

Perceptual Learning Evidence for Tuning to Spectrotemporal Modulation in the Human Auditory System

Andrew T. Sabin,¹ David A. Eddins,^{3,4} and Beverly A. Wright^{1,2}

¹Department of Communication Sciences and Disorders and ²Knowles Hearing Center, Northwestern University, Evanston, Illinois 60208, and

³Department of Communication Sciences and Disorders and ⁴Global Center for Hearing and Speech Research, University of South Florida, Tampa, Florida 33620

Natural sounds are characterized by complex patterns of sound intensity distributed across both frequency (spectral modulation) and time (temporal modulation). Perception of these patterns has been proposed to depend on a bank of modulation filters, each tuned to a unique combination of a spectral and a temporal modulation frequency. There is considerable physiological evidence for such combined spectrotemporal tuning. However, direct behavioral evidence is lacking. Here we examined the processing of spectrotemporal modulation behaviorally using a perceptual-learning paradigm. We trained human listeners for ~ 1 h/d for 7 d to discriminate the depth of spectral (0.5 cyc/oct; 0 Hz), temporal (0 cyc/oct; 32 Hz), or upward spectrotemporal (0.5 cyc/oct; 32 Hz) modulation. Each trained group learned more on their respective trained condition than did controls who received no training. Critically, this depth-discrimination learning did not generalize to the trained stimuli of the other groups or to downward spectrotemporal (0.5 cyc/oct; -32 Hz) modulation. Learning on discrimination also led to worsening on modulation detection, but only when the same spectrotemporal modulation was used for both tasks. Thus, these influences of training were specific to the trained combination of spectral and temporal modulation frequencies, even when the trained and untrained stimuli had one modulation frequency in common. This specificity indicates that training modified circuitry that had combined spectrotemporal tuning, and therefore that circuits with such tuning can influence perception. These results are consistent with the possibility that the auditory system analyzes sounds through filters tuned to combined spectrotemporal modulation.

Introduction

Most natural sounds have peaks and valleys of intensity that vary across both frequency and time. Accurate perception of these spectrotemporal modulations is necessary for fundamental auditory skills such as the discrimination and grouping of auditory objects (Bregman, 1990; Woolley et al., 2005) and the perception of speech (Elhilali et al., 2003). It has been proposed that the perception of spectrotemporal modulation is mediated by a bank of filters, with each filter tuned to a particular combination of a spectral and a temporal modulation frequency (Chi et al., 1999, 2005). At present, there is considerable physiological, but little behavioral, evidence for such spectrotemporal tuning.

A schematic of the proposed spectrotemporal modulation filter bank is shown in Figure 1. Each spectrogram depicts the spectrotemporal modulation to which a particular filter is tuned. This model has been used to describe human performance on monaural phase-detection (Carlyon and Shamma, 2003) and speech-perception (Elhilali et al., 2003) tasks. These demonstrations

suggest that combined spectrotemporal processing could underlie perception, but do not provide a direct test of that possibility. Current evidence for combined spectrotemporal tuning comes only from physiological observations. Such tuning has been observed in individual neurons in inferior colliculus (Versnel et al., 2009), thalamus (Miller et al., 2002), and auditory cortex (Kowalski et al., 1996; Depireux et al., 2001), as well as in population responses identified with fMRI in human auditory cortex (Langers et al., 2003; Schönwiesner and Zatorre, 2009). Investigations focusing on a single modulation dimension have also revealed tuning along the tested dimension (for temporal, see Langner and Schreiner, 1988; for spectral, see Shamma et al., 1995). Behavioral support for modulation tuning comes from modulation-frequency-specific effects of masking (for temporal, see Bacon and Grantham, 1989; for spectral, see Saoji and Eddins, 2007), adaptation (for temporal, see Wojtczak and Viemeister, 2003; for spectral, see D.A. Eddins and R. Harwell, unpublished observations), expectation (for temporal, see Wright and Dai, 1998), and training (for temporal, see Fitzgerald and Wright, 2005, 2011; for spectral, see Sabin et al., 2012) on modulation perception. However, in all of these cases only one dimension of modulation was manipulated, thereby preventing potential observations of combined spectrotemporal tuning.

As a potential means to observe behavioral evidence of combined spectrotemporal tuning in humans, we examined the influence of training on the discrimination of modulation depth. We reasoned that if training with a particular combination of

Received Nov. 15, 2011; revised Feb. 19, 2012; accepted March 8, 2012.

Author contributions: A.T.S., D.A.E., and B.A.W. designed research; A.T.S. performed research; A.T.S., D.A.E., and B.A.W. analyzed data; A.T.S. and B.A.W. wrote the paper.

This work was supported by NIH/NIDCD Grants F31DC009549 (A.T.S.) and R01DC004453 (B.A.W.). We thank Drs. Julia Huyck, Nicole Marrone, and Yuxuan Zhang for helpful discussion.

Correspondence should be addressed to Andrew T. Sabin, Department of Communication Sciences and Disorders, 2240 Campus Drive, Northwestern University, Evanston, IL 60208. E-mail: a-sabin@northwestern.edu.

DOI:10.1523/JNEUROSCI.5732-11.2012

Copyright © 2012 the authors 0270-6474/12/326542-08\$15.00/0

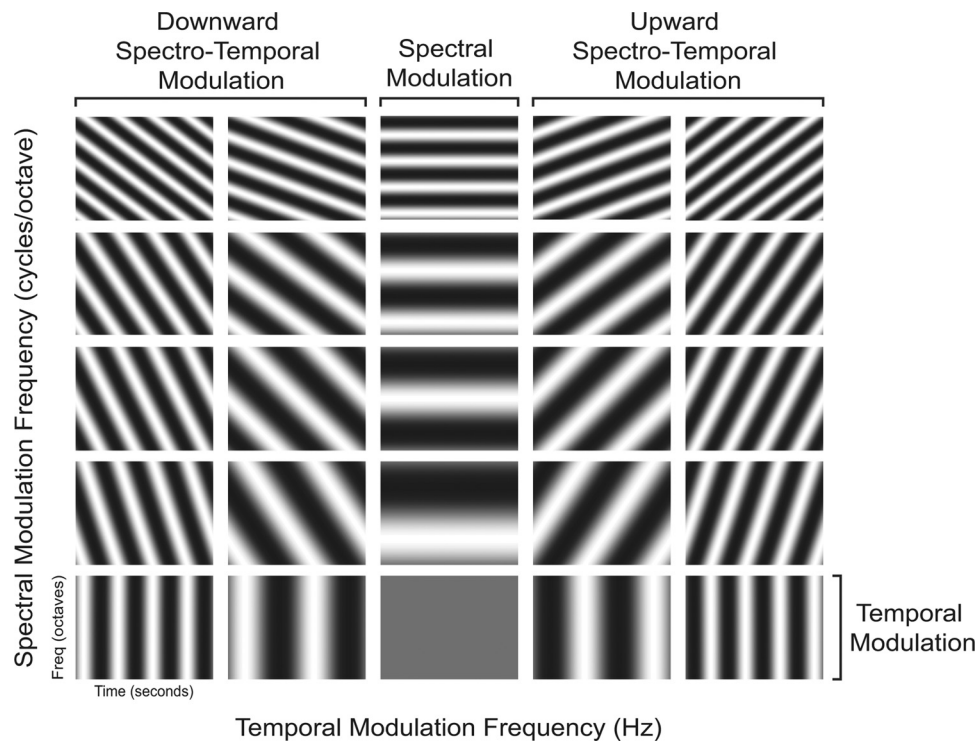


Figure 1. Spectrotemporal modulation filter bank. A schematic of hypothetical filters in a spectrotemporal modulation filter bank. Each spectrogram depicts the spectrotemporal modulation to which a particular filter is tuned. Each filter has a preferred spectral modulation frequency (vertical spacing of bars), temporal modulation frequency (horizontal spacing of bars), and direction (up or down: direction of bar tilt). These filters are tuned to spectral-alone (middle column), temporal-alone (bottom row), or spectrotemporal (downward: left two columns; upward: right two columns) modulation. Note that the filters that are selective for isolated spectral or temporal modulation frequencies simply constitute one slice of this spectrotemporal modulation filter bank where the preferred modulation frequency in the other dimension is zero.

spectral and temporal modulation frequencies modified processing that was tuned to that combination, then training-induced improvements would not generalize to any untrained combinations. Such specificity could arise from a change in a modulation filter itself, or in the weight assigned to that filter by a decision maker. We report that training to discriminate the depth of spectral-only, temporal-only, or spectrotemporal modulation led to depth-discrimination learning that was largely restricted to the trained combination of modulation frequencies and had different influences on the ability to detect spectrotemporal modulation ranging from improvement to worsening. This result provides some behavioral support for the proposal that filters tuned to combinations of spectral and temporal modulation underlie modulation perception.

Materials and Methods

Overview. We tested three separate groups of trained listeners and one group of controls. All trained listeners participated in an initial screening, a pretest, a training phase, and a posttest. During the screening, we measured pure-tone detection thresholds at octave frequencies from 250 to 8000 Hz. In the pretest, performance was evaluated on five modulation-perception conditions, including the trained one. For the trained listeners, the training phase consisted of seven daily practice sessions (each ~1 h in length) in which thresholds were measured repeatedly on a single modulation-depth-discrimination condition. The trained condition differed across the three trained groups. The posttest followed the training phase and was identical to the pretest. 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.

Conditions. The pretests and posttests consisted of four modulation-depth-discrimination conditions and one modulation-detection condition. Each modulation was characterized by both a spectral modulation

frequency [in cycles/octave (*cyc/oct*)] and a temporal modulation frequency (in Hz). The four depth-discrimination conditions were spectral modulation alone (0.5 *cyc/oct*, 0 Hz), temporal modulation alone (0 *cyc/oct*, 32 Hz), upward spectrotemporal modulation (0.5 *cyc/oct*, 32 Hz), and downward spectrotemporal modulation (0.5 *cyc/oct*, –32 Hz). The detection condition was upward spectrotemporal modulation (0.5 *cyc/oct*, 32 Hz). We selected 0.5 *cyc/oct* and 32 Hz as the trained modulation frequencies because we anticipated that we would observe learning with them. Each of these frequencies falls at a point near the edge of its respective modulation-detection transfer function. There is previous evidence from both the auditory (A.T. Sabin, unpublished observations) and visual (Huang et al., 2008) systems that the amount of learning increases as the modulation frequency approaches the edge of the modulation-detection transfer function, at least for modulation detection. Each of the three trained groups practiced a different depth-discrimination condition from the pretests and posttests: spectral modulation alone ($n = 8$), temporal modulation alone ($n = 8$), or upward spectrotemporal modulation ($n = 8$). Tests of vowel and consonant identification in noise were also included in the pretests and posttests, but those data are not reported here.

Tasks and procedure. For the modulation-depth-discrimination conditions, listeners had to distinguish a modulated noise stimulus with a 50 dB depth (standard) from a one with a shallower depth (signal) (Fig. 2A). Stimuli were presented using a three-interval, forced-choice method. On each trial, the standard was presented in two intervals, and the signal in one, with the signal interval determined randomly. Listeners indicated which of the three intervals contained the signal by using a computer mouse to click labeled buttons on a visual display. After every trial, visual feedback was provided indicating whether the response was correct or incorrect.

The signal modulation depth (peak-to-valley difference in decibels) was adjusted adaptively across trials to estimate the modulation-depth-discrimination threshold. Modulation-depth adjustment followed a three-down/one-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 0 dB and was adjusted in steps of 6 dB until the third reversal; subsequent steps were 2 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. The difference between this average and the standard modulation depth is reported as the depth-discrimination threshold. Blocks were excluded from analysis if they contained fewer than seven total reversals (2% in total) or any trials that took longer than 20 s (from the first observation interval through the response, 3.4% in total).

For the modulation-detection condition, listeners had to distinguish an unmodulated noise stimulus (standard) from a modulated one (signal) (Fig. 2*B*), and the modulation depth of the signal was varied adaptively to estimate the detection threshold. The procedure was the same as for depth-discrimination except that the initial signal modulation depth was 20 dB. The average of the signal modulation depths at the included reversals is reported as the detection threshold.

During the pretests and posttests, listeners completed five threshold estimates (300 trials) for each modulation condition. The order of the conditions was randomized across listeners, but held constant between the pretests and posttests for each individual listener. During the training phase, trained listeners completed 12 threshold estimates (720 trials) for the single trained condition on each session. The pretest and first day of training were conducted on consecutive days, as were the final day of training and the posttest. The training sessions occurred on most weekdays. The pretests and posttests were separated by an average of 20.3 d for the trained listeners and 22.3 d for the controls.

Stimulus synthesis and presentation. The protocol for stimulus generation was adapted from previous reports (Kowalski et al., 1996) and primarily used code from the neural systems laboratory toolbox. The spectrotemporal envelope (S) of each stimulus was a two-dimensional sinusoid defined across both time (t) and frequency (x ; number of octaves above the fundamental frequency) as follows: $S(t,x) = 1 + A \sin(2\pi wt + 2\pi \Omega x + \Phi)$, where A is the linear modulation depth, w is the temporal modulation (in Hz), Ω is the spectral modulation (in cycles/octave), and Φ is the starting phase of the sinusoid. The fundamental frequency was 100 Hz and the components were spaced in steps of 2.5 Hz up to 16,314 Hz. The starting phase was randomized. To generate the modulated noise, the spectrotemporal envelope was multiplied by random numbers drawn from a Gaussian distribution. The modulated noise was passed through a Butterworth filter with cutoff frequencies of 400 and 3200 Hz and a slope of -32 dB/octave. The duration of each stimulus was 125 ms including 10 ms raised cosine on/off ramps. Due to computational constraints, the stimuli were synthesized before testing. To introduce some randomization, we precomputed 25 instances of each combination of spectral modulation frequency, temporal modulation frequency, and modulation depth. On a given presentation, one of those 25 stimuli was randomly selected and played to the listener. The presentation level on each observation interval was roved ± 8 dB around a spectrum level of 35 dB SPL. The level and starting-phase randomizations were used to reduce the availability of local intensity cues in the spectral-modulation condition, but were used in all conditions for uniformity.

All stimuli were presented using custom software written in MATLAB and played through a 16-bit digital-to-analog converter (DD1) followed by an anti-aliasing filter with a 16 kHz cutoff frequency (FT6-2), a programmable attenuator (PA4), a sound mixer (SM3), and a headphone driver (HB6) (all from Tucker-Davis Technologies). The sounds were

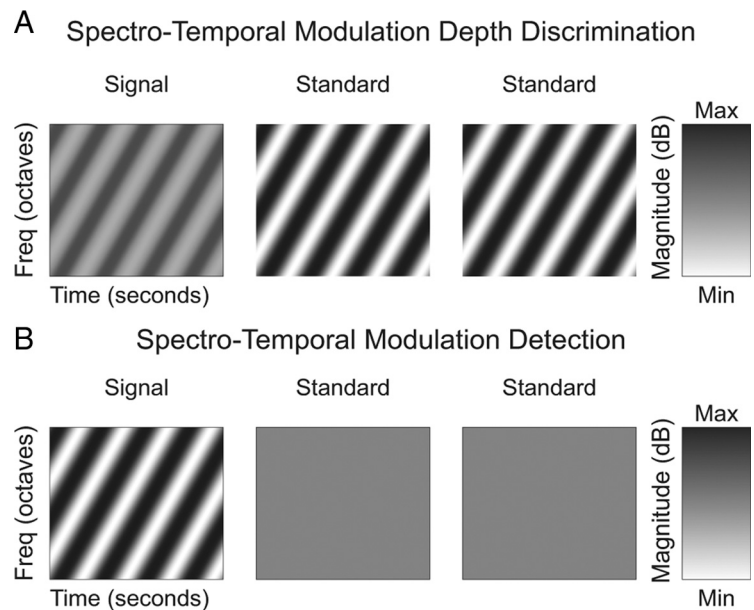


Figure 2. Tasks. Three sounds were presented on each trial. The listener's task was to choose the sound (signal) that was different from the other two (standards). **A**, In the modulation-depth-discrimination tasks (spectrotemporal shown), the signal had a shallower modulation depth (left) than the standards (middle and right). The more similar the modulation depths that could be discriminated, the better the threshold. **B**, In the modulation-detection task, the signal was modulated (left) and the standards were not (middle and right). The shallower the signal modulation depth that could be detected, the better the threshold.

presented through the left earpiece of Sennheiser HD265 circumaural headphones. Listeners were tested in a sound-attenuated room.

Listeners. Thirty-two participants (26 female) between 18 and 36 years of age served as listeners. All had normal hearing sensitivity in the left ear [≤ 20 dB HL from 250 to 8000 Hz (ANSI, 2004)] and no previous experience with psychoacoustic tasks. All listeners gave informed consent and were financially compensated for their participation. All procedures were approved by the Institutional Review Board at Northwestern University. The data from listeners whose pretest thresholds were greater than 2 SDs above the mean of all listeners on a particular condition were removed from the analyses of that condition (3.1% of the entire dataset).

Results

Learning on the trained conditions

Multiple sessions of depth-discrimination training aided performance on spectral- (Fig. 3*A*, triangles), temporal- (Fig. 3*B*, diamonds), and spectrotemporal- (Fig. 3*C*, squares) modulation. We evaluated the influence of each training regimen by computing, separately for each trained condition, a group (trained vs control) \times session (pre vs post) ANOVA, using time as a repeated measure. In each case, there was a significant group by session interaction (all $p < 0.03$). The trained listeners improved between the pretests and posttests (paired t test, all $p < 0.004$; Fig. 3*A–C*, filled symbols) while the controls did not (all $p > 0.26$; Fig. 3*A–C*, open symbols).

Each trained group improved over the course of the experiment, though the temporal-modulation-trained listeners reached asymptote before the other two trained groups. For each trained group, this improvement was confirmed by 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 (all $p < 0.003$). Further, for each individual, we computed the slope of a regression line fitted to each threshold estimate over the \log_{10} of the session number (Fig. 3*D–F*, left). The population of slopes for each trained group was significantly negative (all $p < 0.007$), again indicating improvement in each group. However, when the same set of analyses were repeated with the pretest removed (Fig. 3*D–F*, right), the

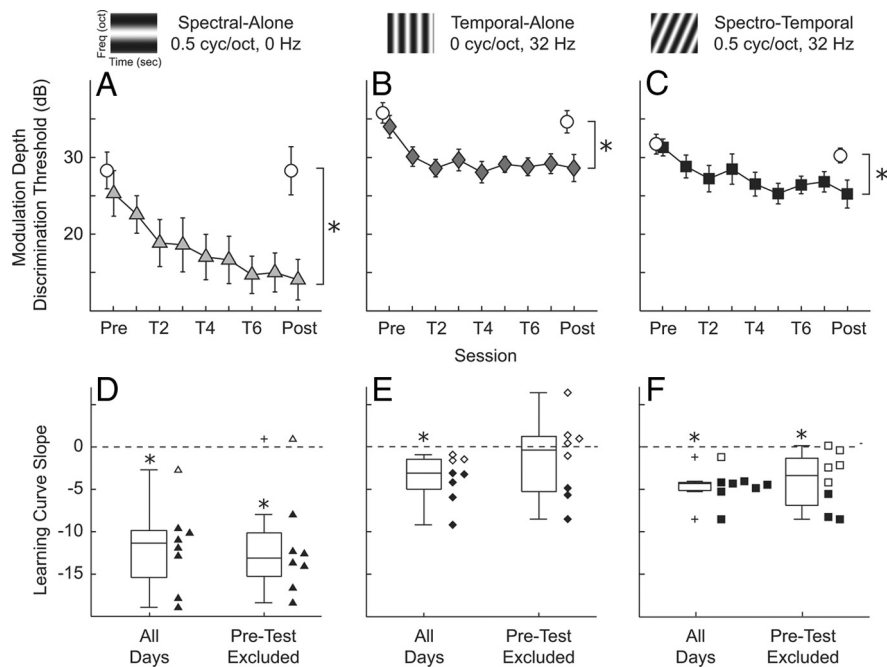


Figure 3. Performance on the trained depth-discrimination conditions. **A–C**, The group average modulation-depth-discrimination thresholds (79% correct performance) as a function of testing session for listeners trained on spectral (**A**, triangles; $n = 8$), temporal (**B**, diamonds; $n = 8$), or spectrotemporal (**C**, squares; $n = 8$) modulation. Results are also shown for controls who received no training (circles; $n = 8$). Spectrograms of each trained stimulus are depicted at the top of each column. Error bars indicate ± 1 SEM. Asterisks indicate a significant interaction ($p < 0.05$) of a group (trained vs control) \times time (pre vs post) ANOVA using time as a repeated measure. **D–F**, The slopes of individual regression lines fitted to all threshold estimates versus the \log_{10} of the session number for each listener in the spectral- (**D**), temporal- (**E**), and spectrotemporal- (**F**) modulation trained groups. Filled symbols indicate that the slope for that listener was significant and negative. The box plots 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 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 pretest (right in each panel). Asterisks indicate that the population of slopes was significantly less than zero ($p < 0.05$) according to a one-sample t test. Training led to improvement for all three trained modulations, but with different time courses.

spectral- and spectrotemporal-modulation-trained listeners still showed improvement (all $p < 0.04$), while the temporal-modulation-trained listeners did not (all $p > 0.42$). These analyses suggest that participating in the pretest brought the temporal-modulation-trained listeners to asymptotic performance. Interestingly, the controls, who also participated in the same pretest, showed no improvement, indicating that, by the posttest, the controls lost any improvements resulting from exposure to the pretest, while the practice sessions served to maintain those improvements in the temporal-modulation-trained listeners.

The relative contributions of within- and between-session improvements that occurred during the training phase differed among the trained groups. To examine when on a given day training-phase improvements emerged, we compared within- to between-session improvements for the two groups that showed improvement during this phase (the spectral- and spectrotemporal-modulation-trained listeners). For each listener, we computed the means of the first three and last three threshold estimates for each training session, and used those values to calculate performance changes within sessions (first 3 minus last 3 estimates) as well as between sessions (last 3 estimates of current session minus first three of the subsequent session). The relative magnitudes of these changes were evaluated separately for each trained group using a change (within-session vs between-session) \times session (all training days) ANOVA with both change and session as repeated measures. For spectral-modulation-trained listeners, there was a main effect of change ($F_{(1,42)} = 11.3$;

$p = 0.012$), which arose because the listeners tended to get worse within sessions ($t_{(55)} = -2.1$; $p = 0.04$) and to improve between sessions ($t_{(55)} = 3.7$; $p < 0.001$). There was no main effect of session and no change \times session interaction for this group (all $p > 0.031$). Thus, the improvements of the spectral-modulation-trained listeners emerged during the time between sessions. In contrast, for the spectrotemporal-modulation-trained listeners, there were no significant main effects or interactions (all $p > 0.25$), indicating that for these listeners, there was no marked distinction between within- and between-session improvements.

Generalization to untrained conditions

Depth discrimination

While practice led to improved depth discrimination on each trained condition, in no case did that learning generalize to an untrained depth-discrimination condition. We evaluated generalization, separately for each untrained condition, with the same criterion that we used to assess learning on the trained condition—a significant group (trained vs control) \times session (pre vs post) interaction of an ANOVA, using session as a repeated measure. By this criterion, there was no evidence of generalization. None of trained groups generalized their improvement to the trained conditions of the other groups (all $p > 0.12$; Fig. 4A–C) or to a depth-discrimination condition on which none of the groups trained—downward spectrotemporal modulation (all $p > 0.13$; Fig. 4D). Further, for each condition, the effect size (partial η^2) of the group \times session interaction was >0.43 when the condition was trained, but <0.21 when it was not (Fig. 5).

Though none of the trained groups improved more than controls on any of untrained depth-discrimination conditions, the trained groups did show some improvement that was uniform across these conditions. We further examined performance on the untrained conditions separately for each group (3 trained and 1 control) by computing a session (pre vs post) \times condition (all untrained depth-discrimination conditions) ANOVA with session as a repeated measure. For each trained group, there was a main effect of session (all $p < 0.001$) and no session \times condition interaction (all $p > 0.55$). In contrast, for the controls, neither the main effect of session nor the session \times condition interaction was significant (all $p > 0.19$). Thus, the trained listeners improved somewhat on the untrained depth-discrimination conditions, but not by amount that was distinguishable from the controls.

Detection of spectrotemporal modulation

Finally, the influence of practicing modulation depth-discrimination on the ability to detect upward spectrotemporal modulation (Fig. 6) differed across the three trained groups. On this detection condition, compared with controls, the spectral-modulation-trained listeners learned ($F_{(1,14)} = 9.8$; $p = 0.007$), the temporal-modulation-trained listeners showed no change ($F_{(1,14)} =$

0.50, $p = 0.49$), and the spectrotemporal-modulation-trained listeners got worse ($F_{(1,13)} = 4.8$; $p = 0.048$). The decrements in performance on detection in the spectrotemporal-modulation-trained listeners were correlated with improvements in depth-discrimination for the same stimulus (their trained condition; $r = -0.80$, $p = 0.03$; Fig. 6E). There were no such tradeoffs between detection and discrimination of this stimulus in the remaining three groups (all $p > 0.13$; Fig. 6B–D) or between the detection of this stimulus and the discrimination of the respective trained stimuli of the other two trained groups (all $p > 0.47$).

Discussion

Combined spectrotemporal processing

The present results provide behavioral evidence that stimuli can be processed by a mechanism that is selective for particular combinations of spectral and temporal modulation frequencies. We trained listeners to discriminate the depth of spectral, temporal, or upward spectrotemporal modulation and tested the generalization of their learning to the trained conditions of the other groups as well as to the depth discrimination of downward spectrotemporal modulation. Three influences of training were dependent on the spectrotemporal characteristics of the trained stimulus. First, and most importantly, there was no generalization of training-induced improvement to the untrained depth-discrimination conditions. Learning on isolated spectral (0.5 cyc/oct; 0 Hz) or temporal (0 cyc/oct; 32 Hz) modulation did not generalize to either direction of spectrotemporal modulation (0.5 cyc/oct; ± 32 Hz), even though the spectrotemporal conditions shared one modulation frequency with each of the isolated modulations (Fig. 3C,D). There was also no generalization from spectrotemporal modulation to the isolated modulations or to the opposite direction of spectrotemporal modulation (Figs. 4A,B; 5D).

Second, there was a unique relationship between detection and discrimination when both tasks were assessed using stimuli that had the same combination of spectral and temporal modulation frequencies. Learning on depth discrimination of spectrotemporal modulation led to worsening on (negative generalization to) the detection of that spectrotemporal modulation, and the individual magnitudes of these changes were significantly correlated. This relationship was specific to the combined spectrotemporal modulation. Learning on spectral modulation depth discrimination led to improvements, rather than worsening, on the detection of spectrotemporal modulation. This generalization implies a connection between these two conditions, but one that differs from the connection between the two spectrotemporal conditions. Learning on temporal modulation depth discrimination did not influence performance on the detection condition, implying separate processing of these two conditions.

Depth Discrimination

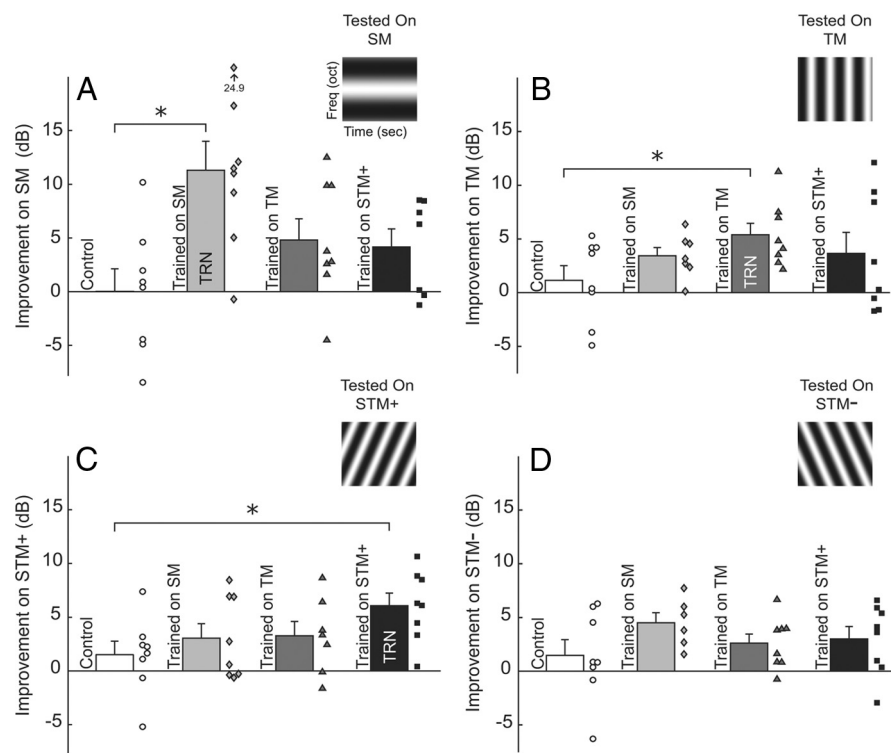


Figure 4. Performance on all of the modulation-depth-discrimination conditions. The difference in threshold between the pretest and posttest (pre minus post) for each group (bars) and listener (symbols) on the depth-discrimination conditions. **A–D**, Results are shown for spectral modulation (SM; **A**) and temporal modulation (TM; **B**), as well as for upward (**C**) and downward (**D**) spectrotemporal modulation (STM+ and STM–, respectively). For each modulation, the magnitude of improvement is shown for the listeners, if any, who were trained on that modulation (TRN), for those who were trained on other modulations, and for those who received no training. Spectrograms of the tested stimuli are displayed near the top right corner of each panel. Error bars indicate ± 1 SEM. Asterisks indicate a significant interaction ($p < 0.05$) of a group (trained vs control) \times time (pre vs post) repeated-measures ANOVA. The learning on the trained conditions did not generalize to any untrained depth-discrimination conditions.

Third, the time courses of improvement on modulation depth discrimination differed for different combinations of spectral and temporal modulation frequencies. Asymptotic performance was reached by the first training day for temporal modulation, but later in the training phase for spectral and spectrotemporal modulation. Further, the more gradual improvements occurred primarily between sessions for spectral modulation, but not for spectrotemporal modulation.

It is unlikely that these outcomes would have occurred if training had modified an aspect of processing in which the spectral and temporal components were represented separately. Consider, for example, the lack of generalization to untrained depth-discrimination conditions. If training on spectrotemporal modulation had modified separate processing of spectral and/or temporal modulation, learning would have generalized to spectral-alone and/or temporal-alone modulation. Similarly, if training on these isolated modulations had modified such separate processing, learning would have generalized, at least in part, to spectrotemporal modulation. Thus, though separable processing of spectral-alone or temporal-alone modulation may occur somewhere along the auditory pathway, it appears that the training-induced improvements observed here resulted from modifications to the processing of combined spectrotemporal modulation.

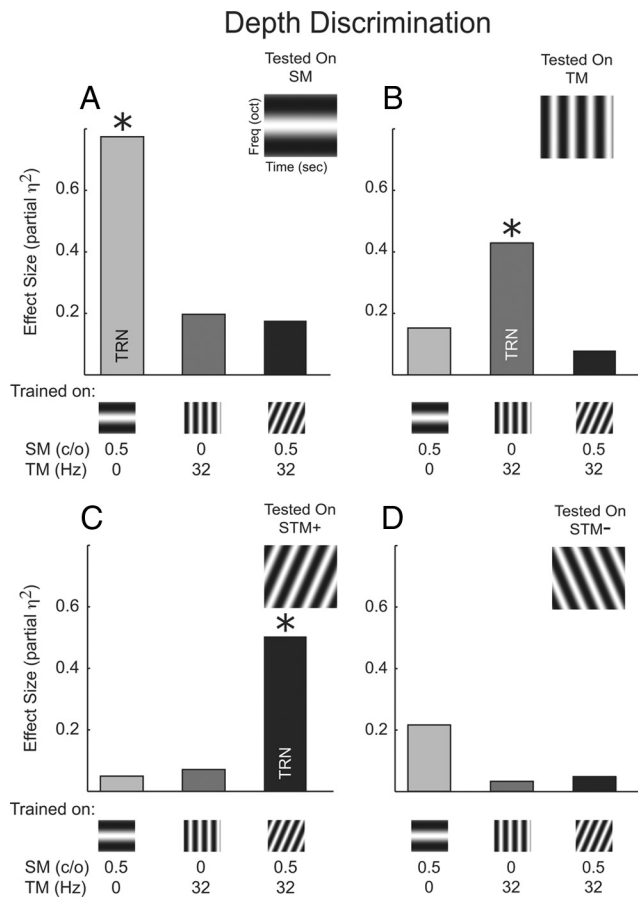


Figure 5. Effect sizes for all trained versus control comparisons on the modulation-depth-discrimination conditions. Partial η^2 effect sizes of the group (trained vs control) \times session (pre vs post) interaction in ANOVA using time as a repeated measure. **A–D**, Effect sizes are shown for spectral modulation (SM; **A**) and temporal modulation (TM; **B**), as well as for upward (**C**) and downward (**D**) spectrotemporal modulation (STM+ and STM–, respectively). For each modulation, the effect size is shown for the listeners who were trained on that modulation (TRN) as well as for those who were trained on other modulations. Asterisks indicate that the interaction was significant ($p < 0.05$). For each depth-discrimination condition, the effect size was at least twice as large when it was trained than when it was not.

The present behavioral support for combined spectrotemporal processing is reminiscent of the combined spectrotemporal tuning observed in numerous physiological investigations. As mentioned in the Introduction, individual neurons that are tuned to particular combinations of spectral and temporal modulation frequencies have been observed at cortical (Kowalski et al., 1996) and subcortical (Versnel et al., 2009) levels in animals, and similar tuning has been observed in regions of human primary and secondary auditory cortex (Langers et al., 2003; Schönwiesner and Zatorre, 2009). Thus, there are many candidate neural populations that could have been involved in the learning. However, the specificity of learning to the trained direction of spectrotemporal modulation helps narrow the possibilities. Neurons that preferentially respond to one direction of spectrotemporal modulation are more common in higher than lower areas of auditory cortex (Loftus and Sutter, 2001). Therefore, it is more likely that the training-induced modification involved higher rather than lower areas of auditory cortex. The training could have led to a refinement in these modulation-tuned neurons themselves or in the weight-

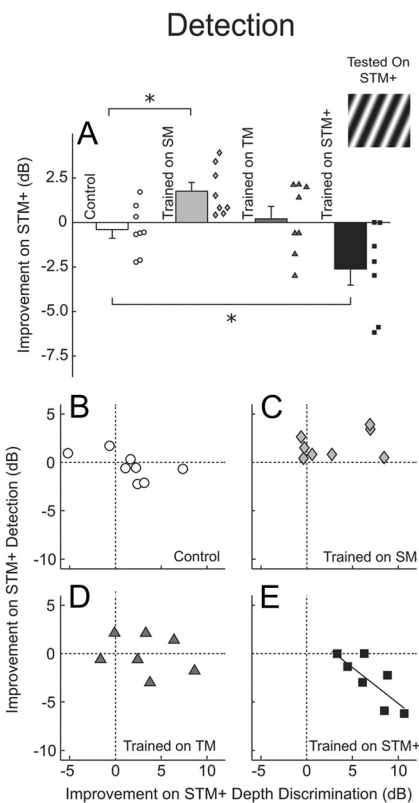


Figure 6. Performance on the untrained spectrotemporal modulation detection condition. **A**, As in Figure 3, but for the detection (rather than discrimination) of upward spectrotemporal modulation. **B–E**, The difference in threshold between the pretest and posttest (pre minus post) for discrimination (abscissa) and detection (ordinate) of upward spectrotemporal modulation. Results are plotted separately for listeners trained to discriminate the depth of spectral (**B**), temporal (**C**), or spectrotemporal (**D**) modulation as well as for controls (**E**). Discrimination training led to worsening on detection but only when both tasks used the same modulation.

ing assigned to inputs from these neurons by a decision maker (for more general discussions of these alternatives, see Doshier and Lu, 1998; Ahissar and Hochstein, 2004).

Despite the specificity of learning to the trained condition, each trained group showed a modest improvement across all of the untrained depth-discrimination conditions (Fig. 4). Thus, the trained groups appear to have learned some general aspect of modulation depth discrimination. One possibility is that this uniform improvement came from an increased ability to ignore the randomization of presentation level. Since the same level randomization was used on all conditions, an increased ability to ignore this randomization would lead to improvement on all untrained conditions. Another possibility is that the listening strategy for the modulation itself changed over the course of training. For example, listeners might have initially chosen the interval that was most correlated to an internal template of an unmodulated noise, and then later chosen the interval that was least correlated to a template of the trained standard. Refinements of the unmodulated template with practice would lead to modest uniform improvement across all depth-discrimination conditions, while refinements of the template of the trained standard would lead to improvements that were specific to the trained condition. It is also worth noting that the magnitude of the improvement between the pretests and posttests on the untrained

conditions was comparable to that between the pretest and the first day of training on the trained conditions. This indicates that the pretest itself was sufficient to generate some learning on untrained conditions. It is possible that further training simply maintained these improvements in the trained groups, and that these improvements declined in the controls due to the lack of training. For example, each training session may have aided in the process of making the pretest-induced learning more permanent (consolidation). Alternatively, the learning arising from the pretest itself may have lasted only into the next day, with each subsequent day of training—including the final day—inducing similarly temporary learning.

Detection versus discrimination

The current results also replicate and extend existing literature indicating that improvements in modulation discrimination can come at the cost of worsening on detection. Training on depth discrimination of spectrotemporal modulation led to learning on that condition, but worsening on the detection of that same modulation (Fig. 5E). A related tradeoff has been observed in auditory temporal modulation perception where listeners who practiced discriminating modulation rate improved on that task, but got worse at detecting the same stimulus (Fitzgerald and Wright, 2005) [though detection training did not influence discrimination (Fitzgerald and Wright, 2011)]. Similarly, in the visual system, adaptation to a sinusoidal grating led to better orientation discrimination but poorer detection (Regan and Beverley, 1985). The present data extend this previous work by demonstrating that the discrimination/detection tradeoff can be specific to the trained stimulus. This tradeoff did not occur when depth-discrimination training and detection testing used different modulations (Fig. 5C,D). Training to discriminate spectral modulation actually led to improvements in detection of spectrotemporal modulation.

There are several potential listening strategies, which if adopted over the course of training, could have led to a discrimination/detection tradeoff that is specific to the trained modulation. For example, trained listeners could have learned to focus on the processing of the trained modulation at deep modulation depths and to ignore that of shallow depths. Adoption of this listening strategy would improve depth discrimination because this task requires listeners to distinguish deep modulation depths (signal) from deeper ones (standard), but worse detection because this task depends upon the ability to distinguish shallow modulation depths (signal) from no modulation at all (standard). Alternatively, training could have led listeners to adopt a strategy to choose the interval that is least correlated to an internal template of the trained standard (as suggested above). This strategy would aid depth discrimination because the signal interval in this task (the shallower modulation depth) has the weakest correlation to that template, but hurt detection because the signal interval in this task (the modulated stimulus) has the strongest correlation to that template.

Summary and conclusion

A fundamental function of any sensory system is to extract information from the patterns of activity distributed across the sensory periphery. The auditory system must analyze the complex patterns of time-varying cochlear activity that are evoked by natural sounds such as speech. The current data demonstrate that perceptual learning can be specific to particular combinations of spectral and temporal modulation fre-

quency, and thus provide some of the first direct behavioral evidence that a mechanism with such combined tuning can influence perception. These data are consistent with the proposal that auditory perception is mediated a bank of modulation filters that are tuned to particular combinations of spectral and temporal modulation frequency, and further suggest that this proposal should allow the contributions of each filter to be influenced separately by previous experience. They also suggest that experience-induced physiological modifications that are restricted to neurons tuned to particular combined spectrotemporal modulations could be observed. Finally, these results imply that training on spectrotemporal modulation might improve performance on real-world perceptual tasks, but only if the training involves the particular combination of spectral and temporal modulation frequencies that limits performance of those tasks.

References

- Ahissar M, Hochstein S (2004) The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci* 8:457–464.
- ANSI (2004) American national standards specification for audiometers. New York: American National Standards Institute.
- Bacon SP, Grantham DW (1989) Modulation masking: effects of modulation frequency, depth, and phase. *J Acoust Soc Am* 85:2575–2580.
- Bregman A (1990) Auditory scene analysis: the perceptual organization of sound. Cambridge, MA: MIT.
- Carlyon RP, Shamma S (2003) An account of monaural phase sensitivity. *J Acoust Soc Am* 114:333–348.
- Chi T, Gao Y, Guyton MC, Ru P, Shamma S (1999) Spectro-temporal modulation transfer functions and speech intelligibility. *J Acoust Soc Am* 106:2719–2732.
- Chi T, Ru P, Shamma SA (2005) Multiresolution spectrotemporal analysis of complex sounds. *J Acoust Soc Am* 118:887–906.
- Depireux DA, Simon JZ, Klein DJ, Shamma SA (2001) Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophysiol* 85:1220–1234.
- Doshier BA, Lu ZL (1998) Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proc Natl Acad Sci U S A* 95:13988–13993.
- Elhilali M, Chi T, Shamma S (2003) A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. *Speech Communication* 41:331–348.
- Fitzgerald MB, Wright BA (2005) A perceptual learning investigation of the pitch elicited by amplitude-modulated noise. *J Acoust Soc Am* 118:3794–3803.
- Fitzgerald MB, Wright BA (2011) Perceptual learning and generalization resulting from training on an auditory amplitude-modulation detection task. *J Acoust Soc Am* 129:898–906.
- Huang CB, Zhou Y, Lu ZL (2008) Broad bandwidth of perceptual learning in the visual system of adults with anisometric amblyopia. *Proc Natl Acad Sci U S A* 105:4068–4073.
- Kowalski N, Depireux DA, Shamma SA (1996) Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *J Neurophysiol* 76:3503–3523.
- Langers DR, Backes WH, van Dijk P (2003) Spectrotemporal features of the auditory cortex: the activation in response to dynamic ripples. *Neuroimage* 20:265–275.
- Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J Neurophysiol* 60:1799–1822.
- Levitt H (1971) Transformed up-down methods in psychoacoustics. *J Acoust Soc Am* 49 [Suppl 2]:467+.
- Loftus WC, Sutter ML (2001) Spectrotemporal organization of excitatory and inhibitory receptive fields of cat posterior auditory field neurons. *J Neurophysiol* 86:475–491.
- Miller LM, Escabi MA, Read HL, Schreiner CE (2002) Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J Neurophysiol* 87:516–527.

- Regan D, Beverley KI (1985) Postadaptation orientation discrimination. *J Opt Soc Am A* 2:147–155.
- Sabin AT, Eddins DA, Wright BA (2012) Perceptual learning of auditory spectral modulation detection. *Exp Brain Res*. Advance online publication. doi:10.1007/s00221-012-3049-0.
- Saoji AA, Eddins DA (2007) Spectral modulation masking patterns reveal tuning to spectral envelope frequency. *J Acoust Soc Am* 122:1004–1013.
- Schönwiesner M, Zatorre RJ (2009) 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* 106:14611–14616.
- Shamma SA, Versnel H, Kowalski N (1995) Ripple analysis in ferret primary auditory cortex. I. Response characteristics of single units to sinusoidally rippled spectra. *Aud Neurosci* 1:233–254.
- Versnel H, Zwiers MP, van Opstal AJ (2009) Spectrotemporal response properties of inferior colliculus neurons in alert monkey. *J Neurosci* 29:9725–9739.
- Wojtczak M, Viemeister NF (2003) Suprathreshold effects of adaptation produced by amplitude modulation. *J Acoust Soc Am* 114:991–997.
- Woolley SM, Fremouw TE, Hsu A, Theunissen FE (2005) Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat Neurosci* 8:1371–1379.
- Wright BA, Dai H (1998) Detection of sinusoidal amplitude modulation at unexpected rates. *J Acoust Soc Am* 104:2991–2996.