



Published in final edited form as:

Exp Brain Res. 2012 May ; 218(4): 567–577. doi:10.1007/s00221-012-3049-0.

Perceptual learning of auditory spectral modulation detection

Andrew T. Sabin¹, David A. Eddins³, and Beverly A. Wright^{1,2}

¹Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL 60208

²Knowles Hearing Center, Northwestern University, Evanston, IL 60208

³Department of Communication Sciences and Disorders, University of South Florida, Tampa, FL 33620

Abstract

Normal sensory perception requires the ability to detect and identify patterns of activity distributed across the receptor surface. In the visual system, the ability to perceive these patterns across the retina improves with training. This learning differs in magnitude for different trained stimuli and does not generalize to untrained spatial frequencies or retinal locations. Here we asked whether training to detect patterns of activity across the cochlea yields learning with similar characteristics. Differences in learning between the visual and auditory systems would be inconsistent with the suggestion that the ability to detect these patterns is limited by similar constraints in these two systems. We trained three groups of normal-hearing listeners to detect spectral envelopes with a sinusoidal shape (spectral modulation) at either 0.5, 1, or 2 cycles/octave and compared the performance of each group to that of a separate group that received no training. On average, as the trained spectral modulation frequency increased, the magnitude of training-induced improvement and the time to reach asymptotic performance decreased, while the tendency for performance to worsen within a training session increased. The training-induced improvements did not generalize to untrained spectral modulation frequencies or untrained carrier spectra. Thus, for both visual-spatial and auditory-spectral modulation detection, learning depended upon and was specific to analogous features of the trained stimulus. Such similarities in learning could arise if, as has been suggested, similar constraints limit the ability to detect patterns across the receptor surface between the auditory and visual systems.

INTRODUCTION

One of the primary functions of a sensory system is to detect and identify patterns of activity distributed across the receptor surface. In the visual system, the activity pattern across the retina reflects the distribution of light in space and provides a primary cue for visual object recognition. In the auditory system, the analogous information is conveyed through the activity pattern across the cochlea, reflecting the peaks and valleys of sound level spread across audio frequency (the spectral envelope). The ability to detect and discriminate visual patterns improves with practice, indicating that the perception of these patterns is malleable (e.g., Mayer, 1983, Sowden et al., 2000, Sowden et al., 2002, Adini et al., 2004, Polat et al., 2004, Yu et al., 2004, Wenger and Rasche, 2006, Zhou et al., 2006, Huang et al., 2008, Huang et al., 2009), but it is not known whether the characteristics of this visual learning are mirrored in the auditory system. Differences in learning between these two systems would imply that different constraints underlie improvements in the ability to detect these patterns, while similarities would be consistent with the idea (Shamma, 2001) that the underlying constraints are comparable. Here, to enable comparison to visual learning, we investigated the extent to which training-induced improvements on an auditory spectral-modulation detection task are influenced by and specific to basic characteristics of the trained stimulus.

In the visual system, training-induced improvements in the ability to detect luminance-defined spatial patterns differ in magnitude for different trained stimuli and fail to generalize to most untrained stimulus features. In the majority of training experiments of visual contrast detection, participants were asked to distinguish a uniform-contrast image from a sinusoidal grating. The amplitude of that sinusoid was varied adaptively to determine the minimum contrast required to detect the grating. Performance on this contrast-detection task gradually improved across multiple days of practice (Mayer, 1983, Sowden et al., 2002, Polat et al., 2004, Wenger and Rasche, 2006, Zhou et al., 2006, Huang et al., 2008). However, the magnitude of the improvement on the trained condition varied across different trained stimuli. Huang et al. (2008), noted large improvements on contrast detection for a spatial frequency of ~ 27 cycles/degree, but only smaller, if any, improvements for a spatial frequency of ~ 10 cycles/degree. Further, the learning was specific to a subset of the features of the stimulus used during training. Sowden et al. (Sowden et al., 2002) reported that improvements were narrowly tuned to the trained spatial frequency (see also Huang et al. (2008)), retinal location, and eye, but generalized broadly to untrained orientations.

In the auditory system, while there is some indication that sensitivity to spectral envelope shape can improve with practice, there has been no investigation of the influence of the trained stimulus on learning or of the generalization of learning to untrained stimulus features. We are aware of only two previous human auditory investigations of the influence of training on spectral envelope shape perception (Kidd et al., 1986, Drennan and Watson, 2001). In both, listeners were asked to distinguish a reference stimulus comprised of multiple simultaneous pure-tone components from a signal that had an increase in the intensity of one of the components, but was otherwise identical to the reference except for a randomly selected higher or lower overall level (profile analysis, (Green, 1987)). Performance on the trained stimulus improved gradually over multiple days, but comparison to the visual results described above is not possible because neither the influence of the characteristics of the trained stimulus nor generalization were tested. Further precluding this comparison, the type of trained stimuli differed considerably between the visual (sinusoidal grating) and auditory (a complex with a single peak) experiments.

Here we examined how practice affected the ability to detect the presence of each of three different auditory spectral shapes and how improvements in that ability generalized to untrained stimuli. To do so, we trained listeners to detect the presence of auditory sinusoidal spectral modulation (Eddins and Bero, 2007), a task that parallels the one used in the visual contrast-detection training experiments (Sowden et al., 2002). In this task, listeners distinguished a reference noise with a flat spectral envelope (Fig. 1, left) from a signal with a spectral envelope that had a sinusoidal shape on a logarithmic frequency axis (Fig. 1, right). The frequency of the sinusoid, the spectral modulation frequency, was measured in cycles/octave (cyc/oct). To investigate whether the particular properties of the stimulus used during training affect learning on this spectral-modulation detection task, we trained three separate groups of listeners, each with a different stimulus, using a multiple-day training regimen. To determine the pattern of generalization to untrained stimuli on this task, before and after training we tested listeners on stimuli that differed from the trained stimulus in carrier spectrum (i.e., cochlear location) and spectral modulation frequency. The learning patterns were qualitatively similar to those previously observed with visual stimuli suggesting that similar factors may limit improvement in the ability to detect patterns of activity distributed across the receptor surface for these two senses.

METHOD

Overview

Three separate groups of trained listeners and two separate groups of controls participated in this study. All trained listeners participated in an initial screening, a pre-training session, a training phase, and a post-training session. During the screening, pure tone detection thresholds were measured at octave frequencies from 250–8000 Hz. In the pre-training session, performance was evaluated on the trained, and two or three untrained, spectral modulation detection conditions. For all trained listeners the training phase consisted of seven daily practice sessions (each approximately 1 hr in length) in which thresholds were measured repeatedly on a single spectral modulation detection condition. That condition differed across the three trained groups. The post-training session followed the training phase and was identical to the pre-training session. The pre-training session and first day of training were conducted on consecutive days, as were the final day of training and the post-training session. The controls participated in all of the same stages, except for the training phase. Thus, any difference between the trained and control groups can be attributed to the training phase. The pre- and post-tests were separated by an average of 15.6 days for the trained listeners and 14.9 days for the controls. The order of the conditions in the pre- and post-training sessions was randomized across listeners, but held constant between the pre- and post-training sessions for each individual listener.

Conditions

The trained condition differed across the three trained groups. The conditions tested in the pre- and post-training sessions were the same for two of the trained groups, but differed for the third, and thus two different control groups were employed. Two of the trained groups practiced detecting either 0.5 ($n = 8$) or 1 ($n = 12$) cyc/oct spectral modulation spanning 200–1600 Hz. During the pre- and post-training sessions, these two groups were tested on their ability to detect 0.5, 1, and 2 cyc/oct spectral modulation spanning 200–1600 Hz as well as 1 cyc/oct spectral modulation spanning 1600–12800 Hz. One group of controls (Control Group 1: $n = 12$) was tested on the same conditions as the 0.5- and 1-cyc/oct trained listeners. The third trained group practiced detecting 2 cyc/oct ($n = 7$) spectral modulation spanning 400–3200 Hz and were tested on their ability to detect 1, 2, and 4 cyc/oct spectral modulation spanning 400–3200 Hz during the pre- and post-training sessions. Another group of controls (Control Group 2: $n = 8$) was tested on the same conditions as the 2-cyc/oct trained listeners. The data from the 2 cyc/oct trained group and corresponding controls were originally gathered as part of a different investigation. We chose to include those data here because they appeared to fit along a continuum with the other trained spectral modulation frequencies despite the differences in carrier spectrum. Subsets of these listeners were also tested on modulation-masking and speech-identification-in-noise conditions before and after training, however for the purposes of this paper we limit our analyses to the spectral modulation detection performance.

Task and Procedure

In the spectral-modulation detection task, listeners had to distinguish a signal, spectrally modulated stimulus (Fig. 1, right) from a reference, flat-spectrum stimulus (Fig. 1, left). Stimuli were presented using a three-alternative, forced-choice method. On a given trial, three intervals, two containing the reference stimulus and one containing the signal were presented in random order. Listeners indicated which of the three intervals contained the signal stimulus by using a computer mouse to click on a visual display. After every trial, visual feedback was provided indicating whether the response was correct or incorrect.

The modulation depth (peak to valley difference in dB) was adjusted adaptively across trials to estimate the spectral modulation detection threshold. Modulation-depth adjustment followed a 3-down/1-up rule and therefore converged on the 79.4% correct point on the psychometric function (Levitt, 1971). The modulation depths at which the direction of change reversed from decreasing to increasing or vice versa are referred to as reversals. The depth was initially 20 dB and was adjusted in steps of 2 dB until the third reversal; subsequent steps were 0.4 dB. In each block of 60 trials, the first three reversals were discarded, and the modulation depths at the largest remaining even number of reversals were averaged and taken as the spectral modulation detection threshold. Blocks that contained fewer than 7 reversals (5 % in total) or single trials that were longer than 20 sec (from the first observation interval through the response, 2 % in total) were excluded from analysis.

During the pre- and post-tests, listeners completed four threshold estimates (240 trials) for each of the tested modulation conditions. During each session of the training phase, listeners completed twelve threshold estimates (720 trials) for the single trained condition.

Stimulus Synthesis

The protocol for stimulus generation was adapted from a previous study on spectral modulation detection (Eddins and Bero, 2007). All stimuli were generated digitally with a sampling period of 24.4 μ s (40983 Hz). An 8192-point buffer was first filled with a sinusoid computed on a \log_2 frequency axis with the appropriate spectral modulation frequency (0.5, 1, 2, or 4 cyc/oct) and modulation depth (expressed in dB). The sinusoid was first multiplied by an equivalently sized buffer filled with random numbers drawn from a Gaussian distribution, and then multiplied by the magnitude response of a Butterworth filter (-32 dB/octave) with cutoff frequencies that were determined by the condition (200–1600 Hz, 1600–12800 Hz, or 400–3200 Hz). The resulting magnitude response was combined with a random phase spectrum and the real inverse Fourier transform was computed. Given the sampling rate and buffer size, the stimulus had a spectral density (spacing of frequency components) of 2.5 Hz. Once in the time domain, the sound was shaped by a 100-ms amplitude envelope with 10-ms raised cosine on/off ramps. Finally the stimuli were scaled to have the same RMS amplitude. Two steps were taken to help prevent the listeners from basing detection on the use of local level cues (comparing the intensity at a single frequency across intervals). First, the phase of the sinusoid that determined the spectral modulation frequency and depth was randomly selected from a uniform distribution spanning $0-2\pi$, causing the spectral peaks and valleys to be located randomly in frequency. Second, the presentation level on each observation interval was roved ± 8 dB around a spectrum level of 35 dB SPL, which corresponds to an overall presentation level between 66.5 and 75.5 dB SPL based on the bandwidth of the carrier. This synthesis procedure was repeated before each stimulus presentation.

Stimulus Presentation

All stimuli were presented using custom software written in MATLAB and played through a 16-bit digital-to-analog converter (Tucker-Davis-Technologies DD1) followed by an anti-aliasing filter with a 16-kHz cutoff frequency (TDT FT6-2), a programmable attenuator (TDT PA4), a sound mixer (TDT SM3), and a headphone driver (TDT HB6). The sounds were presented through the left earpiece of Sennheiser HD265 circumaural headphones. Listeners were tested in a sound-attenuated room.

Listeners

Forty-seven participants (29 female) between 18 and 40 years of age served as listeners. Listeners had normal hearing sensitivity (< 20 dB HL) in the test ear at octave frequencies from 250 to 8000 Hz as measured in the screening session and no previous experience with

psychoacoustic tasks. We also confirmed normal hearing sensitivity up to 12 kHz in the 1 cyc/oct-trained listeners and controls, but did not test the other groups at these higher frequencies due to limited access to the necessary testing equipment. All listeners gave informed consent and were financially compensated for their participation. All procedures were approved by the Institutional Review Board at Northwestern University. Data from listeners whose pre-training thresholds were greater than two standard deviations above the mean of all listeners on a particular condition were removed from analysis of that condition (3.5% of the entire dataset). This exclusion policy was intended to focus the analysis on the effect of training on the typical naïve listener. Removal of these outlying data points did not change the statistical conclusions, with one exception. For the 1 cyc/oct 200–1600 Hz condition, one control listener started very poorly (17.6 dB; >3 standard deviations above the mean of all listeners) and showed a large improvement by the post-training session. When the data from this listener were included, the performance of the listeners who were trained on that condition did not differ significantly from controls.

RESULTS

Performance on the Trained Conditions

Spectral modulation detection thresholds improved with training, but the influence of practice depended upon the trained stimulus. The listeners who were trained using the lowest spectral modulation frequency (0.5-cyc/oct spectral modulation spanning a 200–1600 Hz carrier) improved gradually across multiple sessions. The thresholds of these listeners decreased by 4.9 dB, from 14.4 dB (the highest average pre-training threshold of the three trained groups) to 9.5 dB (Fig. 2A, circles). This improvement was confirmed by both a significant negative slope of a single line fitted to the population of within-listener daily mean thresholds over the \log_{10} of the session number (slope = -4.7 , $p < 0.0001$), and a significant one-way analysis of variance (ANOVA) using session number as a repeated measure ($F_{8,56} = 6.4$, $p < 0.0001$). Both statistics were calculated across all sessions, including the pre- and post-training tests. The controls for this trained group (Control Group 1) also improved between the pre- and post-training tests ($t_{11} = 2.4$, $p = 0.04$) (Fig. 2A, diamonds). However, the magnitude of improvement of the trained listeners was larger than that of the controls, as determined by an analysis of covariance (ANCOVA) computed on the post-q7). As can be seen in the average learning curve, the largest improvement in the trained listeners occurred between the pre-training test and the first training session. Nevertheless, the across-session improvement was still significant when the pre-training test was removed from the analyses (slope = -3.36 ; $p = 0.028$; ANOVA $p = 0.008$).

The listeners who were trained to detect the intermediate spectral modulation frequency (1-cyc/oct spectral modulation spanning a 200–1600 Hz carrier) also improved with practice, but with most of the learning occurring early in training. The thresholds of these listeners decreased from 9.8 dB (the second highest average pre-training threshold) to 7.1 dB, an improvement of 2.7 dB (Fig. 2B, triangles). As a group, these listeners improved significantly when performance was evaluated across all of the sessions including the pre- and post-training tests (slope = -2.1 ; $p = 0.01$; ANOVA; $p = 0.001$). They also improved more than the controls for this trained group (Control Group 1) (ANCOVA; $F_{1,20} = 4.3$, $p = 0.05$) who themselves did not improve between the pre- and post-training tests ($t_{10} = 0.69$, $p = 0.50$). However, when the pre-training test was omitted from the analyses the trained listeners showed no improvement across sessions (slope = -0.76 ; $p = 0.55$; ANOVA; $p = 0.38$), indicating that most of their improvement occurred between the pre-training test and the first training session. The performance of the controls on this condition is particularly interesting in this context. The controls participated in the same pre-training test that seems to have induced the learning in the trained listeners, but did not improve between the pre- and post-training tests. Thus, it appears that by the post-training test the controls lost any

improvements resulting from exposure to the pre-training test, while the practice sessions served to maintain those improvements in the trained listeners.

The listeners who were trained to detect the highest spectral modulation frequency (2-cyc/oct spectral modulation spanning a 400–3200 Hz carrier) did not improve. These listeners began with the lowest average pre-training threshold (6.9 dB) and ended at nearly the same value (6.6 dB) (Fig. 2C, squares). They showed no significant learning either across all sessions (slope = -0.99 ; $p = 0.21$; ANOVA; $p = 0.47$) or when the pre-training test was omitted from the analyses (slope -1.7 ; $p = 0.15$; ANOVA; $p = 0.44$). The controls for this trained group (Control Group 2) also showed no improvement between the pre- and post-training tests ($t_7 = 1.67$, $p = 0.14$).

Analyses of the individual learning-curve slopes support the same conclusions as those reached through the analyses of the average learning curves. For each trained stimulus and individual listener, we computed the slope of a regression line fitted to each threshold estimate over the \log_{10} of the session number. Each slope is a point in Fig. 3. For the 0.5-cyc/oct-trained listeners, the proportion of slopes that were significantly different from zero and negative (filled circles) was the same (0.75) when all sessions were included in the analysis (Fig. 3A, left column) as when the pre-training test was omitted (Fig. 3A, right column). In both cases, the population of slopes was significantly less than zero (pre-training test included: $p = 0.002$; excluded: $p = 0.01$). For the 1-cyc/oct trained listeners, the proportion of significantly negative slopes was greater when the pre-training test was included (0.58) compared to excluded (0.17) (Fig. 3B) and the population of slopes was significantly less than zero only when the pre-training test was included (included: $p = 0.02$; excluded: $p = 0.30$). For the 2-cyc/oct-trained listeners, the same proportion of slopes were significantly negative both with and without the pre-training test, but in neither case was the population of slopes less than zero (all $p > 0.25$, Fig. 3C).

Finally, another difference in the influence of training across the three modulation frequencies was that performance within session did not change consistently for the lowest trained frequency but worsened for the other two frequencies. We investigated whether performance changed systematically within training sessions by computing, for each trained listener on each session, the mean of the first three and the last three threshold estimates. The group averages are plotted in Figure 4A–C. We evaluated within-session performance using a 2 time (first vs. last) by 7 session ANOVA with time as a repeated measure. There was no consistent within-session change in performance for the 0.5-cyc/oct trained listeners (time: $F_{1,49} = 0.86$, $p = 0.36$; time \times session: $F_{6,49} = 1.78$; $p = 0.12$). In contrast, the listeners who practiced either of the two higher modulation frequencies showed a consistent within-session worsening (1 cyc/oct: time: $F_{1,77} = 13.79$, $p < 0.001$; time \times session: $F_{6,77} = 0.07$, $p = 0.99$) (2 cyc/oct: time: $F_{1,42} = 11.7$, $p = 0.001$; time \times session: $F_{6,42} = 0.5$, $p = 0.8$). Therefore, performance deteriorated within sessions for those frequencies for which there was no change in performance across training sessions.

Performance on the Untrained Conditions

No trained group learned significantly more than controls on any untrained spectral modulation detection conditions. We evaluated whether training led to improvements on untrained conditions (generalization) using the same criterion we used to determine the effect of training on the trained condition (a significant ANCOVA between the trained listeners and controls, using pre-training performance as a covariate). The data displayed in Figure 5 show individual and group-mean thresholds at the post-training test, after adjusting the values based on their relationship to the pre-training thresholds. In each panel, the dashed line indicates the average pre-training threshold, and the horizontal box represents the 95% confidence interval of the mean of the controls' post-training thresholds. The

controls improved between the pre- and post-training tests only on the two conditions with the highest pre-training thresholds (0.5 cyc/oct, 200–1600 Hz ($p = 0.04$) and 1-cyc/oct, 1600–12800 Hz ($p < 0.001$) all other $p > 0.14$). The trained groups did not distinguish themselves from the controls on any untrained spectral modulation frequency (Fig. 5, middle columns; all $p > 0.18$) and/or carrier spectrum (Fig. 5, right column; all $p > 0.50$). This lack of training-induced improvement on untrained conditions occurred for the two groups that learned more than controls on their trained condition (0.5 cyc/oct- and 1 cyc/oct-trained listeners) as well as for the group that did not (2 cyc/oct-trained listeners) (Fig. 5, left column). Thus, according to our criterion for generalization, the effect of training was specific to the trained spectral modulation detection condition. Further, for the two trained groups who learned more than controls on their trained conditions (0.5 and 1 cyc/oct), the effect sizes of the between-group (trained vs. control) analyses on every untrained condition were less than half of those on the trained conditions (Fig. 6).

DISCUSSION

The present data demonstrate that the influence of training spectral modulation detection in normal-hearing adults for multiple days is dependent upon the characteristics of the trained stimulus. On average, both the magnitude of training-induced improvement and the time to reach asymptotic performance decreased as the trained spectral modulation frequency increased from 0.5 to 2 cyc/oct. Pre-training thresholds also decreased as spectral modulation frequency increased. Within the training sessions, performance consistently worsened for 1 and 2 cyc/oct, but did not change for 0.5 cyc/oct. Finally, in no case did trained listeners improve significantly more than controls on (generalize to) an untrained spectral modulation frequency or carrier spectrum.

Basis for Improvements

It appears that the present improvements on the trained conditions actually reflect an improved sensitivity to spectral modulation despite several potential alternatives. We considered (and ultimately rejected) two alternative accounts of the present learning. First we asked whether the observed learning might have arisen solely from an improvement in the ability to ignore the randomization of stimulus features. In the current procedure, the spectral modulation phase and presentation level were randomized to minimize the use of local (audio-frequency specific) intensity cues. Thus, one possibility is that these randomizations initially distracted listeners from the target spectral modulation detection task, but that the ability to ignore these randomizations increased with training. However, if that were the case, there would have been similar learning on all trained conditions and complete generalization to all untrained conditions, because each of the conditions used the same randomizations. Instead, listeners only improved on a subset of the trained and tested conditions, making this alternative unlikely.

Second, we asked whether the observed learning might simply have resulted from improved memory of the reference spectrum. Previously observed training-induced improvements in profile analysis (Kidd et al., 1986, Drennan and Watson, 2001) have been attributed to memorization of the reference stimulus, because listeners still improve with practice on a novel reference spectrum after reaching asymptotic performance with a trained reference (Kidd et al., 1986). However, if the present learning were due to the memorization of the reference (a flat-spectrum bandpass noise), it would have generalized to all conditions that employed the same reference. Instead there was no such generalization. The lack of support for these alternative accounts, leads us to the idea that the present learning resulted from the enhancement of sensitivity to spectral modulation.

Stimulus Dependence of Learning

The differences in the learning patterns across the three trained stimuli and lack of generalization among them suggests that different factors may limit performance for different spectral modulation frequencies. Though the trained stimuli differed in carrier spectrum (the spectrum of the trained stimulus for the 2-cyc/oct group was 1 octave higher than that for the 0.5- and 1- cyc/oct groups), there are at least two reasons to think that the differences in learning were instead due to the spectral modulation frequency. First, the rate and magnitude of learning as well as the pattern of within-session performance differed between the 0.5- and 1-cyc/oct trained groups even though the carrier spectrum was the same (200–1600 Hz) for both trained stimuli (Figs. 2–4). In addition, the controls improved more at 0.5 than at 1 cyc/oct with that same carrier spectrum. These differences show that the spectral modulation frequency itself can influence learning separately from the carrier spectrum. Second, there was no consistent relationship between carrier spectrum and improvement. In a direct comparison of improvement with different carriers but the same spectral modulation frequency (the only one possible in this data set), the magnitude of improvement at 1 cyc/oct in controls was greater with the higher (1600–12800 Hz) than the lower (200–1600 Hz) carrier spectrum ($t_6 = 2.58$, $p = 0.03$). This result indicates that the carrier spectrum can affect the magnitude of pre-test induced learning on spectral modulation detection. However, for training-induced learning, the pattern was reversed. The improvement was greater with the lower (200–1600 Hz at 0.5 or 1 cyc/oct) rather than the higher (400–3200 Hz at 2 cyc/oct) carrier spectra (2 group t-tests; all $p < 0.04$). The opposite influence of increasing the frequency range of the carrier spectrum for pre-test and training-induced learning suggests that the carrier spectrum is unlikely to have been the dominant feature that determined the learning pattern. Instead, it appears that the different training-induced learning outcomes with the present three trained stimuli were determined primarily by the spectral modulation frequency.

The stimulus dependence of learning observed here is consistent with evidence that different factors limit spectral modulation detection at different spectral modulation frequencies. Eddins and Bero (Eddins and Bero, 2007) quantified the modulation depth in the excitation pattern (an approximation of the peripheral representation, (Moore and Glasberg, 1987)) at the spectral modulation detection threshold for a range of spectral modulation frequencies. They reasoned that if spectral modulation detection performance for all stimuli were limited by the modulation depth in the excitation pattern (the peak to valley difference), then the threshold level of depth in that pattern would be identical across all detectable spectral modulation frequencies, but this was not the case. For spectral modulation frequencies greater than 2 cyc/oct, listeners were highly sensitive to spectral modulation, requiring about 1 dB of modulation depth in the excitation pattern for detection. However, as the spectral modulation frequency decreased below 2 cyc/oct, detection required increasingly greater modulation depth in the excitation pattern, reaching about 7 dB at 0.25 cyc/oct. A similar analysis reported by Summers and Leek (Summers and Leek, 1994) revealed the same pattern. These analyses suggest that a factor beyond the depth of modulation in the excitation pattern affects the detection of spectral modulation. One possibility is that this modulation is detected by a mechanism that compares the output levels across audio frequency channels to find the peaks and valleys in the excitation pattern, and that the mechanism's capacity to make these comparisons decreases as the channel separation between the peaks and valleys increases (i.e., spectral modulation frequency decreases). If we extend this idea to the current investigation, we can account for the results by assuming that the capacity to compare across nearby channels is already at asymptotic performance before training, but that the capacity to compare across more distant channels can improve with practice.

Interpreted in this context, the current results further suggest that the mechanism modified by training was selective for both spectral modulation frequency and carrier spectrum. If the mechanism were not selective for these features, improvements would have generalized broadly. It is of interest to note that neurons tuned to combinations of spectral modulation frequency and carrier spectrum have been documented in ferret and cat auditory cortex (Schreiner CE and BM, 1994, Shamma et al., 1995, Kowalski et al., 1996, Calhoun and Schreiner, 1998, Klein et al., 2000, Keeling et al., 2008). Selectivity for spectral modulation frequency using stimuli with the same carrier has also been observed in humans. For such stimuli, the ability to detect a target spectral modulation in the presence of an interfering modulation decreases as the modulation frequencies of the target and interferer become more similar (Saoji and Eddins, 2007). There is also evidence that portions of human auditory cortex identified with fMRI are tuned to specific ranges of spectral modulation frequency (Langers et al., 2003, Schonwiesner and Zatorre, 2009). Thus the current training may have resulted from increased sensitivity to modulation in units such as these (or other units with similar selectivity), or from optimizing the weights of such units on a more central decision maker (for discussions of these two views of perceptual learning see (Doshier and Lu, 1998) and (Ahissar and Hochstein, 2004)).

Practical Implications

The present results demonstrate that the detection of spectral modulation at low frequencies (< 1 cyc/oct) can be improved with training, and thus suggest that training might lead to improvement on real-world tasks for which performance is limited by the ability to detect modulation at these frequencies. The ability to detect low spectral modulation frequencies appears to be important for several real-world tasks. For example, in individuals with cochlear implants, the ability to detect low (< 0.5 cyc/oct) spectral modulation frequencies is positively correlated with the ability to identify speech sounds (Litvak et al., 2007, Saoji et al., 2009). This result suggests that for cochlear implant users, improved spectral modulation detection at these frequencies might lead to improvements in speech perception. In addition, in normal-hearing listeners, vertical sound localization appears to depend upon detection of low spectral modulation frequencies < 1 cyc/oct; (Macpherson and Middlebrooks, 2003, Qian and Eddins, 2008), suggesting that improved sensitivity to these modulations might aid performance on this task. However, the specificity of training-induced learning to the trained spectral modulation frequency and carrier spectrum implies that spectral modulation detection training will only be effective if listeners train at the point along these dimensions that is most crucial for the target real-world task.

Comparison to Learning on Visual Contrast Detection

The influence of practice on auditory spectral modulation detection documented here shares several qualitative similarities to that previously reported for the analogous visual task (contrast detection of a sinusoidal grating). In both cases the influence of training depended upon characteristics of the trained stimulus, such that improvements were largest for modulation frequencies where naïve performance was poorest (Huang et al., 2008, Huang et al., 2009). The modulation transfer functions representing naïve performance are “bowl-shaped” as a function of spatial (DeValois and DeValois, 1988) or spectral modulation (Summers and Leek, 1994, Eddins and Bero, 2007) frequency, with performance being best for middle frequencies and worse at either extreme. Training at frequencies near the edges of the transfer function led to large improvements in performance while training at frequencies in the flat portion led to little, if any, improvement in the visual (Huang et al., 2008) as well as the auditory (here) systems. Further, contrast detection improvements in the visual system were specific to the trained spatial frequency and retinal location (Sowden et al., 2002, Huang et al., 2008), and the current auditory improvements were specific to the analogous features: spectral modulation frequency and cochlear location (i.e., carrier spectrum).

Finally, in both cases, the specificity of training-induced improvements resembled the selectivity of neurons in the primary sensory cortex associated with the trained modality (visual (e.g., Tootell et al., 1981) or auditory (e.g., Kowalski et al., 1996)). Thus it appears that improvements in the ability to detect the presence of patterns of activity distributed across the sensory epithelium might be mediated by similar mechanisms in the auditory and visual systems.

Acknowledgments

This work was supported by grants F31DC009549 (ATS) and R01DC004453 (BAW). We thank Cara Depalma for assistance in data collection. Nicole Marrone, Yuxuan Zhang, and Julia Huyck provided helpful comments on earlier drafts of this manuscript.

References

- Adini Y, Wilkonsky A, Haspel R, Tsodyks M, Sagi D. Perceptual learning in contrast discrimination: the effect of contrast uncertainty. *J Vis.* 2004; 4:993–1005. [PubMed: 15669907]
- Ahissar M, Hochstein S. The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci.* 2004; 8:457–464. [PubMed: 15450510]
- Calhoun BM, Schreiner CE. Spectral envelope coding in cat primary auditory cortex: linear and non-linear effects of stimulus characteristics. *Eur J Neurosci.* 1998; 10:926–940. [PubMed: 9753160]
- DeValois, R.; DeValois, K. *Spatial Vision.* New York: Oxford Univ. Press; 1988.
- Dosher BA, Lu ZL. Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proc Natl Acad Sci U S A.* 1998; 95:13988–13993. [PubMed: 9811913]
- Drennan WR, Watson CS. Sources of variation in profile analysis. I. Individual differences and extended training. *J Acoust Soc Am.* 2001; 110:2491–2497. [PubMed: 11757938]
- Eddins DA, Bero EM. Spectral modulation detection as a function of modulation frequency, carrier bandwidth, and carrier frequency region. *J Acoust Soc Am.* 2007; 121:363–372. [PubMed: 17297791]
- Green, DM. *Profile Analysis : Auditory Intensity Discrimination.* Oxford, UK: Oxford University Press; 1987.
- Huang CB, Lu ZL, Zhou Y. Mechanisms underlying perceptual learning of contrast detection in adults with anisotropic amblyopia. *J Vis.* 2009; 9:24, 21–14. [PubMed: 20053087]
- Huang CB, Zhou Y, Lu ZL. Broad bandwidth of perceptual learning in the visual system of adults with anisotropic amblyopia. *Proc Natl Acad Sci U S A.* 2008; 105:4068–4073. [PubMed: 18316716]
- Keeling MD, Calhoun BM, Kruger K, Polley DB, Schreiner CE. Spectral integration plasticity in cat auditory cortex induced by perceptual training. *Exp Brain Res.* 2008; 184:493–509. [PubMed: 17896103]
- Kidd G Jr, Mason CR, Green DM. Auditory profile analysis of irregular sound spectra. *J Acoust Soc Am.* 1986; 79:1045–1053. [PubMed: 3700859]
- Klein DJ, Depireux DA, Simon JZ, Shamma SA. Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *J Comput Neurosci.* 2000; 9:85–111. [PubMed: 10946994]
- Kowalski N, Depireux DA, Shamma SA. Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *Journal of neurophysiology.* 1996; 76:3503–3523. [PubMed: 8930289]
- Langers DR, Backes WH, van Dijk P. Spectrotemporal features of the auditory cortex: the activation in response to dynamic ripples. *NeuroImage.* 2003; 20:265–275. [PubMed: 14527587]
- Levitt H. Transformed up-down methods in psychoacoustics. *J Acoust Soc Am.* 1971; 49(Suppl 2): 467. [PubMed: 5541744]
- Litvak LM, Spahr AJ, Saoji AA, Fridman GY. Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners. *J Acoust Soc Am.* 2007; 122:982–991. [PubMed: 17672646]

- Macpherson EA, Middlebrooks JC. Vertical-plane sound localization probed with ripple-spectrum noise. *J Acoust Soc Am.* 2003; 114:430–445. [PubMed: 12880054]
- Mayer MJ. Practice improves adults' sensitivity to diagonals. *Vision research.* 1983; 23:547–550. [PubMed: 6880052]
- Moore BC, Glasberg BR. Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. *Hear Res.* 1987; 28:209–225. [PubMed: 3654390]
- Polat U, Ma-Naim T, Belkin M, Sagi D. Improving vision in adult amblyopia by perceptual learning. *Proc Natl Acad Sci U S A.* 2004; 101:6692–6697. [PubMed: 15096608]
- Qian J, Eddins DA. The role of spectral modulation cues in virtual sound localization. *The Journal of the Acoustical Society of America.* 2008; 123:302–314. [PubMed: 18177160]
- Saoji AA, Eddins DA. Spectral modulation masking patterns reveal tuning to spectral envelope frequency. *The Journal of the Acoustical Society of America.* 2007; 122:1004–1013. [PubMed: 17672648]
- Saoji AA, Litvak L, Spahr AJ, Eddins DA. Spectral modulation detection and vowel and consonant identifications in cochlear implant listeners. *The Journal of the Acoustical Society of America.* 2009; 126:955–958. [PubMed: 19739707]
- Schonwiesner M, Zatorre RJ. Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. *Proc Natl Acad Sci U S A.* 2009; 106:14611–14616. [PubMed: 19667199]
- Schreiner CE, BMC. Spectral envelope coding in cat primary auditory cortex: properties of ripple transfer functions. *Audit Neurosci.* 1994; 1:39–61.
- Shamma S. On the role of space and time in auditory processing. *Trends Cogn Sci.* 2001; 5:340–348. [PubMed: 11477003]
- Shamma S, Versnel H, Kowalski N. Ripple analysis in ferret primary auditory cortex. I. Response characteristics of single units to sinusoidal rippled spectra. *Audit Neurosci.* 1995; 1
- Sowden PT, Davies IR, Roling P. Perceptual learning of the detection of features in X-ray images: a functional role for improvements in adults' visual sensitivity? *J Exp Psychol Hum Percept Perform.* 2000; 26:379–390. [PubMed: 10696624]
- Sowden PT, Rose D, Davies IR. Perceptual learning of luminance contrast detection: specific for spatial frequency and retinal location but not orientation. *Vision research.* 2002; 42:1249–1258. [PubMed: 12044757]
- Summers V, Leek MR. The internal representation of spectral contrast in hearing-impaired listeners. *J Acoust Soc Am.* 1994; 95:3518–3528. [PubMed: 8046143]
- Tootell RB, Silverman MS, De Valois RL. Spatial frequency columns in primary visual cortex. *Science.* 1981; 214:813–815. [PubMed: 7292014]
- Wenger MJ, Rasche C. Perceptual learning in contrast detection: presence and cost of shifts in response criteria. *Psychon Bull Rev.* 2006; 13:656–661. [PubMed: 17201366]
- Yu C, Klein SA, Levi DM. Perceptual learning in contrast discrimination and the (minimal) role of context. *J Vis.* 2004; 4:169–182. [PubMed: 15086307]
- Zhou Y, Huang C, Xu P, Tao L, Qiu Z, Li X, Lu ZL. Perceptual learning improves contrast sensitivity and visual acuity in adults with anisometropic amblyopia. *Vision research.* 2006; 46:739–750. [PubMed: 16153674]

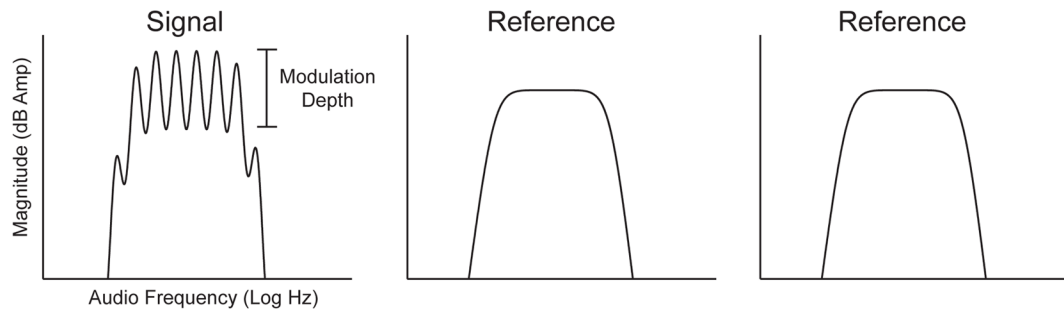


Figure 1. Spectral modulation detection

Schematic diagrams of the stimuli in the spectral-modulation detection task. Listeners had to distinguish a noise with a sinusoidal spectral shape over a logarithmic frequency axis (signal; left) from noises with a flat spectrum (references; middle and right) in each three-interval forced-choice trial. The signal interval was chosen at random. The modulation depth was varied adaptively to determine the spectral modulation detection threshold.

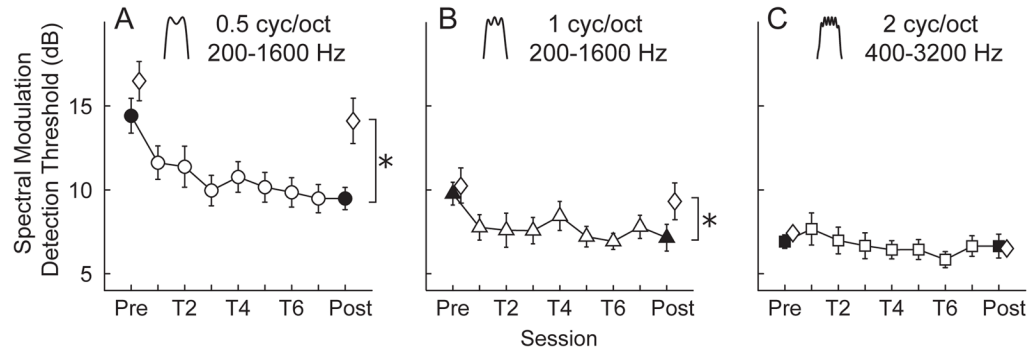


Figure 2. Learning Curves

The group average spectral modulation detection thresholds (79% correct performance) as a function of testing session for listeners trained at 0.5 (A, circles; $n=8$), 1 (B, triangles; $n=12$), and 2 (C, squares; $n=7$) cyc/oct as well as for controls who received no training (diamonds; $n=8-12$). The schematics near the top left corner of each panel illustrate the trained stimulus for that panel, with audio frequency on the abscissa and magnitude on the ordinate (axes not shown; see Figure 1). Error bars indicate one standard error of the mean. Asterisks indicate a significant difference ($p < 0.05$) between trained listeners and controls based on analyses of covariance using pre-training performance as a covariate. The 0.5- and 1-cyc/oct trained listeners learned significantly more than controls between the pre- and post-tests, while the 2-cyc/oct trained listeners did not.

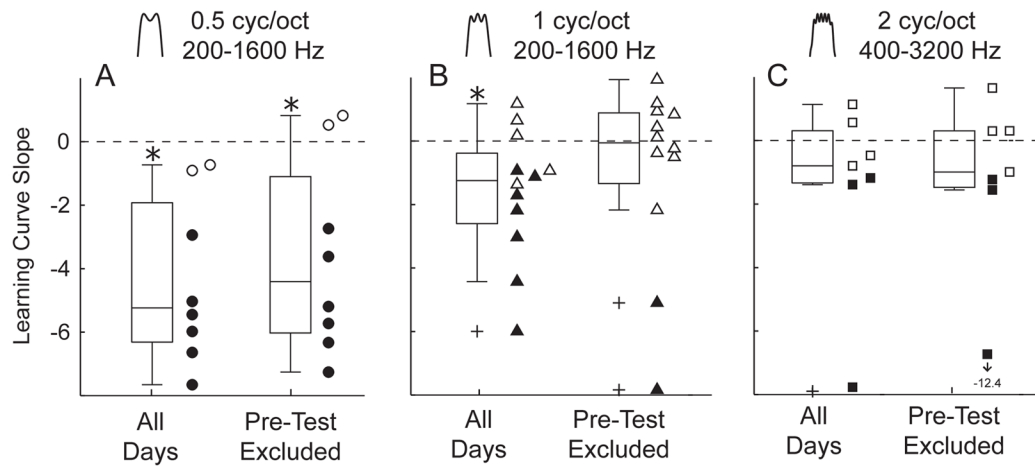


Figure 3. Learning Curve Slopes

The slopes of regression lines fitted to the daily mean thresholds over the \log_{10} of the session number for each listener in the 0.5- (A), 1- (B), and 2- (C) cyc/oct trained groups. Filled symbols indicate that the slope for that listener was significant and negative. The boxplots to the left of the individual points reflect the distribution of points, where 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 (excluding outliers). Slopes were computed either across all sessions (left in each panel) or across all sessions excluding the pre-test (right in each panel). Asterisks indicate that the population of slopes was significantly ($p < 0.05$) less than zero according to a one-sample t-test. When all sessions were considered, the population of slopes was significantly negative for the 0.5- and 1-cyc/oct trained groups, but not the 2-cyc/oct trained group. Only the population of slopes for the 0.5-cyc/oct trained group was significantly negative when the pre-training session was excluded.

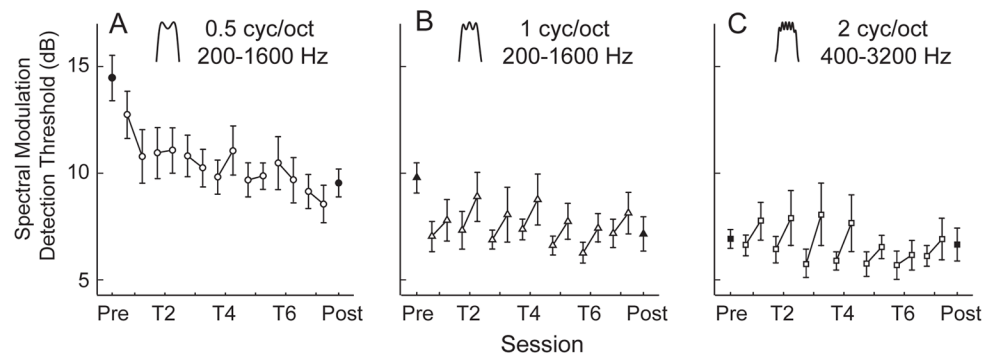


Figure 4. Within-Session Performance

The group-average spectral modulation detection thresholds from the beginning and the end of each training session. The averages of the first three (left open symbols) and last three (right open symbols) threshold estimates from the same training session are connected by a line. The average pre- and post-training thresholds are also shown (filled symbols). The results are plotted separately for the 0.5- (A, circles), 1- (B, triangles), and 2- (C, squares) cyc/oct trained listeners. Error bars indicate one standard error of the mean. The 0.5-cyc/oct trained listeners showed no consistent within-session performance change, while the 1- and 2- cyc/oct-trained listeners consistently got worse within each session.

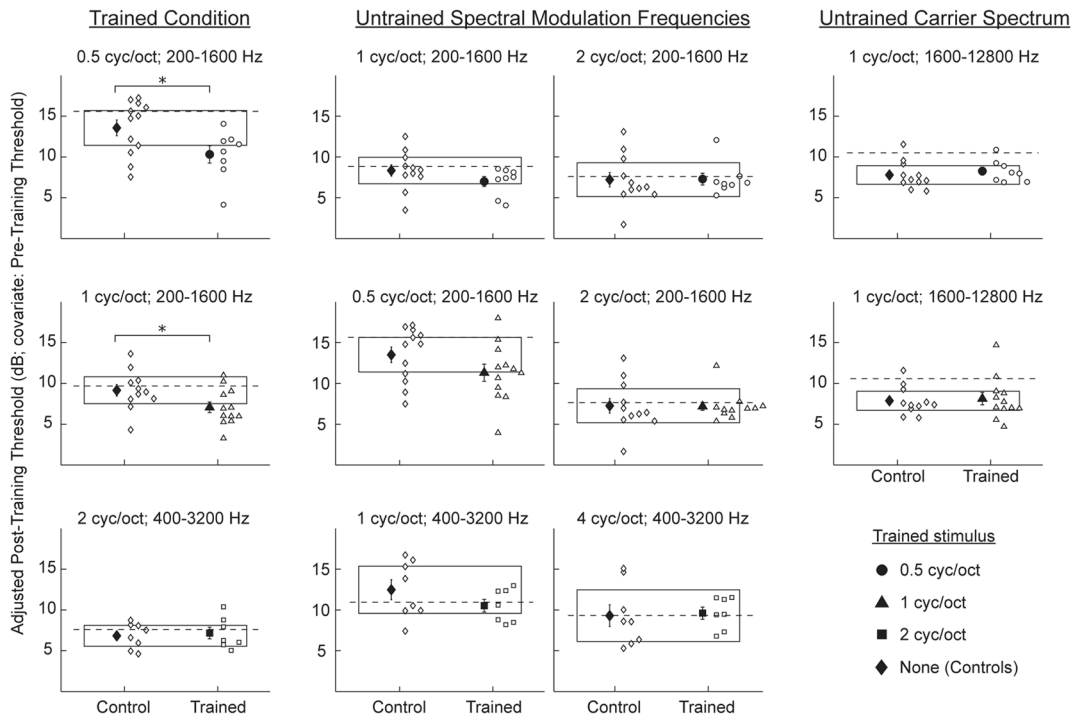


Figure 5. Untrained Spectral Modulation Detection Conditions

Post-training spectral modulation detection thresholds for each tested condition for the 0.5- (circles; top row), 1- (triangles; middle row), and 2- (squares; bottom row) cyc/oct trained listeners as well as for the controls (diamonds). Results are shown for the group averages (filled symbols) and the individual listeners (unfilled symbols). The thresholds were adjusted using pre-training performance as a covariate. The horizontal boxes represent the 95% confidence interval of the post-training performance of the controls, and the dashed lines represent the mean pre-training performance across groups. Error bars indicate one standard error of the mean. Asterisks indicate a significant ($p < 0.05$) difference between trained listeners and controls based on an analysis of covariance using pre-training performance as a covariate. Performance was evaluated for the trained condition (left column), untrained spectral modulation frequencies (middle columns), and an untrained carrier spectrum (right column). The top right panel is for both an untrained spectral modulation frequency and an untrained carrier spectrum. There were no untrained conditions on which trained listeners distinguished themselves significantly from controls.

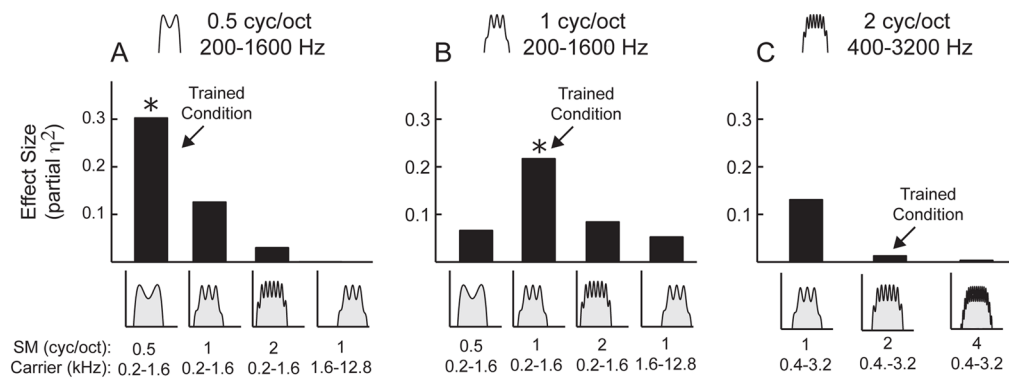


Figure 6. Effect Sizes for All Trained vs Control Comparisons

Partial η^2 effect sizes for analyses of covariance (ANCOVA) comparing post-training thresholds between trained listeners and controls using pre-training thresholds as a covariate. Effect sizes are shown for each condition tested in the 0.5-cyc/oct (A), 1-cyc/oct (B), and 2-cyc/oct (C) trained groups. Asterisks indicate a significant ($p < 0.05$) ANCOVA. For the two groups that learned on their trained conditions (0.5 and 1 cyc/oct), the effect sizes on those conditions were at least twice as large as the effect sizes on the untrained conditions tested in the same group.