al, 1992; O'Regan and Humbert, 1989; Schlauch and Rose, 1990). This bias appears to be a result of the serial dependency of trials in an adaptive sequence (Kaernbach, 2001) and is exacerbated by other factors, such as the type of staircase, the choice of step size and the number of trials performed (Dai, 1995; Saberi and Green, 1996; Leek *et al.*, 1992). However, obtaining a psychophysical measure of the absolute bias seen in adaptive methods is difficult, unless a full psychometric function is obtained simultaneously. Some studies have reported a 'relative' bias between different adaptive methods (e.g. Rowan *et al.*, 2006), whereas others have found that thresholds obtained from adaptive psychometric functions are similar to those obtained from reversals (Amitay *et al.*, 2006) and to those obtained using a constant-stimulus technique (Dai, 1995).

In this study, we compare the relative impact of using a constant-stimulus versus an adaptive training strategy on the learning time-course and generalization pattern seen after ILD discrimination training with SAM tones. We also used these experiments to extend recent findings on the specificity of perceptual learning by examining whether or not training with stimuli containing binaural cues that favor one side of auditory space can enhance performance for stimuli either with a different carrier frequency or presented on the opposite side.

II. METHOD

A. Subjects

A total of 22 listeners took part in the first experiment, in which we compared adaptive and constant-stimulus algorithms for measuring ILD discrimination thresholds, and 31 in the second experiment, which examined the specificity of ILD learning. Seven listeners took part in both experiments, having completed the learning experiment first. All subjects were assessed before testing to ensure normal bilateral audiometric thresholds (i.e. within 20 dB of normal, from 125 Hz to 8 kHz). Ages were in the range 19-42, with a mean age of 24.5 years. All subjects were right-handed. Subjects were recruited through departmental emails and public notices, and received payment for their contribution. All subjects provided informed consent before participation in the study. Ethical approval was provided by the Central University Research Ethics Committee of the University of Oxford.

B. Apparatus

The software package Matlab (Mathworks, Inc., r.14) was used for stimulus generation, response collection and data organization. Further analyses were conducted in S-Plus (Insightful Corp, r. 6.2). Stimuli were passed to a Tucker-Davis Technologies (Alachua, FL) RM1 mobile processor for D/A conversion, and presented to subjects over headphones (Sony MDR-V500; Tokyo, Japan). Subjects responded by clicking a mouse on one of two response alternatives presented on a flat-screen monitor, which was also used to cue each response period and to provide trial-by-trial feedback. All testing took place in a double-walled sound-attenuating chamber.

C. Stimuli

Prior to testing, head-related transfer functions for 4 kHz and 0.5 kHz pure tones were measured from one of the authors using a similar procedure to that reported previously in the literature (Wightman and Kistler, 1989). From these measurements we estimated the ILD/ITD values associated with 4 kHz and 0.5 kHz tones presented at approximately 5° (for all familiarization runs conducted before testing) and 30° (for the experiment proper) from the midline, and these values were used to offset the overall perceived position of stimulus presentation. These offset ITD and ILD values are presented in Table I.

On each trial, two sequential sinusoidally amplitude modulated (SAM) tones, with a carrier frequency of 4 kHz or 0.5 kHz, modulation depth of 100% and a modulation rate of 80 Hz, were presented binaurally. These stimuli were presented with offset ITDs and ILDs appropriate for sound source directions at either 5° or 30° from the midline. However, while the ITD was held constant throughout the presentation of the stimulus pair, the ILD was varied by a small increment, Δ ILD, between the two stimuli. Positive increments (larger ILD on the second stimulus presentation of the pair) gave the percept of the sound source shifting away from the midline, whereas negative increments produced a perceived shift towards the midline. The size of the increment was varied throughout the experiments, either adaptively using the staircase methods described below or chosen randomly from a constant stimulus set, to determine the subjects' sensitivity to small Δ ILDs. We wished to avoid the use of one fixed standard ILD as one of the stimuli in each trial, as this has been viewed as a potential source of the discrimination learning described in other studies (Karmarkar and Buonomano, 2003). The ILD of the first interval stimulus was therefore the offset value plus half the Δ ILD and that of the second interval stimulus the offset value minus half the Δ ILD, or *vice versa*, depending on the intended, pseudorandomly determined direction of the ILD shift. Thus, the Δ ILDs presented to the subject straddled the offset ILD value appropriate for a sound source located at either approximately 5° or 30° from the midline.

Stimuli were generated digitally, with a sample rate of 48 kHz. Stimulus duration was 400 ms, with an inter-stimulus interval of 500 ms. The SAM tones were ramped on and off in phase with the stimulus envelope, giving rise/fall times of 6 ms. Stimulus intensity was calibrated such that the maximum sound level (offset ILD + $0.5 \times \Delta$ ILD) to be presented to either ear would be 74 dB SPL, and overall presentation level was pseudorandomized at attenuations of 0 dB, -5 dB and -10 dB to prevent the subjects from using monaural level cues, rather than binaural level difference cues.

D. Procedure

Early pilot work using a single-staircase paradigm highlighted the issue of subjects becoming aware of the adaptive algorithm (Fig. 1), a problem that has been described in some other studies (e.g. Rammsayer, 1992). It appeared that, even though subjects were capable of performing the task, once they had discovered the algorithm they found it distracting. At times this led to elevated thresholds, because the subjects diverted their attention to trying to 'second guess' the algorithm, rather than focusing solely on the physical attributes of the stimulus. In the case of the subject represented in Fig. 1, the single-staircase adaptive tracks regularly reversed around stimulus values that were well above the true threshold, based on the performance achieved before the presentation strategy was detected. We therefore devised a similar algorithm that used two staircases running alternately on each consecutive trial. For trial n , stimulus placement was dependent not upon the adaptive track pertaining to $n-1$ trials, but to $n-2$ trials (see Fig. 2). This allowed for consecutive trials in which the Δ ILD value presented was regularly different from one trial to the next and was not dependent upon performance on the previous trial, thus reducing the effects of listener expectancy. Each staircase followed a 3-Down, 1-Up algorithm, with three correct answers leading to a decrease in absolute Δ ILD value between the two stimuli, and one incorrect answer leading to an increase. For each staircase, this method targeted the 79.4% correct point on the psychometric function (Levitt, 1971).

Pilot data had indicated the presence of serious ceiling effects in untrained subjects (probably caused by the difficulty in performing an ILD discrimination task away from the midline, especially for SAM tones with a 0.5 kHz carrier), and adaptive tracks that frequently reached a Δ ILD of 0 dB in highly trained subjects, thus preventing us from obtaining an accurate threshold. We therefore changed the step size logarithmically using a factor of 0.2057 (Saber, 1995). This procedure results in psychometric functions that are

parallel, so that thresholds depend on the performance of the subjects within the range of ILDs tested rather than on the underlying slope of the function. This also prevented the stimulus level from ever reaching zero, reducing the impact of floor effects on the estimated threshold. Logarithmic step sizes were considered to be unnecessary for the constant-stimulus technique, as the selection of stimulus values was predetermined, so that all subjects were exposed to the full range of stimulus values used to plot the psychometric function.

The starting Δ ILD for the first run was 4 dB and the starting value for the next run was set two log-steps above the geometric mean of the reversal-derived thresholds from each staircase. A maximum ceiling Δ ILD of 10.05 dB was applied, equivalent to four log-steps above the starting Δ ILD. Subjects completed 13 reversals per staircase, after which the run was terminated. The constant-stimulus method presented randomized Δ ILD values that ranged from 0 to 4 dB in 0.4 dB steps.

The listener's task was identical across experimental stages, groups and stimulus algorithms. Two SAM tones were presented sequentially on each trial, as described above, differing by only a Δ ILD that was determined by the experimental algorithm. In every condition, binaural cue values were displaced from the midline as described above by the introduction of the base ILD and ITD value, upon which the stimulus ILD values were superimposed. Subjects were required to indicate which of the two stimuli had been furthest from the midline.

For experiment 1, both constant-stimulus and adaptive thresholds were extracted by fitting psychometric functions to the data obtained in each run. For constant-stimulus thresholds, trials in which presentation of the most peripheral or lateral stimulus occurred in the first interval were denoted by positive ILD values, and those in which it occurred in the second interval by negative values. Psychometric functions were then fitted to the data using probit analysis (Finney, 1971; see Fig. 3A). Because we modeled the data based on the distribution of responses for which the more peripheral stimulus was perceived to have been presented first (i.e. "stimulus shifted toward midline" between the first and second intervals), threshold values equivalent to that yielded by the 3-Down, 1-Up adaptive method (79.4%) were obtained by taking the mean of the stimulus values yielding 79.4% and 20.6% correct on the fitted function.

For the adaptive method, psychometric functions were fitted to the data from both adaptive tracks in each run (see Fig. 3B), and then ILD threshold for each was determined as above. The threshold for a given run was defined as the arithmetic mean of these two values. We also extracted reversal-derived thresholds by taking the geometric mean of the final 10 reversals in each staircase, and averaging the values obtained from the 2 interleaved staircases.

For experiment 2, thresholds were extracted as described above, but for the constant-stimulus data we used all of the data from a given session (6 runs) to calculate the daily thresholds. Adaptive thresholds for a given session were defined as the mean of the thresholds obtained from all of the runs performed by the listener during that session.

E. Experimental organization and design

1. Threshold bias in the psychometric function—This experiment was intended to shed light upon the relative integrity of the data obtained from the dual-staircase and constant-stimulus methods, and to provide an indication of the effects of any slope or inherent methodological bias upon measured adaptive thresholds. Subjects were randomly selected to perform one run with each method, with base binaural cue values that were offset

to 5° left or right of the midline. Half the subjects performed the constant-stimulus run followed by the adaptive run and *vice versa*. Because the number of trials in an adaptive run is known to contribute to the bias in the slope of the psychometric function (Kaernbach, 2001; Leek *et al.*, 1992), the number of trials in each constant-stimulus run was made roughly equivalent to the mean number of trials in the adaptive runs, for which the number of trials was variable. This ensured that any bias in thresholds obtained from psychometric functions that was due to differences between the methods could be more accurately assessed, and also ensured that each run lasted a similar length of time. Each constant-stimulus run contained 121 trials, with 11 presentations at each stimulus value. The mean number of trials in each adaptive run was 131.7.

2. Perceptual learning—Listeners were randomly assigned to either the constant-stimulus or dual-staircase algorithm, and also randomly assigned to be either a training subject or a control subject within that algorithm. The experiment was conducted in four phases. All subjects took part in a familiarization phase 2-3 days before the pre-test, designed to acquaint them with the task requirements and the stimuli as far as possible, without inducing binaural learning. To this end, listeners completed 8 runs, each comprising 120 ILD discrimination trials in a session lasting 2-3 hours, with base binaural cue values that were offset to 5° and a constant Δ ILD of 4 dB. This allowed for learning of the task requirements with an easy cue value and at a relatively central position in auditory space, where binaural cue resolution should be well developed (Mills, 1960; Yost, 1974; Yost and Dye, 1998). Runs centered on ILD values corresponding to either 5° left or right of the midline were interleaved throughout this familiarization session, allowing subjects to become accustomed to switching sides during a session in preparation for the pre-test. Subjects were required to score 90% in at least two familiarization runs on each side before progressing to the pre-test.

During the pre-test phase (3-3.5 hours), baseline ILD discrimination thresholds were obtained from each participant for each of three conditions: 1) 4 kHz SAM offset 30° to the left; 2) 4 kHz SAM offset 30° to the right; 3) 0.5 kHz SAM offset 30° to the left. Conditions were always run in that order to ensure that order effects were similar across subjects and groups. Before each condition was run, a short (30-trial) practice run was administered, with a 5° offset and Δ ILD values of 4 dB.

After the pre-test, training subjects underwent 8 days of ILD discrimination training on condition 1 (4 kHz, 30° left). Training took place over the course of 9-10 days, and lasted for approximately 1 hour per day. Control subjects did not participate in the training phase. The post-test phase was identical to the pre-test, and was completed by all subjects 11-14 days after the pre-test.

For the constant-stimulus method, 66 trials were completed in each run, with 6 presentations at each stimulus level. Six runs were completed in each session at all times. For the dual-staircase method, two runs of 13 reversals per staircase were completed per condition during the pre-test and post-test; during training, four runs were completed per session.

F. Analysis

Any fitted psychometric functions that were over-dispersed were excluded from the analysis (in the case of the bias experiment, this meant that a subject's entire dataset would need to be excluded). Over-dispersion is observed if the variance in a dataset is significantly higher than the variance assumed in the probit model. It can arise when the performance of the subject drifts considerably throughout the course of the run, e.g. as a result of wandering attention or fatigue. We therefore considered over-dispersion as diagnostic of potentially unreliable data, and excluded from further analysis runs for which the value of the deviance

for the fitted function was greater than twice the number of degrees of freedom for error. If one of the pair of psychometric functions in an adaptive run was over-dispersed, then that entire run was rejected. Under these criteria, 4 subjects were dropped from the analysis for the bias experiment, leaving 18 subjects to be used in the analysis. Of the remaining data, the degrees of freedom for error for each constant-stimulus function was always 19 and the mean deviance was 20.37 (s.d. = 6.74). The mean degrees of freedom for error for the adaptive functions was 11.33 (s.d. = 4.17), and the mean deviance was 10.5 (s.d. = 4.68).

In the learning experiment, we adopted one further exclusion criterion: a subject's innate inability to perform the task. Subjects whose average pre-test ILD thresholds were greater than two standard deviations above the mean threshold obtained by all (trained and control) listeners within a method were excluded from all analyses. On this basis, three subjects were excluded from the learning experiment, leaving 29 subjects for the analysis. Because in the learning experiment the data from all six runs within a session were used to calculate our constant-stimulus thresholds, the data were well-described by the fitted psychometric functions, and exhibited no over-dispersion. The number of degrees of freedom for error was again always 19, and the mean deviance was 20.36 (s. d. = 7.21), a value that is extremely close to that obtained in the bias experiment and is therefore indicative of good agreement between the two experiments in terms of goodness-of-fit. The mean standard error of threshold estimate per psychometric function was 0.14 dB in condition 1 (on average, 10.5% of threshold obtained for a given run), 0.33 dB in condition 2 (on average, 10.7% of threshold obtained), and 0.61 dB in condition 3 (on average, 16.5% of threshold obtained). For the adaptive method the total proportion of rejected functions was 9.74%. For the remaining functions the mean degrees of freedom was 10.22 (s. d. 2.9) and the mean deviance was 10.69 (s. d. 4.37). The mean standard error of threshold estimate per psychometric function was 0.38 dB in condition 1 (on average, 25.2% of threshold obtained for a given run), 0.59 dB in condition 2 (on average, 29.4% of threshold obtained), and 0.68 dB in condition 3 (on average, 26.6% of threshold obtained).

The extracted data were then subjected to statistical testing that was appropriate to the analysis in question.

III. RESULTS

A. Assessing the adaptive method

1. Threshold bias in the psychometric function—To assess the amount of bias in the adaptive thresholds, we computed the linear regression of constant-stimulus thresholds against thresholds obtained with the adaptive method (estimated either using reversals or probit fits, as described above). Figure 4 compares the thresholds obtained using the two estimation techniques with the corresponding constant-stimulus thresholds. Adaptive thresholds derived by fitting psychometric functions (Fig. 4A) tended to be slightly lower than those obtained from reversals (Fig. 4B). However, for both estimation techniques, regression analysis showed that the intercepts were not significantly different from zero. We therefore constrained the regression intercept term for each to be 0. This yielded r^2 values of 0.96 for each estimation technique, indicating that the relationship between the scores obtained in the constant-stimulus runs and those calculated using either estimation technique from the adaptive data was highly consistent. We then tested the slopes of each of these functions to see if they were significantly different from 1. For the reversal-derived thresholds, this difference was not significant ($t(33) = 0.76$, $p = 0.45$). However, the data derived using thresholds calculated from psychometric functions exhibited a slope that was significantly different from 1 ($t(33) = 3.21$, $p < 0.01$), indicating that, as measured threshold increased, these scores exhibited significant bias compared with the corresponding constant-stimulus scores.

To provide confirmation that the threshold bias described above persisted in the learning data, we also compared the adaptive thresholds derived from psychometric functions with those derived from reversals in the perceptual learning experiment. In the latter case, we collapsed the data across pre-test and post-test for all three experimental conditions. These data provided both a wide range of thresholds and experimental conditions for which this relationship could be compared. Outliers in the learning dataset were identified using a leverage test. If the leverage of any datapoint exceeded $2(p/n)$, where n is the number of observations and p is the number of parameters in the model (i.e. 2), that datapoint was excluded from the analysis (Hoaglin and Welsch, 1978).

Figure 5 compares the psychometric function- and reversal-derived thresholds obtained from both the bias investigation (Fig. 5A; data derived from Fig. 4) and the learning experiment (Fig. 5B). The regression parameters are almost identical to each other, suggesting that the difference between the two estimation techniques seen in Fig. 4 also applied to the learning data. Neither of the functions shown in Fig. 5 have intercepts that are significantly different from 0, so we constrained the intercepts for both to be the same and fitted models with differing slopes to the data obtained for each. These slopes did not differ significantly ($t(184) = 1.88$, $p = 0.06$). It can be seen in Fig. 5B that the slope of the regression line for the learning dataset shows a small but clear tendency for thresholds derived from psychometric functions to be systematically (mean = 8.4%) lower than those obtained from reversals, across the threshold range.

In order to investigate whether or not the psychometric function fitting technique was contributing to the bias observed in these data, we also fitted the pre-test and post-test learning data obtained from the adaptive and constant-stimulus methods using a logit function. Adaptive thresholds calculated using a probit function were, on average, 1.6 times larger than their logit-calculated counterparts, whilst constant-stimulus thresholds were 1.7 times larger when calculated using the probit function. Paired-sample t-tests showed that both the constant-stimulus thresholds ($t(83) = 11.23$, $p < 0.0001$) and the adaptive thresholds ($t(168) = 13.44$, $p < 0.0001$) were significantly lower when calculated using the logit function. To compare how much the choice of fitting technique affected each training method, we calculated difference scores for the thresholds from each fitting technique (probit-derived threshold – logit-derived threshold). A two-sample t-test indicated that the difference scores obtained for each training method were not significantly different from each other ($t(251) = 0.42$, $p = 0.64$).

These analyses imply that the particular fitting method chosen could significantly affect the amount of bias seen for both constant-stimulus and adaptive thresholds. The logit fits yielded lower thresholds, probably due to the relative “straightness” of the slope of the logit function, indicating that they would match the reversal-derived thresholds even less closely. Equivalent threshold differences were found for the two training methods with probit and logit fitting, suggesting that the choice of fitting method alone could not account for the bias we saw in the adaptive thresholds. Thus, the bias seen in experiment 1 is most likely caused by the adaptive method itself.

Because the reversal-derived thresholds more closely matched the constant-stimulus thresholds obtained in experiment 1 and therefore appeared to be free of bias, the adaptive learning thresholds obtained in experiment 2 were calculated from reversals only.

B. Perceptual learning

Across the training period, the overall mean number of trials for each adaptive staircase within a run was 62.5. The mean number of adaptive trials per subject on each day was 501, compared with the fixed value of 396 for the constant-stimulus method.

1. Time-course of learning in the trained condition—Figure 6A,B shows the smoothed individual and group mean ILD learning curves for both methods in the trained condition. Subjects who did not show a numerical improvement in threshold between the pre-test and post-test ($n = 1$ for each method) were excluded from further analyses and are not included in this figure. All other subjects showed a gradual, if variable, reduction in ILD thresholds across the training period. In order to characterize the rate of change in improvement for the two methods, we calculated the normalized percentage change in threshold (defined as $(\text{pre-test threshold} - \text{session } X \text{ threshold}) / \text{pre-test threshold} \times 100$) across sessions for all subjects (Fig. 6C,D). We then fitted linear models, either with or without a quadratic term, to the pooled group mean learning curves using least squares regression. The constant-stimulus data (Fig. 6C) were well fitted by a model with no quadratic term ($r^2 = 0.9$, $t = 8.32$, $p < 0.0001$) and when a quadratic term was added it was found not to be significant ($t = 1.08$, $p = 0.31$), indicating that a linear trend best described the change in thresholds across sessions. The adaptive method, on the other hand, yielded data for which the simplest fit was obtained using a regression model that contained a quadratic term (Fig. 6D; $r^2 = 0.9$, $t = 2.78$, $p < 0.05$).

Because in the adaptive method the number of trials varied from one run to the next, we also fitted the individual normalized learning curves from both techniques against the number of trials completed per session. Individually, the fitted models for the adaptive method were significantly better when they contained a quadratic term in five out of eight cases. In contrast, none of the constant-stimulus learning curves showed a significantly better fit when a quadratic term was included.

These data imply a dissociation between the two methods in terms of the rate of improvement in ILD discrimination performance during training, with the constant-stimulus method exhibiting a linear trend and the adaptive method yielding an improvement that was largely curvilinear. Comparison of Figs. 6A and 6B shows that the adaptive method generally produced thresholds that were slightly elevated compared to those measured with the constant-stimulus method. However, as can be seen from Fig. 6C,D, the overall change in performance between the two methods was very similar.

2. Learning and generalization—The mean thresholds before and after training for each individual subject are shown as pre- vs post-threshold scatter plots in Fig. 7. Trained subjects are plotted in black, untrained controls in gray. Figure 7A shows the scores for stimuli with a 4 kHz carrier lateralized to the left, i.e. they illustrate performance on the trained stimulus condition. It is apparent that the clear majority of trained subjects exhibited improved thresholds after training (all but two of the black symbols lie well below the $x = y$ diagonal). In the “4 kHz Right” condition, the datapoints for the trained cohort lie closer to, and are more evenly scattered around, the $x = y$ diagonal (Fig. 7B), indicating that any improvement gained by training on the left led to only small or no improvement on the right, i.e. there was little or no transfer to the opposite hemifield. By contrast, in the “0.5 kHz Left” condition, we again observed that, in all but one case, the post-test threshold was smaller than the pre-test value (Fig. 7C), suggesting that training with a 4 kHz carrier SAM tone had led to improved performance even when tested with an untrained 0.5 kHz carrier SAM tone. In all three conditions, data from the untrained, control subjects are evenly spread around the $x = y$ diagonal, indicating no systematic difference between pre-test and post-test thresholds (Fig. 7A-C).

t-tests revealed no significant differences between the methods at pre-test for all listeners in any condition, or for either trained or control listeners at post-test in any condition. In order to increase the power of our analysis, we therefore pooled the data from the two methods for each condition and calculated, for each subject, the normalized percentage threshold change

relative to pre-test threshold, again defined as (pre-test threshold – post-test threshold) / pre-test threshold $\times 100$. The observed mean percentage threshold changes were below zero for all groups and all conditions, but only the trained subjects in the “4 kHz Left” ($t(16) = 7.49$, $p < 0.0001$) and the “0.5 kHz Left” ($t(16) = 3.12$, $p < 0.01$) conditions had means that were significantly below zero (confirmed with sign tests; $p = 0.002$ and $p = 0.0003$, respectively). We saw the same pattern when this analysis was conducted using the absolute change in threshold ($t(16) = 5.25$, $p < 0.0001$, and $t(16) = 4.42$, $p < 0.001$ for these two conditions, respectively). This finding was again confirmed with sign tests ($p = 0.002$ for both analyses). Furthermore, t-tests indicated that, for the “4 kHz Left” ($t(27) = 4.18$, $p < 0.001$) and the “0.5 kHz Left” ($t(27) = 2.2$, $p < 0.05$) conditions, the mean percentage change in threshold was significantly greater in the trained cohort than in the untrained controls (confirmed using Wilcoxon rank-sum tests; $p = 0.006$ and $p = 0.0005$, respectively; see Fig. 8). Again, this pattern of generalization was observed when we analyzed the absolute change in threshold (“4 kHz Left”: $t(27) = 2.81$, $p < 0.01$; “0.5 kHz Left”: $t(27) = 3.31$, $p < 0.01$), and was confirmed using Wilcoxon rank-sum tests ($p = 0.006$ and $p = 0.002$, respectively). No significant differences were seen in the “4 kHz Right” condition for these analyses.

The mean change in threshold in the trained “4 kHz Left” condition was -0.95 dB, or a 46% improvement relative to the mean initial threshold. In the “0.5 kHz Left” frequency-transfer condition we observed a -1.32 dB mean change in threshold, corresponding to a 30% improvement. This change was not significantly different from those seen at “4 kHz Left” ($t(16) = 1.72$, $p = 0.11$), indicating that the reduction in thresholds observed with a different carrier frequency was equivalent to that obtained with the trained stimulus.

We also compared the percentage change in threshold between trained and control listeners for each method. The adaptive method yielded significant differences between these groups for the trained “4 kHz Left” condition ($t(13) = 2.59$, $p < 0.05$; confirmed with rank-sum test: $p = 0.02$), but not the “0.5 kHz Left” condition ($t(13) = 1.12$, $p = 0.14$). However, when the absolute change in threshold was used for this analysis, the “4 kHz Left” condition approached significance ($t(13) = 1.67$, $p = 0.06$) and the “0.5 kHz Left” condition was significant ($t(13) = 2.05$, $p < 0.05$; confirmed with rank-sum test: $p = 0.03$). The constant-stimulus percentage change scores were significantly different between the trained and control groups for the “4 kHz Left” condition ($t(12) = 3.27$, $p < 0.01$; confirmed with rank-sum test: $p = 0.005$). In the “0.5 kHz Left” condition a significant difference was also seen between groups ($t(12) = 2.08$, $p < 0.05$), although this was not confirmed with a rank-sum test ($p = 0.1$). When we conducted this analysis using the absolute change in threshold, both the “4 kHz Left” condition ($t(12) = 3.42$, $p < 0.01$; confirmed with rank-sum, $p = 0.01$) and the “0.5 kHz Left” condition ($t(12) = 2.62$, $p < 0.05$; confirmed with rank-sum, $p = 0.02$) condition yielded significant differences between trained and control listeners. We attribute the apparent contradictions in some of these within-methods analyses to the relatively small sample sizes used in these tests, and the accompanying reduction in statistical power. For both methods, no significant differences were observed between trained and control listeners in the “4 kHz Right” condition using any of the above analyses.

Together these results provide strong evidence that training with a 4 kHz SAM tone presented in the left hemifield led to significant improvements in ILD discrimination thresholds, and that these improvements transferred to an untrained carrier frequency presented in the same hemifield, but not to stimuli with the same carrier frequency presented in the opposite hemifield.

IV. DISCUSSION

In this study, we used SAM tones to investigate the time-course and specificity of learning on an ILD discrimination task. Thresholds were estimated using either a standard constant-stimulus method or a novel adaptive task. A methodological investigation was undertaken to compare adaptive thresholds obtained from fitted psychometric functions against those calculated by averaging the reversals in the adaptive track. A comparison was made with thresholds obtained from the same subjects using the constant-stimulus method, and adaptive thresholds were found to be unbiased only when they were calculated from reversals. We also observed differences in the rate of perceptual learning between the two behavioral methods across the training period. However, global learning outcomes in the trained condition were highly comparable in each case. ILD discrimination learning for SAM stimuli whose virtual position was lateralized to the left hemifield of auditory space did not generalize to the opposite hemifield, implying functional segregation of ILD processing for stimuli originating on opposite sides of the midline. However, perceptual transfer of ILD discrimination learning did occur for SAM stimuli presented to the same hemifield used for training, but with a different carrier frequency. Given that ILD learning does not generalize across frequency when pure tones are used (Wright and Fitzgerald, 2001), our results suggest that the amplitude envelope might be a more important cue for ILD discrimination learning than the fine-structure of the stimulus.

A. Methodological assessment

The threshold bias observed when behavioral thresholds were calculated from psychometric functions taken from the adaptive staircase data is consistent with the results of simulations by Leek *et al.* (1992). They found that slope bias in the adaptive psychometric function depends on its underlying true slope, such that, especially for shorter runs, shallower true slopes (and by implication in our analysis, higher thresholds) exhibit more bias than steeper ones (or lower thresholds). This phenomenon is potentially reflected in the thresholds shown in Fig. 4A, where it is clear that the higher the constant-stimulus threshold, the more biased is the corresponding adaptive threshold. This is problematic for studies of learning in general, in which the parameters of the psychometric function are likely to change over a period of time and therefore may exhibit more or less bias from one measurement to the next.

Based on the thresholds derived from reversals, the average bias in the adaptive thresholds derived from psychometric functions equated to 8.4% (Fig. 5). Rowan *et al.* (2006) indirectly assessed relative threshold bias for two adaptive methods (one of which was the 3-Down, 1-Up rule, and both of which are thought to exhibit bias) to be 6.5%. This value is less than that found in our study, as would be expected if the amount by which those adaptive thresholds were biased relative to the underlying psychometric function was within the range observed here.

In contrast to these results, Amitay *et al.* (2006) found little difference between adaptive thresholds obtained from psychometric functions or reversals. Furthermore, Dai (1995) reported that psychometric functions obtained using the constant-stimulus method and both 2-down, 1-up and 3-down, 1-up adaptive methods were equally accurate. However, this conclusion was based upon a comparison between the psychometric functions obtained from each method, and a “total” psychometric function computed by combining the functions derived from all three. If any inherent slope or threshold bias had been present in the adaptive data reported in that study, it may have also been present in the “total” psychometric function.

Simulations carried out by Saberi and Green (1996) indicate that bias in reversal-derived thresholds can be introduced by an inherent imbalance in the relative probabilities of the stimulus level increasing or decreasing at equal distances above or below the target threshold. However, the reversal-derived thresholds in our study showed good agreement with those derived from the constant-stimulus method. In light of this, it seems reasonable to suggest that adaptive thresholds obtained from short runs can be calculated from reversals without compromising the conclusions obtained to any significant degree.

B. Learning in the trained condition

Our results suggest that global learning outcomes for ILD discrimination in the trained condition were relatively unaffected by the choice of psychophysical training method used to induce and measure them. Nevertheless, the learning curves shown in Fig. 6A and B indicate slightly higher thresholds for the adaptive method than for the constant-stimulus method in the earlier training sessions. It is possible that in these sessions the adaptive method was not providing enough exposure to a sufficient range of stimulus levels for maximal learning to occur (e.g. Trahiotis et al, 1990). The relative overall change in performance for the two methods did not, however, differ significantly and post-test thresholds were comparable between them.

The most striking difference between the methods is the temporal pattern of learning shown for each; the rather linear improvement seen for the constant-stimulus method contrasts with the exponential shape of the adaptive learning curve, which shows a large initial drop in threshold but appears to asymptote as training progressed. A similar non-linear improvement in performance was also observed by Wright and Fitzgerald (2001) when ILD discrimination thresholds were measured using an adaptive staircase procedure. The rapid early improvement could reflect either task-specific or early perceptual learning (e.g. Hawkey *et al.*, 2004) (which seems unlikely, given that substantial familiarization with the task was provided before training began). Because an equivalent initial reduction in thresholds was not observed in the constant-stimulus data, it seems more likely that the overall shape of the adaptive learning curve was influenced by the combined strategies of 1) the logarithmically reducing step size, and 2) starting Δ ILD values for each run that were taken as two log-steps above the threshold obtained in the previous run. This could have caused large changes in measured threshold at the beginning of training, when the step size between successive stimulus levels was large, and progressively smaller changes, and therefore an asymptotic learning curve, as subjects became more expert with the task and the range of absolute stimulus levels to which they were exposed became smaller.

The thresholds obtained with the constant-stimulus method still appeared to be slowly improving after eight days of training, suggesting that further ILD learning was possible for the stimuli used in this experiment. If that is the case, then the overall implication is that perceptual learning of ILD discrimination *per se* was relatively linear across time, with small but consistent changes in the degree of learning seen from day to day. Interestingly, the pre-test and post-test thresholds measured in the present study using either training method were lower than those reported by Wright and Fitzgerald (2001), who studied ILD learning with tonal stimuli. Thus, it appears that the use of SAM stimuli may confer an advantage over tonal stimuli for ILD discrimination learning, both in terms of absolute performance thresholds and potentially the degree of learning that is permitted by the underlying neural hardware.

C. Generalization to untrained conditions

Overall, our data do not offer any support for generalization of ILD discrimination learning from one hemifield to the other, implying that perceptual learning for highly lateralized

stimuli may utilize perceptual hardware that is functionally separate. This finding is consistent with electrophysiological (e.g. Moore *et al.*, 1984; Brugge *et al.*, 1996; Mrsic-Flogel *et al.*, 2005; Campbell *et al.*, 2006; Harrington *et al.*, 2008) and behavioural studies (Kavanagh and Kelly, 1987; Kelly and Kavanagh, 1994; Malhotra *et al.*, 2004) in animals, and with psychophysical data in humans (Boehnke and Phillips, 1999), showing that each side of auditory space is represented largely contralaterally at higher levels of the auditory system. However, Zatorre and Penhune (2001) examined spatial hearing in human patients with unilateral temporal lobe lesions and found a dissociation in the pattern of localization deficit when the lesion occurred on the right side of the brain compared with similar damage to the left. They reported that most of the individuals with right auditory cortex lesions were impaired in their ability to perform either a spatial localization or discrimination task in both left and right hemifields. If, as suggested by the results of this study, the right auditory cortex does represent both sides of space, we might expect transfer of ILD discrimination learning across the midline. However, the damaged regions in these patients were relatively large, possibly representing a more extensive region of space than those influenced by sound localization training. Accordingly, our data do not show transfer across the midline, implying that the underlying changes in ILD processing are found at a level of the auditory system where sounds are represented contralaterally.

In contrast to the lack of hemisphere transfer, we found that ILD discrimination learning for high-frequency complex stimuli did generalize to SAM tones with the same modulation rate but a much lower carrier frequency. The same result has been reported by Zhang and Wright (2005), but contrasts with the earlier demonstration by Wright and Fitzgerald (2001) of frequency-specific learning with pure tones. Generalization of perceptual learning is usually interpreted to be indicative of changes at higher loci in the sensory pathway, where information about the trained and generalized stimuli is shared (Gilbert *et al.*, 2001; Gilbert, 1994). Thus, while plasticity of ILD processing for tonal stimuli could potentially occur within any tonotopically-organized structure, beginning with the superior olivary complex where binaural cues are first computed (Yin, 2002), the lack of learning specificity for SAM tones implies the involvement of neurons that integrate information across different frequency channels. This does not mean, however, that binaural learning with more complex sounds, which are lateralized more accurately than pure tones (Bernstein and Trahiotis, 2002) is necessarily restricted to a level of the auditory pathway where frequency selectivity and tonotopicity are lost. Thus, electrophysiological recording studies have shown that neurons in both the inferior colliculus (Biebel and Langner, 2002) and primary auditory cortex (Schulze and Langner, 1999) can be driven by SAM tones with carrier frequencies that lie well outside the excitatory frequency response area defined with tonal stimuli. Although Li *et al.* (2006) found that such stimuli failed to activate inferior colliculus neurons, they did observe responses synchronized to the stimulus envelope when SAM tones with carrier frequencies beyond the excitatory response area were paired with pure tones at the neurons' characteristic frequency. The capacity of these neurons to detect the envelope of the stimulus across a wide range of carrier frequencies is consistent with the transfer of ILD discrimination learning to an untrained carrier frequency (Fig. 8).

Further insights into the neural substrate of auditory perceptual learning will require investigation of the physiological changes that take place as performance improves with training. Although electrophysiological recording studies in animals have so far focused mainly on plasticity of neuronal responses to non-spatial sound attributes, the results of an auditory evoked potential study in humans suggest that short-term training can alter the spatial representation in the auditory cortex (Spierer *et al.*, 2007). While adding to the growing body of evidence that perceptual improvements are associated with cortical plasticity (Dahmen and King, 2007), this does not rule out the possibility that these changes actually originate at a lower level, such as the inferior colliculus.

D. Concluding remarks

Provided that thresholds are calculated from reversals in order to avoid adaptive bias, the effects of training on the measured thresholds were comparable for adaptive and constant-stimulus methods for determining ILD discrimination thresholds. Nevertheless, subtle but important differences persisted between the methods that implicate the influence of the adaptive algorithm upon the measured time-course of learning. Finally, the transfer of learning to low frequency ILD discrimination shows that exposure to stimuli with a periodically fluctuating envelope may allow such changes to occur across a wide range of frequencies, whereas the lack of transfer across the midline appears to restrict the site of plasticity to brain regions where spatial cues are represented contralaterally. In both cases, our findings are consistent with the physiological properties of auditory neurons in the midbrain and cortex.

Acknowledgments

This work was supported by the Wellcome Trust through a Principal Research Fellowship to A. J. King, BBSRC grant BB/D009758/1 to J. W. H. Schnupp and A. J. King, and by the Oxford Centre for Hearing and Balance. We are grateful to John Culling for discussion.

REFERENCES

- Amitay S, Irwin A, Hawkey DJC, Cowan JA, Moore DR. A comparison of adaptive procedures for rapid and reliable threshold assessment and training in naïve listeners. *J. Acoust. Soc. Am.* 2006; 119:1616–10361625. [PubMed: 16583906]
- Bernstein LR, Trahiotis C. Enhancing sensitivity to interaural delays at high frequencies by using ‘transposed stimuli,’. *J. Acoust. Soc. Am.* 2002; 112:1026–1036. [PubMed: 12243151]
- Biebel UW, Langner G. Evidence for interactions across frequency channels in the inferior colliculus of awake chinchilla. *Hear. Res.* 2002; 169:151–168. [PubMed: 12121748]
- Boehnke SE, Phillips DP. Azimuthal tuning of human perceptual channels for sound location. *J. Acoust. Soc. Am.* 1999; 106:1948–1955. [PubMed: 10530019]
- Boothroyd A. Recovery of speech perception performance after prolonged auditory deprivation: case study. *J. Am. Acad. Aud.* 1993; 4:331–337.
- Brugge JF, Reale RA, Hind JE. The structure of spatial receptive fields of neurons in primary auditory cortex of the cat. *J. Neurosci.* 1996; 16:4420–4437. [PubMed: 8699253]
- Campbell RAA, Schnupp JWH, Shial A, King AJ. Binaural-level functions in ferret auditory cortex: evidence for a continuous distribution of response properties. *J. Neurophysiol.* 2006; 95:3742–3755. [PubMed: 16510777]
- Constantinides H, Rose MM, Moore DR. Training in binaural hearing: towards its use in clinical practice. *Int. Congress Series.* 2003; 1254:481–485.
- Dahmen JC, King AJ. Learning to hear: plasticity of auditory cortical processing. *Curr. Opin. Neurobiol.* 2007; 17:456–464. [PubMed: 17714932]
- Dai H. On measuring psychometric functions: a comparison of the constant-stimulus and adaptive up-down methods. *J. Acoust. Soc. Am.* 1995; 98:3135–3139. [PubMed: 8550938]
- Finney, DJ. Probit analysis. 3rd edn. Cambridge University Press; Cambridge, U.K.: 1971.
- Gilbert CD, Sigman M, Crist RE. The neural basis of perceptual learning. *Neuron.* 2001; 31:681–697. [PubMed: 11567610]
- Gilbert CD. Early perceptual learning. *Proc. Natl. Acad. Sci. USA.* 1994; 91:1195–1197. [PubMed: 8108386]
- Harrington IA, Stecker GC, Macpherson EA, Middlebrooks JC. Spatial sensitivity of neurons in the anterior, posterior, and primary fields of cat auditory cortex. *Hear. Res.* 2008 (Epub ahead of print).
- Hawkey DJC, Amitay S, Moore DR. Early and rapid perceptual learning. *Nature Neurosci.* 2004; 7:1055–1056. [PubMed: 15361880]

- Hoaglin DC, Welsch RE. The hat matrix in regression and ANOVA. *Am. Statistician*. 1978; 32:17–22.
- Hofman PM, Van Riswick JGA, Van Opstal AJ. Relearning sound localization with new ears. *Nature Neurosci*. 1998; 1:417–421. [PubMed: 10196533]
- Kacelnik O, Nodal FR, Parsons CH, King AJ. Training-induced plasticity of auditory localization in adult mammals. *PLoS Biol*. 2006; 4:e71, 0627–0638. [PubMed: 16509769]
- Kaernbach C. Slope bias of psychometric function derived from adaptive data. *Percept. Psychophys*. 2001; 63:1389–1398. [PubMed: 11800464]
- Karmarkar UR, Buonomano DV. Temporal specificity of perceptual learning in an auditory discrimination task. *Learning and Memory*. 2003; 10:141–147. [PubMed: 12663752]
- Kavanagh GL, Kelly JB. Contribution of auditory cortex to sound localization by the ferret (*Mustela putorius*). *J. Neurophysiol*. 1987; 57:1746–1766. [PubMed: 3598629]
- Kelly JB, Kavanagh GL. Sound localization after unilateral lesions of inferior colliculus in the ferret (*Mustela putorius*). *J. Neurophysiol*. 1994; 71:1078–1087. [PubMed: 8201403]
- Kujala T, Karma K, Ceponiene R, Belitz S, Turkila P, Tervaniemi M, Naatanen R. Plastic neural changes and reading improvement caused by audiovisual training in reading-impaired children. *PNAS*. 2001; 28:10509–10514. [PubMed: 11517333]
- Leek MR. Adaptive procedures in psychophysical research. *Percept. Psychophys*. 2001; 63:1279–1292. [PubMed: 11800457]
- Leek MR, Hanna TE, Marshall L. Estimation of psychometric functions from adaptive tracking procedures. *Percept. Psychophys*. 1992; 51:247–256. [PubMed: 1561050]
- Levitt H. Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am*. 1971; 49:467–477. [PubMed: 5541744]
- Li H, Sabes JH, Sinex DG. Responses of inferior colliculus neurons to SAM tones located in inhibitory response areas. *Hear. Res*. 2006; 220:116–125. [PubMed: 16945495]
- Malhotra S, Hall AJ, Lomber SG. Cortical control of sound localization in the cat: unilateral cooling deactivation of 19 cerebral areas. *J. Neurophysiol*. 2004; 92:1625–1643. [PubMed: 15331649]
- McArthur GM, Bishop DVM. Auditory perceptual processing in people with reading and oral language impairments: current issues. *Dyslexia*. 2001; 7:150–170. [PubMed: 11765982]
- Merzenich MM, Jenkins WM, Johnston P, Schreiner C, Miller SL, Tallal P. Temporal processing deficits of language-learning impaired children ameliorated by training. *Science*. 1996; 271:77–81. [PubMed: 8539603]
- Mills AW. Lateralisation of high-frequency tones. *J. Acoust. Soc. Am*. 1960; 32:132–134.
- Moore DR, Hutchings ME, Addison PD, Semple MN, Aitkin LM. Properties of spatial receptive fields in the central nucleus of the cat inferior colliculus. II. Stimulus intensity effects. *Hear. Res*. 1984; 13:175–188. [PubMed: 6715264]
- Mrsic-Flogel TD, King AJ, Schnupp JWH. Encoding of virtual acoustic space stimuli by neurons in ferret primary auditory cortex. *J. Neurophysiol*. 2005; 93:3489–3503. [PubMed: 15659534]
- O'Regan JK, Humbert R. Estimating psychometric functions in forced-choice situations: significant biases found in threshold and slope estimations when small samples are used. *Percept. Psychophys*. 1989; 46:434–442. [PubMed: 2813028]
- Philibert B, Collet L, Vesson JF, Veuillet E. The auditory acclimatization effect in sensorineural hearing-impaired listeners: evidence for functional plasticity. *Hear. Res*. 2005; 205:131–142. [PubMed: 15953523]
- Rammsayer TH. An experimental comparison of the weighted up-down method and the transformed up-down method. *Bull. Psychon. Soc*. 1992; 30:425–427.
- Rowan D, Lutman ME. Learning to discriminate interaural time differences at low and high frequencies. *Int. J. Audiol*. 2007; 46:585–94. [PubMed: 17922348]
- Rowan D, Hinton K, Mackenzie E. Comparison of Levitt- and Zwischlocki-type adaptive procedures for stimulus placement in human listeners. *J. Acoust. Soc. Am*. 2006; 119:3538–3541. [PubMed: 16838494]
- Saberi K, Green DM. Adaptive psychophysical procedures and imbalance in the psychometric function. *J. Acoust. Soc. Am*. 1996; 100:528–536. [PubMed: 8675846]

- Saberi K. Some considerations on the use of adaptive methods for estimating interaural-delay thresholds. *J. Acoust. Soc. Am.* 1995; 98:1803–1806. [PubMed: 7560514]
- Schlauch RS, Rose RM. Two-, three-, and four-interval forced-choice staircase procedures. Estimator bias and efficiency. *J. Acoust. Soc. Am.* 1990; 88:732–740. [PubMed: 2212297]
- Schulze H, Langner G. Auditory cortical responses to amplitude modulations with spectra above frequency receptive fields: evidence for wide spectral integration. *J. Comp. Physiol. A. Sens. Neural Behav. Physiol.* 1999; 181:651–664.
- Silman S, Gelfand SA, Silverman CA. Late-onset auditory deprivation: Effects of monaural versus binaural hearing aids. *J. Acoust. Soc. Am.* 1984; 76:1357–1362. [PubMed: 6512097]
- Spierer L, Tardif E, Sperdin H, Murray MM, Clarke S. Learning-induced plasticity in auditory spatial representations revealed by electrical neuroimaging. *J. Neurosci.* 2007; 27:5474–5483. [PubMed: 17507569]
- Tallal P, Miller SL, Bedi G, Byma G, Wang X, Nagarajan SS, Schreiner C, Jenkins WM, Merzenich MM. Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science.* 1996; 271:81–84. [PubMed: 8539604]
- Trahiotis C, Bernstein LR, Buell TN, Spektor Z. On the use of adaptive procedures in binaural experiments. *J. Acoust. Soc. Am.* 1990; 87:1359–1361. [PubMed: 2324398]
- Van Wanrooij MM, Van Opstal AJ. Relearning sound localization with a new ear. *J. Neurosci.* 2005; 25:5413–5424. [PubMed: 15930391]
- Wightman FL, Kistler DJ. Headphone stimulation of free-field listening. I. Stimulus synthesis. *J. Acoust. Soc. Am.* 1989; 85:858–867. [PubMed: 2926000]
- Wright BA, Fitzgerald MB. Different patterns of human discrimination learning for two interaural cues to sound-source location. *PNAS.* 2001; 98:12307–12312. [PubMed: 11593048]
- Wright BA, Zhang Y. A review of learning with normal and altered sound-localization cues in human adults. *Int. J. Aud.* 2006; 45(Supplement 1):S92–S98.
- Yin, TCT. Neural mechanisms of encoding localization cues in the auditory brainstem. In: Oertel, D.; Fay, RR.; Popper, AN., editors. *Integrative functions in the mammalian auditory pathway.* Springer; New York: 2002.
- Yost WA. Discrimination of interaural phase differences. *J. Acoust. Soc. Am.* 1974; 55:1299–1303. [PubMed: 4846726]
- Yost WA, Dye R. Discrimination of interaural differences of level as a function of frequency. *J. Acoust. Soc. Am.* 1988; 83:1846–1851. [PubMed: 3403800]
- Zattore RJ, Penhune VB. Spatial localization after excision of human auditory cortex. *J. Neurosci.* 2001; 21:6321–6328. [PubMed: 11487655]
- Zhang Y, Wright BA. Different interaural level difference processing with complex sounds and pure tones. *J. Acoust. Soc. Am.* 2005; 117:2562.
- Zhang Y, Wright BA. Similar patterns of learning and performance variability for human discrimination of interaural time differences at high and low frequencies. *J. Acoust. Soc. Am.* 2007; 121:2207–2216. [PubMed: 17471734]
- Zhou X, Merzenich MM. Developmentally degraded cortical temporal processing restored by training. *Nature Neurosci.* 2008 In press.

Fig.1.

Example of the confounding nature of awareness of the track in an adaptive staircase method for determining the effects of training on ILD discrimination thresholds. The learning curve represents data from one subject on a single-staircase 3D1U training paradigm. The arrow highlights the first session after the subject became aware of the adaptive algorithm. Errorbars represent one standard error of the reversal points in each session.

Fig. 2. Example of a dual-staircase run. The first staircase (stair1) is denoted with a gray line, the second with a black line (stair2). In this example, the staircases diverge after around 30 trials.

Fig. 3. Examples of psychometric functions fitted using probit analysis for the constant-stimulus method (A) and the dual-staircase method (B). In B, the black line and circles represent the first staircase, the grey line and triangles represent the second staircase.

Fig. 4.

Bias in the adaptive thresholds derived from psychometric functions (A) and reversals (B) as a function of corresponding constant-stimulus score. The fits are for the non-constrained linear regression models. The crosses indicate thresholds measured for a single subject. Dashed lines indicate 95% confidence intervals for the linear regression fits. In A, the confidence intervals intersect and are mostly below the unity line, indicating that adaptive thresholds calculated from psychometric functions were generally biased low compared with corresponding constant-stimulus thresholds obtained from the same subjects. In B, the unity line is completely encompassed by the confidence intervals, indicating that adaptive thresholds were relatively unbiased when calculated from reversals.

Fig. 5.

Comparison of adaptive thresholds calculated from psychometric functions and from reversals for both the bias data (A) and the learning data (B). The dashed lines indicate 95% confidence intervals for the fitted functions; the confidence intervals for the learning data are so close to the function itself that they are only barely visible. For both sets of data, the fitted function lies mostly beneath the unity line, indicating that, when calculated from psychometric functions, the adaptive thresholds from both the bias investigation and the learning experiment exhibited similar degrees of bias compared with reversal-derived thresholds.

Fig. 6. Mean absolute (A, B) and normalized percentage change (C, D) thresholds for each training method as a function of training session. In panel B black dashed lines indicate days for which no data were available for a given subject. The thick black line in each plot indicates the group mean for that day. The relative curvature of the mean learning curve for the adaptive data (B, D) can clearly be seen..

Fig. 7.

Pre- and post-test thresholds for all subjects. Results in each of the stimulus conditions (4 kHz or 0.5 kHz carrier, lateralized left or right) are shown in separate panels, as indicated. Subjects trained on the “4 kHz Left” condition are shown in black, untrained controls in grey. Subjects trained with constant-level stimuli (const) are shown by the circles, subjects trained using the adaptive staircase paradigm (stair) are plotted as triangles. In both the trained “4 kHz Left” and untrained “0.5 kHz Left” conditions, the trained subjects clearly tend to cluster below the unity line, indicating that post-test thresholds were lower than pre-test thresholds. In contrast, control listeners in all three conditions, as well as trained listeners in the “4 kHz Right” condition, exhibit thresholds that cluster around the unity line, indicating little change in thresholds between pre-test and post-test.

Fig. 8.

Box plots showing distribution of threshold changes (post-test minus pre-test threshold expressed as % of pre-test score) for the trained and control cohorts in each of the stimulus configurations. It is clear that only the trained listeners in the “4 kHz Left” and “0.5 kHz Left” conditions showed a negative threshold change between pre-test and post-test.

Table I

Offset binaural cue values used ($\pm$ indicates lateralization favoring right or left hemifield).

4 kHz	ILD offset	ITD offset
5°	± 1.90 dB	± 0.06 ms
30°	± 11.42 dB	± 0.34 ms
0.5 kHz	ILD offset	ITD offset
5°	± 0.70 dB	± 0.06 ms
30°	± 4.23 dB	± 0.36 ms