Learning to discriminate frequency modulation rate can benefit and worsen pitch acuity

Matthew G. Wisniewski,[a] Estella H. Liu, Barbara A. Church, and Eduardo Mercado III
Department of Psychology, University at Buffalo, The State University of New York, Park Hall Room 204, Buffalo, New York 14260
mgw@buffalo.edu, hliu6@buffalo.edu, bchurch@buffalo.edu, emiii@buffalo.edu

Abstract: Participants were trained to discriminate frequency modulation rates (FM-rate training) or Gabor patch orientations (visual training) in a same–different task for two different training lengths. Test discriminations involved trains of FM sweeps with identical modulation rates, but different frequencies. FM-rate training enhanced test accuracy (relative to visual) when sweep trains contained frequencies similar to training. For extended FM-rate training, the opposite was true for trains shifted one octave higher. In contrast to previous work, generalization of learning to the untrained dimension (pitch) was not well accounted for by conceptual learning. Mechanisms of stimulus learning may better explain the current cross-dimensional generalization.

PACS numbers: 43.66.Fe, 43.66.Ba [QJF]
Date Received: October 10, 2013 **Date Accepted:** January 9, 2014

1. Introduction

Auditory training can modify the ability to perceive acoustic differences and generalizes to untrained sounds (Wright and Zhang, 2009). Generalization is most often studied by testing untrained conditions in which critical differences between stimuli are the same as during training. For instance, from learning a 100 ms interval marked by 1 kHz tones to the same interval marked by 3.75 kHz tones (Karmarkar and Buonomano, 2003). There is considerably less work examining generalization of learning to discriminations involving training-irrelevant dimensions (e.g., to pitch discrimination after training on temporal distinctions).

Past results on this cross-dimensional generalization (sometimes called task generalization) are mixed. Some report no generalization to untrained dimensions [e.g., from amplitude modulation (AM) rate discrimination to pitch discrimination; van Wassenhove and Nagarajan, 2007], while others show either beneficial or harmful consequences of training. For example, learning to discriminate inter-aural level differences (ILDs) can enhance discrimination of inter-aural time differences (ITDs; Ortiz and Wright, 2010). In contrast, AM rate discrimination training can harm detection of AM sounds from unmodulated sounds (Fitzgerald and Wright, 2005; also see Sabin *et al.*, 2012). Some of these cross-study differences can perhaps be explained by conceptual learning. ILD training may benefit ITD discrimination because both involve attending to location (Ortiz and Wright, 2010), while null effects may occur when such similarities between training and test tasks are not present (van Wassenhove and Nagarajan, 2007). Also, degradation of AM detection after AM-rate training could result if learners adopt a listening strategy that works well for rate discrimination, but poorly for detection (Fitzgerald and Wright, 2005; Sabin *et al.*, 2012).

Stimulus learning aspects of perceptual learning (i.e., learning associated with feature values of trained stimuli; Ortiz and Wright, 2010) could also lead to cross-dimensional

[a]Author to whom correspondence should be addressed.

generalization. Studies focused on stimulus learning suggest that reweighting of connections from sensory representations to higher-level decisional areas (e.g., Petrov *et al.*, 2005), and/or modifications to sensory representations themselves (e.g., Mercado *et al.*, 2001) cause changes in acuity. If sounds used to test generalization to an untrained dimension share features with trained sounds, then benefits could arise when processing of those features has been enhanced at a decisional and/or representational level. If features are not shared, possibly because different neural populations are engaged, learning could decrease acuity. This may happen if learning "tunes out" neurons containing discriminative information along a training-irrelevant dimension in weighted connections (for a similar effect in vision, see Petrov *et al.*, 2005), or if enhancing representations of trained sounds comes at the cost of other representations (e.g., by shifts in the characteristic features of receptive fields; Mercado *et al.*, 2001; Weinberger, 2007).

The current study examines possible stimulus learning effects in cross-dimensional generalization. Listeners were trained to distinguish frequency-modulated sweep trains varying in repetition rate (FM-rate training) or Gabor patches differing in orientation (visual training). In generalization tests conducted before and after training, the relevant discrimination dimension was pitch (i.e., the frequencies spanned by each sweep) rather than rate. Frequencies within test sweep trains either overlapped those heard during training (~0.5–1 kHz) or were an octave higher (~1–2 kHz). If conceptual learning drives generalization, then FM-rate training should produce similar effects as visual training, or worsening for both pitch ranges if listeners adopt a non-optimal listening strategy (Fitzgerald and Wright, 2005). In contrast, stimulus learning may benefit acuity in the trained pitch range due to the similarity of representations utilized during training and testing. Acuity for untrained pitch ranges should be degraded or unchanged given more disparate stimulus-activated neural populations during training and test.

2. Methods

2.1 Participants

Thirty-two normal-hearing students at the University at Buffalo, The State University of New York, participated for course credit.

2.2 Design, stimuli, and apparatus

A mixed 2 (training type) × 2 (training length) × 2 (pitch range) × 2 (test) design was used. Training type (FM-rate or visual) and training length (limited or extended) were the between-participants factors. Levels of pitch range refer to whether or not the frequencies contained in sweeps during tests were similar (trained range) or dissimilar to those experienced during training (untrained range). Levels of test were pre- and post-test.

FM sweep trains (described in Sec. 2.3) varied in repetition rate or frequencies within modulation. Visual stimuli were Gabor patches of different orientations (45° and 49°). Stimuli were generated in MATLAB (Mathworks, Inc., Natick, MA). Experimental procedures were executed using DMDX (Forster and Forster, 2003) and sounds were heard over Audio-Technica ATH-m40fs headphones (Audio-Technica, Stow, OH) at ~65 dB sound pressure level.

2.3 Procedures

Participants indicated whether consecutively presented stimuli were the same or different using marked computer keys. Silence lasting 500 ms and a blank screen separated intervals. For "same" trials of FM-rate training, 12 Hz sweep trains with upward sweeps spanning 0.5–1 kHz were presented in both intervals. On "different" trials one interval contained a 10.4 Hz repetition rate train containing the same frequencies [Fig. 1(A)]. Participants trained with Gabor patches performed an analogous task [Fig. 1(B)]. "same" and "different" trials occurred with equal probability. There were 24 trials per training block (limited = 2 blocks, extended = 16 blocks). Trial order was completely randomized within a block and feedback was given after responding.

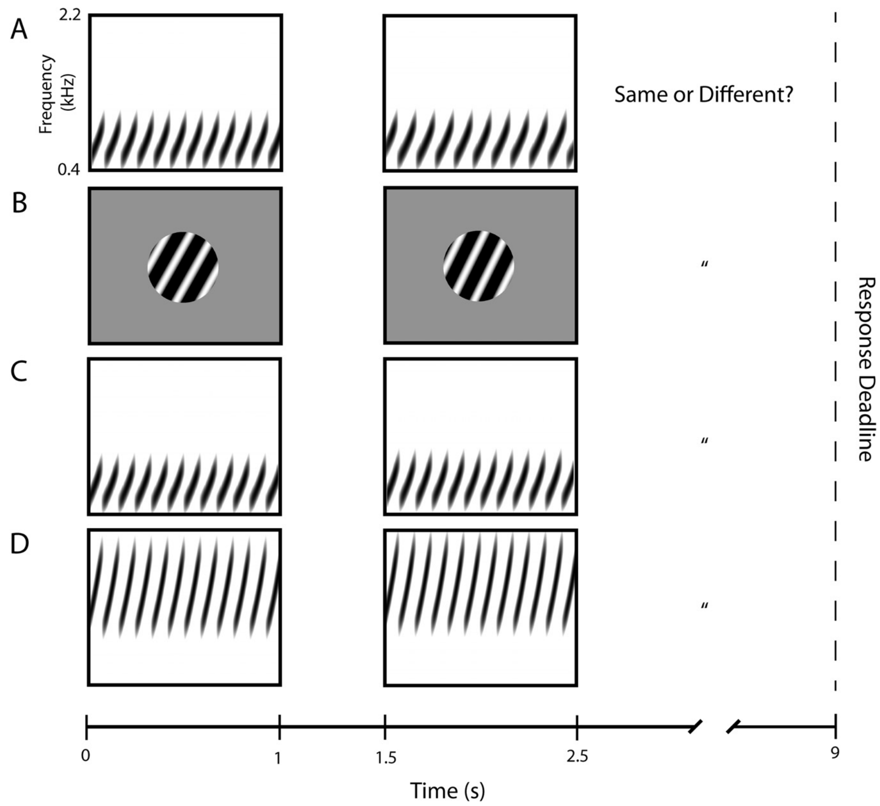

Fig. 1. Depiction of example "different" trials for (A) FM-rate training; (B) visual training; (C) the trained pitch range phase of pre- and post-tests; and (D) the untrained pitch range phase of pre- and post-tests. In each panel, the "standard" stimulus (i.e., the stimulus presented in both intervals on "same" trials) is shown first, followed by one example of a stimulus that could have been presented on "different" trials (see text for descriptions of all possible stimuli). In the actual experiment, order was balanced across intervals. Sounds are depicted as spectrograms with frequency plotted on a linear scale.

Pre- and post-tests had two phases, both of which employed the same–different task described above. All sounds presented in tests had 12 Hz repetition rates. In the trained pitch range phase, intervals of "same" trials were identical to training. One interval during "different" trials contained sweeps spanning 0.51–1.02 kHz, 0.52–1.04 kHz, or 0.53–1.06 kHz [Fig. 1(C)]. In the untrained pitch range phase, "same" trials contained trains with sweeps spanning 1–2 kHz. "Different" trials contained one interval with sweeps spanning frequencies from 1.02–2.04 kHz, 1.04–2.08 kHz, or 1.06–2.12 kHz [Fig. 1(D)]. There were 24 randomized trials per phase (48 trials per test) in which no feedback was given. Half of the participants completed the trained pitch range phase first in tests. The order was reversed for the other half. Participants were told that the overall pitch of stimuli would change between phases. For the limited-training condition, all training and testing was completed in one session. For extended training, the pre-test and half of training occurred on the first day, and the second half of training and the post-test were completed the following day.

3. Results

Percent correct was used as an accuracy measure. For each group, training led to learning [Fig. 2(A)], confirmed by higher last block than first block accuracy, $p < 0.05$ (planned comparison single sided t-tests). Figure 2(B) shows accuracy averaged across "same" and "different" trials for each training type and pitch range in the pre- and

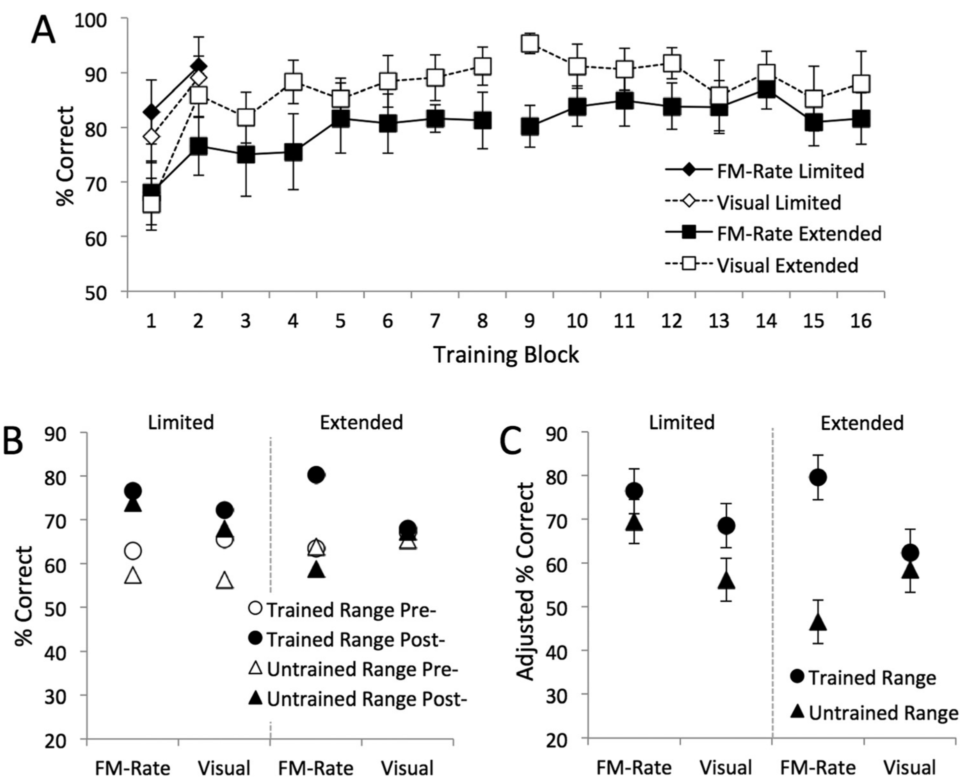

Fig. 2. (A) Mean accuracy in each training block. (B) Mean accuracy averaged across "same" and "different" trial types for each combination of the between and within participant factors. Open markers indicate pre-test and filled markers indicate post-test accuracy. Circles represent the trained pitch range and triangles represent the untrained pitch range. (C) Mean accuracy [adjusted for post-test $\ln(\beta)$ and mean pre-test accuracy] for "different" trials in the post-test. All error bars show standard error of the mean. Error bars are omitted from (B) for clarity.

post-tests. Of critical interest here, however, is performance on "different" trials. "Same" trials for the trained pitch range were identical to those experienced during training, decreasing their significance in measuring generalization. Therefore, statistical analyses focused solely on "different" trial accuracy.[1]

Post-test $\ln(\beta)$, a response bias signal detection parameter (here >0 if "same" biased and <0 if "different" biased), was negatively correlated with post-test accuracy, $r(30) = -0.72$, $p < 0.001$. A mixed model 2 (training type) $\times$ 2 (training length) $\times$ 2 (pitch range) analysis of covariance on post-test accuracy therefore included post-test $\ln(\beta)$ as a covariate. Pre-test accuracy was also used as a covariate because it correlated with post-test accuracy, $r(30) = 0.62$, $p < 0.001$ (Liu *et al.*, 2008). Accuracies (adjusted for the covariates) for "different" trials in the post-test are shown in Fig. 2(C).

Marginal effects of training type, $F(1,26) = 3.88$, $p < 0.10$, $\eta_p{}^2 = 0.13$, and training length, $F(1,26) = 3.34$, $p < 0.10$, $\eta_p{}^2 = 0.11$, were found along with a significant three-way interaction, $F(1,26) = 4.85$, $p < 0.05$, $\eta_p{}^2 = 0.16$, suggesting differential effects of training across conditions and that the null effect hypothesis predicted by conceptual learning should be rejected.[2] All other interactions were non-significant, $F < 2.1$.

Planned contrasts were performed separately for each pitch range to determine whether the interaction was consistent with worsening predicted by conceptual learning (worse performance for FM-rate training) or bidirectional generalization predicted by stimulus learning (better or worse performance for FM-rate training depending on pitch range; see Introduction, Sec. 1). For the trained range [circles in Fig. 2(C)], the conceptual learning contrast (visual limited $\approx$ visual extended $>$ FM-rate limited $\approx$ FM-rate extended) was significant in the opposite direction as predicted, $F(1,26) = 8.05$, $p < 0.01$, ${\eta_p}^2 = 0.36$. That is, FM-rate training was more accurate than visual. The stimulus learning contrast (FM-rate extended > FM-rate limited > visual limited ≈ visual extended) was significant in the hypothesized direction, $F(1,26) = 7.75$, $p < 0.01$, ${\eta_p}^2 = 0.35$, supporting greater accuracy for FM-rate training in the trained pitch range.[3]

Neither the hypothesized conceptual (visual limited ≈ visual extended > FM-rate limited ≈ FM-rate extended), or stimulus learning contrast (visual limited ≈ visual extended > FM-rate limited > FM-rate extended), significantly fit the data in the untrained range [triangles in Fig. 2(C)], $F < 2$. This may reflect the fact that the limited FM-rate group actually outperformed the visual training groups, while the extended FM-rate group performed worse than visual groups. A *post hoc* contrast better fit to this pattern (FM-rate limited > visual extended > visual limited > FM-extended) suggests that this was the case, $F(1,26) = 10.11$, $p < 0.01$, ${\eta_p}^2 = 0.47$.

4. Discussion

If cross-dimensional generalization was driven by attention to repetition rate or learned rate-specific listening strategies, then we would have seen an absent or solely negative effect of learning on pitch discrimination. Neither event occurred. Note that conceptual learning associated with response demands and/or procedure is also an unlikely account of the data given the similarity of FM-rate and visual training tasks in these aspects. For stimulus learning, we predicted benefits for the trained pitch range, but harmful effects for the untrained range. This hypothesis was consistent with trained range performance, but not with the untrained range where acuity was benefited before it was degraded.

One possible explanation for this result is that stimulus representations overlapped largely early in training, perhaps in neurons responding to specific spectrotemporal features (e.g., upward FM; see Sabin *et al.*, 2012), but this overlap decreased with extended training (e.g., Blake *et al.*, 2002). Alternatively, learning could have involved a switch from the utilization of high-level representations tuned to a wide range of frequencies to lower-level representations with finer frequency resolution (Ahissar *et al.*, 2009). In either situation, reweighting might benefit performance initially in the untrained range, but then degrade it when the informative neural populations are no longer shared between tasks. Currently, we can only speculate as to specific mechanisms, as our design was not meant to parse these theories.

In Ortiz and Wright's (2010) cross-dimensional generalization study, benefits from ILD to ITD decreased 24 h post training. Although ILD trained listeners outperformed a naïve control group at this time point, worsening during the delay was proposed to result from consolidation of ILD stimulus learning that came at the cost of ITD performance. Training length is confounded with the possibility of overnight consolidation in the current study, leaving the possibility of a consolidation contribution to worsening seen here. Another possibility is that stimulus exposure underlies our cross-dimensional generalization (e.g., Wright *et al.*, 2010). That is, mere exposure to a 12 Hz (0.5–1 kHz) train, rather than rate training, might enhance acuity along the pitch dimension for that sound, but degrade it for others. These consolidation and exposure-related effects are stimulus-learning explanations in that they are related to features of sounds, not learning of relevant dimensions, procedures, or response demands. Whether explicit rate training, consolidation, or exposure led to the generalization patterns seen here are key questions for future behavioral and theoretical investigations. Nevertheless, the current work demonstrates the utility of studying cross-dimensional generalization from a stimulus learning perspective, and suggests new research directions for further developing perceptual learning theory.

Acknowledgments

This work was supported by NSF Grant No. SBE 0542013 to the Temporal Dynamics of Learning Center. The authors thank P. Q. Pfordresher, J. Sawusch, and M. Dent for comments. We also thank J. Boomer for providing Gabor patch stimuli.

References and links

[1]An analysis of "same" trials yielded no significant effects, $F < 2$.

[2]Removing covariates in alternative analyses did not alter mean patterns. The three-way interaction is also significant if covariates are omitted, $F(1,28) = 5.59$, $p < 0.05$, $\eta_p^2 = 0.17$. There were no significant differences in either of the covariates between groups at either pitch range, $F < 2$. Note that d' controls for bias, but necessitates using data from both trial types. If only the performance on 'Same' trials is modified by learning (not generalization) d' will change, making it an inappropriate dependent measure.

[3]A significant amount of unaccounted variability existed for the trained range stimulus learning contrast, $F(2,26) = 9.16$, $p < 0.001$, likely reflecting unpredicted differences in visual groups and a small accuracy increase from limited to extended FM-rate training.

Ahissar, M., Nahum, M., Nelken, I., and Hochstein, S. (**2009**). "Reverse hierarchies and sensory learning," Philos. Trans. R. Soc. London, Ser. B **364**, 285–299.

Blake, D. T., Strata, F., Churchland, A. K., and Merzenich, M. M. (**2002**). "Neural correlates of instrumental learning in primary auditory cortex," Proc. Nat. Acad. Sci. U.S.A. **99**, 10114–10119.

Fitzgerald, M. B., and Wright, B. A. (**2005**). "A perceptual learning investigation of the pitch elicited by amplitude-modulated noise," J. Acoust. Soc. Am. **118**, 3794–3803.

Forster, K. I., and Forster, J. C. (**2003**). "DMDX: A windows display program with millisecond accuracy," Behav. Res. Methods Instrum. Comput. **35**, 116–124.

Karmarkar, U. R., and Buonomano, D. V. (**2003**). "Temporal specificity of perceptual learning in an auditory discrimination task," Learn. Mem. **10**, 141–147.

Liu, E. H., Mercado, E., III, Church, B. A., and Orduña, I. (**2008**). "The easy-to-hard effect in human (*Homo sapiens*) and rat (*Rattus norvegicus*) auditory identification," J. Comp. Psychol. **122**, 132–145.

Mercado, E., III, Myers, C. E., and Gluck, M. A. (**2001**). "A computational model of mechanisms controlling experience-dependent reorganization of representational maps in auditory cortex," Cogn. Affect. Behav. Neurosci. **1**, 37–55.

Ortiz, J. A., and Wright, B. A. (**2010**). "Differential rates of consolidation of conceptual and stimulus learning following training on an auditory skill," Exp. Brain Res. **201**, 441–451.

Petrov, A. A., Dosher, B. A., and Lu, Z.-L. (**2005**). "The dynamics of perceptual learning: An incremental reweighting model," Psychol. Rev. **112**, 715–743.

Sabin, A. T., Eddins, D. A., and Wright, B. A. (**2012**). "Perceptual learning evidence for tuning to spectrotemporal modulation in the human auditory system," J. Neurosci. **32**, 6542–6549.

van Wassenhove, V., and Nagarajan, S. S. (**2007**). "Auditory cortical plasticity in learning to discriminate modulation rate," J. Neurosci. **27**, 2663–2672.

Weinberger, N. M. (**2007**). "Auditory associative memory and representational plasticity in the primary auditory cortex," Hear. Res. **229**, 54–68.

Wright, B. A., Sabin, A. T., Zhang, Y., Marrone, N., and Fitzgerald, M. B. (**2010**). "Enhancing perceptual learning by combining practice with periods of additional sensory stimulation," J. Neurosci. **30**, 12868–12877.

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