Exp Brain Res
DOI 10.1007/s00221-007-0898-z

RESEARCH ARTICLE

Perceptual learning: how much daily training is enough?

Beverly A. Wright · Andrew T. Sabin

Received: 31 May 2006 / Accepted: 30 January 2007

Abstract The acquisition of many perceptual skills proceeds over a course of days. However, little is known about how much daily training is needed for such learning to occur. Here we investigated this question by examining how varying the number of training trials per day affected learning over multiple days on two auditory discrimination tasks: frequency discrimination and temporal-interval discrimination. For each task, we compared improvements in discrimination thresholds between different groups of listeners who were trained for either 360 or 900 trials per day for 6 days. Improvement on frequency discrimination required >360 trials of training per day while learning on temporal-interval discrimination occurred with 360 training trials per day, and additional daily practice did not increase the amount of improvement. It therefore appears that the accumulation of improvement over days on auditory discrimination tasks may require some critical amount of training per day, that training beyond that critical amount yields no additional learning on the trained condition, and that the critical amount of training needed varies across tasks. These results imply that perceptual skills are transferred from short- to long-term memory (consolidated) daily, but only if a task-specific initiation requirement has been met.

B. A. Wright (✉) · A. T. Sabin
Department of Communication Sciences and Disorders,
Northwestern University, Evanston,
IL 60208-3350, USA
e-mail: b-wright@northwestern.edu

Introduction

Improvements in performance on many perceptual tasks continue over multiple days of practice. For these improvements to accumulate over days, the learning attributable to each training session must be transferred from short- to long-term memory, a process often referred to as consolidation (McGaugh 2000). Most previous behavioral investigations of consolidation have focused on when and how learning that would normally be consolidated can be disrupted (Brashers-Krug et al. 1996; Shadmehr and Brashers-Krug 1997; Krakauer et al. 1999; Walker et al. 2003; Caithness et al. 2004; Seitz et al. 2005). In contrast, of interest here are the requirements of the training itself that are needed for consolidation to occur, in particular, the extent to which the amount of training per session affects the accumulation of improvements across sessions. To our knowledge, this issue has received little previous attention. However, two potential principles about how the training amount per day affects improvement across days can be inferred from a few recent reports.

First, it appears that the participant may be required to engage in at least some critical amount of training within a day for performance to be better on a following day. Supporting this idea are two recent reports in which the improvements normally observed between a single training session and a test session on a subsequent day did not occur when too few practice trials were provided (Hauptmann and Karni 2002; Hauptmann et al. 2005). In these experiments, reaction times on a letter-enumeration task were faster on the second day than on the first only when there was a critical amount of training on the first day.

Second, it appears that additional training beyond the critical amount may be superfluous. This idea arises from the results of recent investigations showing that increasing the number of training trials per day beyond a minimum number that yielded improvement did not augment learning. For multiple-day training, doubling the amount of daily training did not increase the magnitude of improvement in reaction times on a mirror-reading task (Ofen-Noy et al. 2003) or on any of three performance measures on a timed motor-sequence task (Savion-Lemieux and Penhune 2005). Similarly, providing two or more separate practice sessions on a single day yielded no greater improvement than only one practice session on either a visual texture-discrimination task (Karni and Sagi 1993) or a speech-identification task (Roth et al. 2005).

Our primary aim was to test the possibility that these two principles are general rules of skill acquisition by determining whether they would apply to a different learning situation. Specifically, motivated by the desire to establish the requirements for learning on basic auditory perceptual tasks (Wright and Fitzgerald 2005), we wanted to know whether these principles would apply to learning measured by improvements in discrimination thresholds on auditory tasks trained over multiple days. We also wanted to know whether any observed critical daily training amount would be task dependent and whether that amount could be predicted by within-session performance for the particular tasks that we trained. Therefore, we examined how varying the number of training trials per day (360 vs. 900) affected the improvement over 6 days, as well as within each day, on two auditory tasks using the same standard stimulus and measurement paradigm for both. For the tasks, we selected frequency discrimination (Fig. 1a) and temporal-interval discrimination (Fig. 2a) because learning occurs over multiple days on both of them (frequency-discrimination: Campbell and Small 1963; Delhommeau et al. 2002; Demany 1985; Grimault et al. 2003; Irvine et al. 2000; temporal-interval discrimination: Karmarkar and Buonomano 2003; Wright et al. 1997).

The results provide further evidence that improvement over multiple days requires some critical amount of training per day, and that training beyond that critical amount yields no additional learning on the trained condition. They also reveal that the critical amount of daily training can differ across tasks even when the standard stimulus is the same. Finally, they show that the critical amount of training necessary for improvement to accumulate across days appears to be independent of the performance pattern within each session, at least for the two tasks examined here.

Method

Organization of the experiment

The data for the four trained groups reported here were all originally collected as part of a separate series of experiments designed to investigate other issues. All four groups participated in a familiarization session, a pretest, 6–10 training sessions, and a posttest, each of which occurred on separate, usually consecutive days. In the familiarization session all listeners received 1 h of practice on a variety of psychoacoustic tasks to familiarize them with the laboratory setting and the two-presentation, forced-choice procedure. During that hour, detection thresholds were measured for tones in quiet at 0.25, 0.5, 1, 2, 4, and 8 kHz, and for a 1-kHz tone in forward- and backward-masking conditions. In the pre- and posttests listeners completed five threshold estimates (300 trials) on each of six conditions in a random order (1,800 total trials). The six conditions differed slightly across groups, but all conditions employed either a frequency (Fig. 1a) or temporal-interval (Fig. 2a) discrimination task. During all of the 6–10 daily training sessions, listeners completed either 6 (360 trials) or 15 (900 trials) threshold estimates on a single frequency or temporal-interval discrimination condition from the pretest. Thus, there were a total of four groups (2 daily training amounts × 2 tasks). Because the posttest occurred after different numbers of training days across the four trained groups, controlled across-group comparisons of generalization to untrained conditions were not possible. Therefore, we chose to analyze only performance on the trained condition as measured in the pretest, and in the subsequent 6 days of training (the greatest number of days of training completed by all listeners).

Conditions and stimuli

In both trained conditions we presented two brief tones in each presentation of a two-presentation, forced choice trial. In the frequency-discrimination condition, the two tones were separated by the same, fixed interval (t) in both presentations, but had a standard frequency (f) in one presentation and a lower comparison frequency ($f - \Delta f$) in the other. In the temporal-interval condition, the two tones had the same, fixed frequency (f) in both presentations, but were separated by a standard interval (t) in one presentation, and by a longer comparison interval ($t + \Delta t$) in the other. The interval was measured from the onset of the first tone to the onset of the second tone. The onsets of the initial tones in the first and

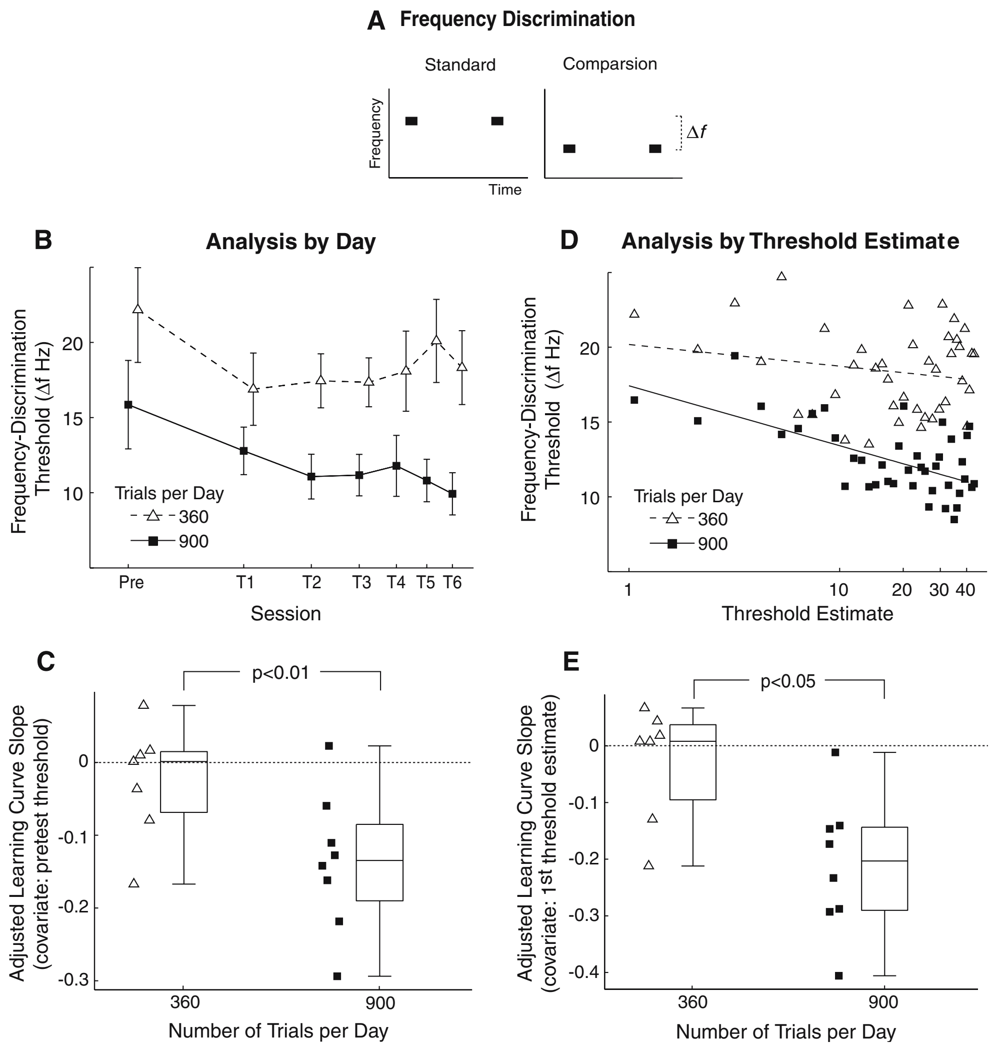

Fig. 1 Learning on frequency discrimination for two different training amounts per day. **a** Schematic diagrams of the stimuli in the standard (*left*) and comparison (*right*) presentations in each forced-choice frequency-discrimination trial. Analysis by day: **b** Group mean frequency-discrimination thresholds (Δf value needed for 79.4% correct discriminations) as a function of the number of sessions. Results are shown separately for listeners who practiced 360 trials/day (*open triangles*; $n = 7$) and those who practiced 900 trials/day (*filled squares*; $n = 8$). *Error bars* indicate ±1 SEM. **c** Regression line slopes of daily mean thresholds versus the log of the session number computed for each listener individually. Results are shown separately for the two trained groups. Slopes were adjusted to account for individual differences in pretest threshold (equation in: Cohen 1988). In each boxplot, the *box* is comprised of *lines* at the upper quartile, median, and lower quartile values, and the *whiskers* extend from each end of the *box* to the maximum and minimum values. Individual data points are horizontally staggered for ease of viewing. Analysis by threshold estimate: **d** Group mean frequency-discrimination threshold as a function of threshold estimate number, plotted separately for each group. **e** Regression line slopes of threshold estimate versus the log of the threshold estimate number computed separately for each listener individually. Results are plotted separately for the two trained groups. Slopes were adjusted to account for individual differences in the first threshold estimate

second presentations were separated by 900 ms. The same standard stimulus ($f = 1$ kHz; $t = 100$ ms) was used in both conditions.

The pretest conditions varied slightly across the listener groups. The two groups who practiced 900 trials/day completed identical pretests consisting of three frequency-discrimination conditions (1 kHz, 50 ms; 1 kHz, 100 ms; 4 kHz, 100 ms) and three temporal-interval-discrimination conditions (1 kHz, 50 ms; 1 kHz, 100 ms; 4 kHz, 100 ms). The pretest for the 360-trial/day frequency-trained listeners was comprised of five frequency-discrimination conditions (1 kHz, 50 ms; 1 kHz, 100 ms; 1 kHz, 200 ms; 4 kHz, 50 ms; 4 kHz, 100 ms) and one temporal-interval-discrimination condition (1 kHz, 100 ms). Finally, for the 360-trial/day interval-trained listeners, the pretest included six interval-discrimination

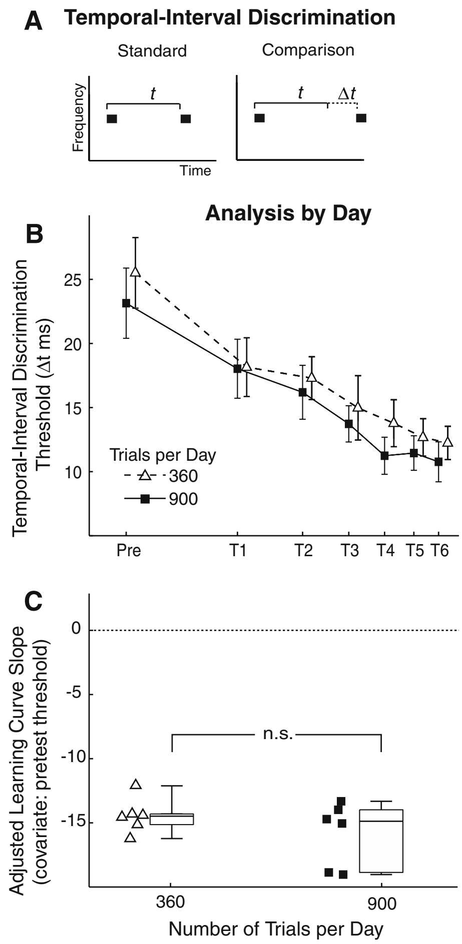

Fig. 2 Learning on temporal-interval discrimination for two different training amounts per day. Same as for Fig. 1a–c, except that the task was temporal-interval discrimination and there were six listeners in each trained group. One listener in the 900 trial/day group did not complete the entire session on training day 2. Her data are excluded from the plots and analysis of learning over a fixed number of threshold estimates

conditions (1 kHz, 50 ms; 1 kHz, 100 ms; 1 kHz, 200 ms; 1 kHz, 500 ms; 4 kHz, 50 ms; 4 kHz, 100 ms). Thus, the pretests for the two groups of frequency-trained listeners shared three frequency-discrimination conditions and one temporal-interval condition, and differed in the two remaining conditions. Likewise, the pretests of the two interval-trained groups shared three temporal-interval conditions, and differed in the three remaining conditions. For all four groups, the order of the pretest conditions was counterbalanced across listeners.

All sounds were generated digitally. Each 86 dB SPL tone was presented in zero phase and had a total duration of 15 ms, including 5-ms raised-cosine on/off ramps. The tone pips were generated using a digital-signal-processing board [Tucker Davis Technologies (TDT) AP2], passed through a 16-bit digital-to-analog converter (TDT DD1), followed by an anti-aliasing filter (8.5-kHz low-pass, TDT FT5), and an attenuator (TDT PA4). The tone pips were then amplified (TDT HB6) and presented over the left ear piece of Sennheiser HD265 headphones in circumaural cushions.

Procedure

We estimated discrimination thresholds using an adaptive procedure with feedback. Listeners pressed a key on a computer keyboard to indicate which of the two randomly selected presentations contained the comparison sound (lower frequency or longer temporal interval). A visual display indicated whether the response was correct or incorrect after every trial throughout the experiment. Threshold estimates were calculated over blocks of 60 trials. The Δf or Δt increased after one incorrect response, and decreased after three consecutive correct responses. When Δf or Δt changed from decreasing to increasing, or vice versa, the value at which that change occurred was labeled a reversal. The first three reversals were discarded, and the mean of the largest remaining even number of reversals was computed. This procedure yielded the value of Δf or Δt that the listener was successfully able to discriminate on 79.4% of trials (Levitt 1971). We refer to that value as the discrimination threshold. Blocks that contained fewer than seven total reversals were excluded from analysis. On the first trial in a block, the comparison frequency or temporal interval was always equal to that of the standard, forcing the listener to guess. The step size was large (5 Hz or 10 ms) until the third reversal, and small (1 Hz or 1 ms) thereafter.

Listeners

Twenty-eight listeners (18 females) with a mean age of 21 years (sd = 3) were paid for their participation. All had normal hearing and no previous experience with psychoacoustic tasks. All procedures were approved by the Northwestern Office for the Protection of Research Subjects. One listener in the 360 trial/day frequency-discrimination group whose pretest threshold was more than 2 standard deviations above the mean of 90 listeners on the same condition was excluded from analysis. Therefore the results reported are for 27 listeners: 360 trial/day groups (frequency: $n = 7$; temporal-interval: $n = 6$), 900 trial/day groups (frequency: $n = 8$; temporal-interval: $n = 6$).

Results

Frequency discrimination

Number of training days fixed

When the number of days of training was held constant at six, listeners who practiced for 900 trials per day showed significant improvements in daily mean thresholds on frequency discrimination while those who practiced for 360 trials per day did not (Fig. 1b). We first examined the performance of the two frequency-trained groups using the log-transformed threshold means from the pretest and each of the 6 days of training for each of the listeners. These means were based on 5 threshold estimates (60 trials per estimate, 300 total trials) per listener for the pretest, and either 6 (360-trial/day) or 15 (900-trial/day) estimates per listener for each training day in all of the analyses. Mean thresholds changed significantly over days for the 900- [from 15.9 to 9.9 Hz; $F(6,42) = 6.63$, $P < 0.0001$], but not for the 360- [from 22.2 to 18.3 Hz; $F(6,36) = 1.38$, $P > 0.05$] trial/day listeners, according to one-way analyses of variance with repeated measures performed separately for each group. Further, the mean thresholds on Training Day 6 were significantly lower for the 900- than the 360-trial/day listeners even when those values were adjusted to take into account individual differences in thresholds on the pretest through an analysis of covariance [$F(1,12) = 9.39$, $P < 0.01$]. We also fitted regression lines, separately for each listener, through the daily mean thresholds against the log of the session number (a power function). For the 900-trial/day group, the slopes of these lines differed significantly from zero [$T(7) = 4.43$, $P < 0.01$] and were negative, indicating learning (Fig. 1c, right side). In contrast, the slopes of the 360-trial/day group did not differ significantly from zero [$T(6) = 0.75$, $P > 0.05$; Fig. 1c, left side]. The slopes also differed significantly between the groups even after using an analysis of covariance to adjust for individual differences in pretest threshold [$F(1,12) = 10.14$, $P < 0.01$]. Thus, the listeners who received more practice per day improved over the 6 days of training while those who received less did not improve. However, because both groups practiced for the same number of days, the greater learning by the 900- than the 360-trial/day listeners cannot be attributed to the number of days of training. This difference instead must arise primarily from either the difference in the total number of training trials (5,700 compared to 2,460) or the number of training trials per day (900 compared to 360).

Total number of training trials fixed

To determine which of these two factors was key, we reexamined the data using analyses parallel to those used for the fixed-days analyses above. However, here, the dependent variable was the individual log-transformed threshold estimate rather than the daily mean estimate, and the total number of training trials, rather than of training days, was held constant between the two groups. The maximum number of threshold estimates on which these comparisons could be made was 41, because that was the total number of estimates from each 360-trial/day listener (5 estimates from the pretest and 6 on each of the 6 training days). Therefore, for these analyses, only the first 41 estimates from the 900-trial/day listeners (through the first 6 estimates on Training Day 3) were considered. We reasoned that if the total number of trials, regardless of how they were distributed across days, was crucial, the 900-trial/day listeners should show no learning over the 41 estimates, just like the 360-trial/day listeners. If, however, the number of trials per day was important, then the 900-trial/day listeners should still show improvement.

Indicating that the key factor was the number of training trials per day, the listeners who received more practice per day improved, while those who received less did not, even when the total number of training trials was held constant between the groups. Performance changed significantly over the first 41 estimates for the 900- [$F(40,280) = 1.72$, $P < 0.01$], but not for the 360- [$F(40,240) = 1.01$, $P > 0.05$] trial/day listeners (Fig. 1d). Even considering individual differences in the first threshold estimate, the 41st estimate was significantly lower for the 900- than the 360-trial/day listeners [$F(1,12) = 9.95$, $P < 0.01$]. In addition, the regression-line slopes of the 900-trial/day listeners differed significantly from zero [$T(7) = 3.40$, $P < 0.05$] and were negative, revealing improvement (Fig. 1e, right side), while those of the 360-trial/day listeners did not differ significantly from zero [$T(6) = 0.93$, $P > 0.05$; Fig. 1e, left side]. The slopes of the two groups, adjusted for differences in the first threshold estimate, also differed significantly from each other [$F(1,12) = 5.95$, $P < 0.05$]. Thus, receiving a sufficient number of training trials per day appears to have been crucial to learning on the frequency-discrimination condition.

Influence of the pretest

Even though the 360-trial/day and the 900-trial/day listeners were exposed to different sets of conditions during the pretest (see Method), this difference had little, if any, effect on the improvement between the pretest and the first day of training. Mean thresholds on the trained frequency-discrimination condition were significantly lower on Training Day 1 than on the pretest for the 360-trial/day [$T(7) = 3.37$, $P < 0.05$], but not for the 900-trial/day [$T(7) = 1.48$, $P > 0.05$] listeners. However, thresholds on Training Day 1 did not differ between the two groups when differences in pretest threshold were taken into account [$F(1,13) = 0.015$, $P > 0.05$]. Thus, any influence of the different mix of pretest conditions for the two groups was fairly minor, and favored the group that did not improve with multiple-hour training.

In addition, removing the pretest from the analyses had little effect on the general conclusions. For the between-group comparisons, omitting the pretest had no effect on the statistical outcomes for any of the by-day or by-block analyses (using T1 thresholds as the covariate). For the within-group analyses, three of the four results that were statistically significant with the pretest either reached or approached significance without it (all P values < 0.085). The one exception was that the thresholds of the 900-trial/day listeners did not change significantly across the individual blocks [$F(35,245) = 1.250$, $P > 0.05$]. The four non-significant results remained so without the pretest.

Within-session analyses

Neither trained group improved significantly within the daily training sessions. To determine whether learning across the daily sessions was related to improvement within the sessions, we reanalyzed the entire data set with an emphasis on the pattern of the threshold estimates obtained within the individual training sessions. We first examined performance separately for each group with a two-way analysis of variance of the individual threshold estimates (900: 15 estimates/day; 360: 6 estimates/day) × 6 training days with repeated measures on day. There was no significant main effect for estimate [900 trials/day: $F(14,105) = 0.37$, $P > 0.05$; 360 trials/day: $F(5, 36) = 0.28$, $P > 0.05$], and no significant day × estimate interaction for either group [900 trials/day: $F(70,525) = 0.91$, $P > 0.05$; 360 trials/day: $F(25,180) = 0.79$, $P > 0.05$]. We also fitted regression lines through the (log-transformed) individual threshold estimates against the log of the estimate number, separately for each listener on each of the six training days. The slopes differed significantly from zero, or nearly so, and were negative only for the 900-trial/day group on the fourth and fifth training days. Further, on the pretest, the individual threshold estimates did not differ significantly from each other, and the slopes of regression lines fitted to those data did not differ significantly from zero, for either group (all $P > 0.41$). Thus, according to these analyses, improvement during a training session was not necessary for learning across sessions on the current frequency-discrimination condition, because even the group that learned over the course of days did not improve systematically within the individual training sessions over the early training days during which the between-session improvements were greatest.

Temporal-interval discrimination

Number of training days fixed

In contrast to the frequency-discrimination results, the daily mean thresholds on temporal-interval discrimination improved significantly over the six training days for both the 360- and the 900-trial/day listeners. We examined performance on the interval-discrimination task using the same analyses as for the frequency-discrimination task, but obtained quite different results. For the interval-discrimination task, mean thresholds changed significantly from the pretest through the 6 days of training with both training regimens [900 trials/day: from 23.1 to 10.7 ms; $F(6,30) = 21.8$, $P < 0.00001$; 360 trials/day: from 25.5 to 12.2 ms; $F(6,30) = 18.84$, $P < 0.00001$; Fig. 2b]. The mean thresholds on Training Day 6, adjusted to take into account individual differences in the pretest threshold, did not differ significantly between the two groups [$F(1,9) = 0.14$, $P > 0.05$]. Further, the regression-line slopes were negative on average and differed significantly from zero for both the 900- [$T(5) = 7.35$, $P < 0.0001$] and the 360- [$T(5) = 8.78$, $P < 0.001$] trial/day listeners, and did not differ significantly between the groups even when individual differences in pretest threshold were taken into account [$F(1,9) = 1.32$, $P > 0.05$; Fig. 2c]. Thus, with 6 days of training, listeners who received less practice per day, and those who received more, improved similarly. This result indicates that 360 trials per day are sufficient for learning on the trained interval-discrimination condition and that additional training beyond 360 trials per day is superfluous.

Influence of the pretest

Including the pretest in the analyses had almost no effect on the conclusions about the influence of multi-hour training on interval discrimination. The exposure to different sets of conditions on the pretest by the two groups seemed to have, at most, only a minor influence on the performance seen on Training Day 1. Mean thresholds were significantly lower on Training Day 1 than on the pretest for both the 360-trial/day listeners [$T(5) = 4.13$, $P < 0.01$], and the 900-trial/day group [$T(5) = 2.88$, $P < 0.05$], and thresholds on Training Day 1 did not differ between the two groups when adjusted to take into account differences in pretest threshold [$F(1,9) = 0.383$, $P > 0.05$]. Further, the same statistical conclusions were reached for all of the by-day and by-block analyses even when the pretest data were removed.

Within-session analyses

Finally, neither trained group showed within-session learning that was predictive of across-session improvement. There was no main effect for estimate (within-day changes in threshold) for either group [900 trials/day: $F(14,60) = 0.50$, $P > 0.05$; 360 trials/day: $F(5,30) = 0.24$, $P > 0.05$]. However, for the 900-trial/day listeners, there was a significant day × estimate interaction [$F(70,300) = 1.35$, $P < 0.05$]. Analysis of this interaction revealed that thresholds only differed across estimates on the first day of training. In addition, the regression-line slopes within sessions differed significantly from zero and were negative only on the first day of training for the 900-trial/day listeners. This isolated within-session learning appears to have had little effect on overall improvement, because both trained groups improved equivalently over the 6 days of training. There also were no statistically significant results on the pretest for either group, either across the individual threshold estimates, or in evaluations of whether the slopes of regression lines fitted to those data differed from zero (all $P > 0.15$). Thus, learning during a training session was not necessary for improvement between sessions on this interval-discrimination condition.

Discussion

The current results are consistent with the idea that listeners must practice at least a critical number of trials per day for improvements to occur over a course of days. This requirement is most clearly demonstrated by the frequency-discrimination data showing that practicing 360 trials per day yielded no learning, while practicing 900 trials per day did, regardless of whether the total number of days (Fig. 1b, c) or of trials (Fig. 1d, e) was held constant. Suggesting a similar requirement for learning on temporal-interval discrimination, Rammsayer (2003) reported no learning in listeners who practiced this task 50 trials per day for 20 days, while here listeners who practiced 360 trials per day for 6 days on a similar condition improved significantly (Fig. 2b, c). Thus, the idea that improvement across days requires a critical amount of practice per day appears to apply to multiple days of training on a perceptual-discrimination task (here), as well as to improvement on a day following a single day of training on a cognitive letter-enumeration task (Hauptmann and Karni 2002; Hauptmann et al. 2005).

It further appears that the critical number of trials needed per day is condition dependent, because there is evidence that it is contingent on both the particular task and stimulus used in training. Clear from the current data is that the critical number required for auditory discrimination learning with a given standard stimulus can differ between tasks. Even though the two conditions trained here used the same standard stimulus, 360 trials per day were sufficient for learning on temporal-interval discrimination, but not on frequency discrimination. Comparison of the current data to a previous report also suggests that the critical number may differ for different stimuli on the same task. Here, listeners who practiced 360 trials per day for 6 days showed no learning on frequency discrimination with a standard stimulus of two 15-ms, 1-kHz tone pips that were separated by 100 ms. In contrast, Roth (2005) reported significant learning in listeners who practiced 350 trials per day for eight days with a 300-ms, 1-kHz standard. Overall, it seems that the task as well as the stimulus used in training can affect the critical number of trials.

The present data are also consistent with the idea that training trials beyond the critical number per day are superfluous to learning on the trained condition. This idea is supported by the temporal-interval results showing that the 360- and the 900-trial/day listeners improved equivalently over the 6 days of training (Fig. 2b, c), suggesting that the trials beyond 360 per day did not contribute significantly to the observed improvement. This result adds to previous reports that increasing the number of training trials per day beyond a minimum number that yielded improvement did not augment learning either after a single day (Karni and Sagi 1993; Roth et al. 2005) or multiple days (Ofen-Noy et al. 2003; Savion-Lemieux and Penhune 2005) of training. However, the conclusion that daily training beyond the critical number of trials provides no benefit to learning should be interpreted with caution. For example, it may be that there are actually two critical numbers of training trials, a smaller one that allows learning to accumulate over days, and a larger one beyond which additional training is superfluous. Further, while trials beyond the critical number do not appear to enhance learning on the trained condition, they may influence some other aspect of learning, such as generalization or retention.

Finally, the present data show no obvious pattern within the individual training sessions that could be used to predict the critical amount of training necessary for improvement to accumulate across days. For both the frequency and the temporal-interval tasks, there was generally no systematic change in performance within each day. In the cases in which there was learning across days, the improvements occurred between, rather than within, the daily training sessions. Similar step-wise improvements have been reported for a visual texture discrimination task (Karni and Sagi 1993). The relatively constant performance within individual training sessions precludes the use of within-day performance as a simple indicator of the number of training trials required for across-day learning. This result differs from those of Hauptman and Karni (2002) and Hauptmann et al. (2005) who reported that only observers who reached asymptotic performance on a letter-enumeration task within the single training session provided showed better performance on the next day. Therefore, the relationship between within- and across-session learning appears to be condition dependent.

Any of a number of factors, individually or in combination, may determine the critical number of trials needed for multiple-day learning on a given condition. For example, it may be that on tasks for which, with effective training, it takes fewer days to reach the best possible performance, a greater amount of daily training is needed. This explanation is consistent with the present data because, compared to the improvements on temporal-interval discrimination, those on frequency discrimination both took fewer days of training to reach asymptote (~2 vs. 4 days) and required more daily training (360 vs. 900 trials). Another possibility is that the critical amount of training required increases as condition difficulty increases. Unfortunately, while differences in difficulty have been shown to affect perceptual learning (Ahissar and Hochstein 1997), their potential contribution to the current results is not clear, because difficulty was not purposefully manipulated. Note, however, that to the extent that difficulty is reflected in the percentage of correct responses, the frequency and temporal-interval conditions were equally difficult because the adaptive procedure used for both targeted the same level of performance (79.4% correct). Nevertheless, it may be that the degree of effort required to reach this percent correct level was greater for the frequency than the interval condition, and that this difference contributed to the greater number of trials needed for improvement on frequency discrimination.

Whatever the contributing factors prove to be, the differences in the critical number of trials needed across conditions must be mediated by differences in the physiology of the neural circuitry engaged by training. Along this line, there is considerable support for the idea that, here, training on the frequency and temporal-interval conditions affected separate neural circuits. At the physiological level, there are numerous reports of hemispheric specialization for the processing of frequency (right hemisphere) and time (left hemisphere) (for review see Zatorre et al. 2002). For example, in one investigation, when listeners made frequency judgments, activation in right posterior auditory cortex increased, while activation instead increased in left posterior auditory cortex when those same listeners performed a duration task with the same set of stimuli (Brechmann and Scheich 2005). Further, it appears that duration, but not frequency, tasks engage the right putamen. This structure processes modality independent information about stimulus timing (Nenadic et al. 2003), forming an interesting connection to the observation that learning on temporal-interval discrimination generalizes from the somatosensory to the auditory system (Nagarajan et al. 1998) and from the auditory to the motor system (Meegan et al. 2000). Also at the behavioral level, learning assessed relative to untrained controls did not generalize between the frequency and temporal-interval tasks (Wright 1998) and followed task-specific patterns of generalization to untrained stimuli (Wright et al. 1997; Wright and Fitzgerald 2005). If learning on these two tasks had modified the same circuitry, then it presumably would have generalized from each task to the other and generalized to the same subsets of untrained stimuli on both tasks. Given this apparent physiological separation, it seems that, of the two circuits, some feature(s) inherent in the one engaged by the present frequency-discrimination training made that circuit less amenable to change, and thus led to the requirement for more trials per day for learning on that condition.

The evidence that the accumulation of improvement on perceptual skills over the course of days requires some critical amount of training per day, and that training beyond that daily amount does not add to the benefit on the trained condition has both practical and theoretical implications. On a practical level, it suggests that perceptual training regimens for therapeutic and non-therapeutic purposes could be made more efficient by establishing and using only the critical number of training trials needed for learning on the condition being trained. It also indicates that care should be taken in assessing the ability to learn a particular task, because while learning may not occur with one number of training trials per day, it may occur with another.

On a theoretical level, this evidence suggests that the consolidation of perceptual learning functions as an all-or-none process. There is no consolidation (at least none that can be identified by the current measures) until the amount of training reaches some critical value, but once that value is reached, consolidation is triggered and there is no benefit from additional training. Most likely, what is actually required to trigger consolidation is the reaching of a critical internal state, brought about by the training. If so, the number of training trials needed for consolidation is apt to differ across individuals. Further, the observation that the critical number of trials differs between tasks even when they are performed with the same standard stimulus, and between different stimuli for the same task, supports the idea that the requirements for consolidation are condition specific. The proposed all-or-none characteristic of consolidation may be one means by which the brain balances the costs and benefits of changing. Excessive plasticity is avoided by changing only in response to sufficient experience, presumably within some restricted time frame, and by limiting the amount of plasticity that can occur as a result of that experience.

Acknowledgments We thank Karen Banai, David Brandel, Julia Huyck, Julia Mossbridge, Jeanette Ortiz, Yuxuan Zhang, and two anonymous reviewers for helpful comments on earlier drafts of this paper. Matt Fitzgerald, Jeanette Ortiz, and Chris Stewart collected much of the data reported here. This work was supported by NIH/NIDCD.

References

Ahissar M, Hochstein S (1997) Task difficulty and the specificity of perceptual learning. Nature 387:401–406

Brashers-Krug T, Shadmehr R, Bizzi E (1996) Consolidation in human motor memory. Nature 382:252–255

Brechmann A, Scheich H (2005) Hemispheric shifts of sound representation in auditory cortex with conceptual listening. Cereb Cortex 15:578–587

Caithness G, Osu R, Bays P, Chase H, Klassen J, Kawato M, Wolpert DM, Flanagan JR (2004) Failure to consolidate the consolidation theory of learning for sensorimotor adaptation tasks. J Neurosci 24:8662–8671

Campbell RA, Small AM (1963) Effect of practice and feedback on frequency discrimination. J Acoust Soc Am 35:1511–1514

Cohen J (1988) Statistical power analysis for the behavioral sciences. In: The concepts of power analysis. Lawrence Erlbaum Assoc, Hillsdale

Delhommeau K, Micheyl C, Jouvent R, Collet L (2002) Transfer of learning across durations and ears in auditory frequency discrimination. Percept Psychophys 64:426–436

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

Grimault N, Micheyl C, Carlyon RP, Bacon SP, Collet L (2003) Learning in discrimination of frequency or modulation rate: generalization to fundamental frequency discrimination. Hear Res 184:41–50

Hauptmann B, Karni A (2002) From primed to learn: the saturation of repetition priming and the induction of long-term memory. Brain Res Cogn Brain Res 13:313–322

Hauptmann B, Reinhart E, Brandt SA, Karni A (2005) The predictive value of the leveling off of within session performance for procedural memory consolidation. Brain Res Cogn Brain Res 24:181–189

Irvine DR, Martin RL, Klimkeit E, Smith R (2000) Specificity of perceptual learning in a frequency discrimination task. J Acoust Soc Am 108:2964–2968

Karmarkar UR, Buonomano DV (2003) Temporal specificity of perceptual learning in an auditory discrimination task. Learn Mem 10:141–147

Karni A, Sagi D (1993) The time course of learning a visual skill. Nature 365:250–252

Krakauer JW, Ghilardi MF, Ghez C (1999) Independent learning of internal models for kinematic and dynamic control of reaching. Nat Neurosci 2:1026–1031

Levitt H (1971) Transformed up-down methods in psychoacoustics. J Acoust Soc Am 49:467–477

McGaugh JL (2000) Memory–a century of consolidation. Science 287:248–251

Meegan DV, Aslin RN, Jacobs RA (2000) Motor timing learned without motor training. Nat Neurosci 3:860–862

Nagarajan SS, Blake DT, Wright BA, Byl N, Merzenich MM (1998) Practice-related improvements in somatosensory interval discrimination are temporally specific but generalize across skin location, hemisphere, and modality. J Neurosci 18:1559–1570

Nenadic I, Gaser C, Volz HP, Rammsayer T, Hager F, Sauer H (2003) Processing of temporal information and the basal ganglia: new evidence from fMRI. Exp Brain Res 148:238–246

Ofen-Noy N, Dudai Y, Karni A (2003) Skill learning in mirror reading: how repetition determines acquisition. Brain Res Cogn Brain Res 17:507–521

Roth DA-E, Kishon-Rabin L, Hildesheimer M, Karni A, Roth DA-Erini, Roth DA (2005) A latent consolidation phase in auditory identification learning: time in the awake state is sufficient. Learn Memory 12:159–164

Savion-Lemieux T, Penhune VB (2005) The effects of practice and delay on motor skill learning and retention. Exp Brain Res 161:423–431

Seitz AR, Yamagishi N, Werner B, Goda N, Kawato M, Watanabe T (2005) Task-specific disruption of perceptual learning. Proc Natl Acad Sci USA 102:14895–14900

Shadmehr R, Brashers-Krug T (1997) Functional stages in the formation of human long-term motor memory. J Neurosci 17:409–419

Walker MP, Brakefield T, Hobson JA, Stickgold R (2003) Dissociable stages of human memory consolidation and reconsolidation. Nature 425:616–620

Wright BA (1998) Generalization of auditory-discrimination learning. In: Association for research in otolaryngology abstracts, p 413

Wright BA, Fitzgerald MB (2005) Learning and generalization on five basic auditory discrimination tasks as assessed by threshold changes. In: Pressnitzer D, de Cheveigne A, McAdams S, Collet L (eds) Auditory signal processing: physiology, psychoacoustics, and models. Springer, New York, pp 510–516

Wright BA, Buonomano DV, Mahncke HW, Merzenich MM (1997) Learning and generalization of auditory temporal-interval discrimination in humans. J Neurosci 17:3956–3963

Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: music and speech. Trends Cogn Sci 6:37–46