Journal of Experimental Psychology:
Learning, Memory, and Cognition
2013, Vol. 39, No. 1, 270–276

0278-7393/12/$12.00 DOI: 10.1037/a0028647

RESEARCH REPORT

Temporal Dynamics in Auditory Perceptual Learning: Impact of Sequencing and Incidental Learning

Barbara A. Church, Eduardo Mercado III, Matthew G. Wisniewski, and Estella H. Liu
University at Buffalo, The State University of New York

Training can improve perceptual sensitivities. We examined whether the temporal dynamics and the incidental versus intentional nature of training are important. Within the context of a birdsong rate discrimination task, we examined whether the sequencing of pretesting exposure to the stimuli mattered. Easy-to-hard (progressive) sequencing of stimuli during preexposure led to a more accurate performance with the critical difficult contrast and greater generalization to new contrasts in the task, compared with equally variable training in either a random or an antiprogressive order. This greater accuracy was also evident when participants experienced the progressively sequenced stimuli in a different incidental learning task that did not involve direct auditory training. The results clearly show the importance of temporal dynamics (sequencing) in learning and show that the progressive training advantages cannot be fully explained by direct associations between stimulus features and the corresponding responses. The current findings are consistent with a hierarchical account of perceptual learning, among other possibilities, but not with explanations that focus on stimulus variability.

Keywords: perceptual learning, cognitive training, discrimination, fading

The ability to distinguish perceptual events often depends on experience. This is known as perceptual learning (Goldstone, 1998). Perceptual learning is often enhanced by progressive training. Early experience with an easy discrimination can facilitate the subsequent learning of a difficult version (e.g., Lawrence, 1952). This phenomenon (referred to as the easy-to-hard effect, transfer along a continuum, or fading) directly contradicts the classical prediction that learning transfer should be optimal when testing exactly matches training (Morris, Bransford, & Franks, 1977). The advantages of progressive training have been demonstrated with a variety of species and sensory domains (e.g., Walker, Lee, & Bitterman, 1990).

Although few studies have directly examined this phenomenon in humans (e.g., Liu, Mercado, Church, & Orduña, 2008; McLaren & Suret, 2000; Orduña, Liu, Church, Eddins, & Mercado, in press), this approach has been integrated into various procedures to maximize training. Progressive training has been used for speech discrimination (Tremblay & Kraus, 2002), nonnative phonemic discrimination (McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002), cognitive skills (Anderson, Corbett, Koedinger, & Pelletier, 1995), and attention (Shalev, Tsal, & Mevorach, 2007), to mention a few. Researchers interested in maximizing learning in humans have often assumed that progressive training is advantageous (for a counterexample see Spiering & Ashby, 2008).

Why progressive training may be advantageous is debated. Early explanations based on associative learning hypothesized that there were more efficient trade off relationships between excitatory and inhibitory response gradients (e.g., Logan, 1966). Elemental-associative and selective-attention theories assume that progression aids either the associative learning or the attentional reweighting of the relevant features (e.g., Jamieson & Morosan, 1989; McLaren, Kaye, & Mackintosh, 1989). Both theories assume that learning requires links between the perceptual inputs and the responses. Progressive learning is advantageous because it aids the formation of the most useful associations or directs attention to the most relevant features for responding. Hierarchical perceptual learning theories, on the other hand, assume that tasks of different difficulty initially engage different cortical levels—progressive training engages higher cortical areas early in training, changing their influence on perceptual learning as the task becomes harder (Ahissar & Hochstein, 1997). This view assumes that progressive training changes the way that stimuli are perceptually represented, and the representations guided by higher cortical levels are believed to be more distinguishable than representations formed at lower levels without higher level input. This view has many similarities to theories of selective-attention. However, because the higher level guidance can still

This article was published Online First May 28, 2012.

Barbara A. Church, Eduardo Mercado III, Matthew G. Wisniewski, and Estella H. Liu, Department of Psychology, University at Buffalo, The State University of New York.

This research was supported by National Institute of Mental Heath Grant MH 67952 and National Science Foundation Science of Learning Center Grant SBE 0542013 to the Temporal Dynamics of Learning Center. We thank Jennifer Schneider and the undergraduate research assistants for help with stimuli and data collection. We thank Ian McLaren for helpful suggestions.

Correspondence concerning this article should be addressed to Barbara A. Church, Department of Psychology, Park Hall, University at Buffalo, The State University of New York, Buffalo, NY 14260. E-mail: bchurch@buffalo.edu

be within the perceptual system, associations with responses are not necessary. All three views predict that progressive learning is advantageous. However, only the hierarchical view predicts that progressive exposure will be advantageous even when the stimuli and responses are not linked.

Unlike the previous perspectives, not all theorists are convinced that the sequence of training difficulty is important. Some have suggested that reports of an advantage simply reflect less frustration during training (Eisenberger, 1992; see Liu et al., 2008, for counter). Others suggest that progressive advantages only occur when they aid the discovery of the task relevant dimension. If humans are told what features are relevant, the sequence of training is unimportant (Casale & Pashler, 2011). A number of recent studies that informed participants about the relevant dimension suggest this hypothesis cannot fully explain the findings (Liu et al., 2008; Orduña et al., in press). However, a somewhat more viable alternative explanation of progressive effects assumes that the advantage actually comes from using more variable stimulus sets during progressive training.

Researchers have focused on the critical role of variable training in maximizing learning. Recent interest in the perceptual variability of training sets arose within studies training phonemic contrasts in second languages (e.g., Lively, Pisoni, & Yamada, 1994). More recently, the issue of greater generalization with stimulus variability has also been addressed in animal learning (e.g.Wright & Katz, 2007), second-language vocabulary (e.g., Barcroft & Sommers, 2005), frequency discrimination (e.g., Amitay, Hawkey, & Moore, 2005), category learning (e.g., Hahn, Bailey, & Elvin, 2005), mathematical problem solving (e.g., Sanders, Gonzalez, Murphy, Pesta, & Bucur, 2002), and even tennis (Douvis, 2005). These studies show that the effects of stimulus variability depend on the type of task (Barcroft & Sommers, 2005; Hahn et al., 2005) and individual–group differences (Amitay et al., 2005; Sanders et al., 2002). In general, training variability leads to better learning of second-language contrasts and is comparable to progressive training (Iverson, Hazan, & Bannister, 2005). However, though variable training increases generalization, retention, and overall skill (Douvis, 2005; Sanders et al., 2002), it also slows initial learning (Hahn et al., 2005) and is more helpful to some than to others (Amitay et al., 2005; Sanders et al., 2002).

Only one previous study has directly compared the effectiveness of variable and progressive training. Iverson et al. (2005) trained Japanese speakers to learn the English /r/–/l/ phoneme contrast using different procedures. One procedure progressively trained listeners to discriminate the third formant (F3) crucial for distinguishing this phoneme contrast. Another procedure simply used variable training with multiple speakers. Both procedures were found to be equally advantageous, despite their vast differences. It remains unclear whether nonprogressive but equally variable training along the F3 dimension would have produced the same efficiency. It is important to disentangle this because progressive training is by definition more variable than constant training. Therefore, the benefits of progressive training may result from the increased variability. The hierarchical, elemental-associative, and selective-attention accounts clearly predict that the temporal sequencing, not just the variability of the training, is vital. However, of these three accounts, only the hierarchical account clearly predicts that this will continue to be true when the discrimination is not trained.

Experiment 1

In Experiment 1, we examined people's ability to discriminate birdsongs played back at varying rates. Four types of training (progressive, constant, random, and antiprogressive) were compared to dissociate the impact of progressive sequencing from increased variability. Half of the participants in each learning condition were intentionally trained (asked to discriminate birdsong rates), and half were incidentally exposed to the stimuli (asked to judge whether they thought they would recognize the particular birdsong during a memory test). This incidental learning procedure is similar to the learning phase of artificial grammar (e.g., Reber, 1967) and dot distortion categorization tasks (e.g., Knowlton & Squire, 1993) in which participants are exposed to stimuli but are not asked to focus on the relevant aspects of the stimuli. This allowed us to determine whether associations between the stimuli and responses were necessary for a progressive advantage.

The theory of reverse hierarchy (RH) predicts that progressive training will produce the best performance because perceptual learning is guided by higher cortical areas, producing more differentiated representations. The constant and random conditions should fall somewhere in between because neither would recruit higher level processes, and the antiprogressive group should show the worst performance because higher cortical levels were being recruited at the wrong time. In addition, because the progressive advantage should reflect changes in stimulus representation as a result of apparent contrast rather than an association with a response, the theory predicts that progressive sequencing should show a similar advantage, whether the discrimination is trained or incidentally exposed. There may be an overall advantage of training, but the pattern of advantage across sequencing types should be the same. On the other hand, if the "progressive" advantage is actually produced by stimulus variability, the progressive, random, and antiprogressive conditions should be similar, all sharing an advantage over the constant condition. If associations or selective attention is necessary, then only the intentionally trained conditions should show a progressive advantage.

Method

Participants. Ninety-nine introductory psychology students from the University at Buffalo, The State University of New York, participated as partial fulfillment of their course requirements. Twelve participants were assigned to each of the eight between-participants conditions (2 learning presentation by 4 sequencing type). Three participants were dropped and replaced for having missing values for more than six trials. Final data analyses included 96 participants.

Stimuli and apparatus. The stimuli were constructed from a publicly available recording of superb lyre bird's song from the Macaulay Library at the Cornell Lab of Ornithology. A 1 s section containing various frequency and amplitude modulations was selected. Variants of the original were created using the time stretch function in Cool Edit Pro 2.0, to make stimulus versions that were 10%, 20%, 30%, 40%, 50%, and 70% faster than the original. Increasing the rate in this fashion led to a number of concurrent changes, such as a shorter stimulus duration and steeper frequency and amplitude modulations. Such complex changes resulted in a

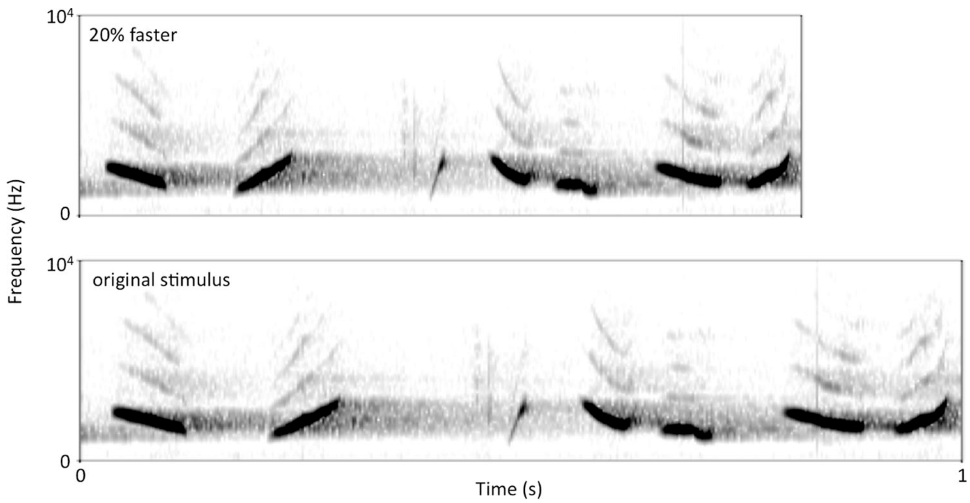

Figure 1. Spectrograms of the original stimulus and the 20% faster stimulus (critical contrast). Other stimuli had similar spectral structure but were time stretched to be 10%, 30%, 40%, 50%, and 70% faster than the original.

variety of acoustic cues that could be used for discrimination. Spectrograms of the original stimulus and the 20% faster version are shown in Figure 1.

Stimuli were presented, and keyboard responses were collected, using DMDX experimental software (Forster & Forster, 2003) running on IBM-compatible desktop computers. Audio-Technica ATH-M40fs headphones presented the stimuli at normal conversational levels.

Design and procedures. A 2 (learning presentation: training vs. exposure) $\times$ 4 (sequencing type: progressive, constant, random, and antiprogressive) between-participants design was used. The dependent variable was discrimination performance with either the critical experienced contrast or novel contrasts not heard during learning.

All participants completed one learning session followed by a test with a break for instructions. Participants were presented with birdsongs and asked to make a two-alternative forced-choice response for each stimulus. During the test and the intentional training, participants decided whether the presented song was the "slow" birdsong (the original recording) or a faster version. Participants responded with keyboard presses, and they were told to guess if they were unsure. They received no feedback during the test. Participants in intentional training conditions received accuracy feedback regarding their trial performance, and the first trial was their first exposure to any stimuli. Participants in incidental exposure conditions heard the same birdsongs but were asked to judge whether they were likely to recognize the stimuli during a later memory test. No feedback was given in exposure conditions. Participants had 5 s to respond to all trials. The intertrial interval was 1.5 s.

Learning sessions had 60 trials. Half were "slow" (the original recording; one speed only), and half were "fast" (faster versions; varied depending on training type). During constant training, the "fast" stimuli were always the 20% faster version. During progressive training, participants cycled through four stages of 15 trials. Each stage compared the "slow" trials with a different speed of "fast" trials, from 70%, 50%, 30%, and 20% faster in the last stage, with no breaks indicating the transitions between stages. During random training, participants heard the same stimuli, but the stimuli were presented in a random order rather than in stages. During antiprogressive training, participants heard the stimuli in the opposite order (20%, 30%, 50%, and 70% faster).

The test was identical across conditions. Half the 100 trials were "slow," and half were "fast." Three versions of "fast" stimuli—10%, 20%, and 40% faster—were used during testing. This allowed for an examination of the effects of sequence on the most difficult experienced discrimination (the critical contrast) and on both easier and harder novel discriminations. All items within learning stages and test were presented in a pseudorandom order (no more than five identical stimuli in a row).

Results

All significance tests were two-tailed with an alpha level of .05. The percentage correct and the discriminability index (d') for each group's test performance on the critical contrast (20% rate difference) and the novel contrasts (10% and 40% rate difference) are presented in Table 1. To correct for potential response biases, all statistical comparisons used d' as the dependent measure.[1]

A 2 $\times$ 4 between-participants analysis of variance (ANOVA) on the critical contrast found a significant main effect of sequencing type, $F(3, 88) = 11.663$, $p < .001$, $\eta^2 = .276$. The main effect of learning presentation was marginally significant, $F(1, 88) = 3.59$, $p = .061$, $\eta^2 = .028$, suggesting a small advantage of intentional training. The interaction was not significant ($F < 1$). Two planned nonorthogonal contrasts were performed to determine whether the data fit the patterns of performance with the different sequencing types predicted by the reverse hierarchy (RH; progressive > constant = random > antiprogressive) and the variability hypoth-

[1] Hit rates and false alarm values were corrected following Wixted and Lee (2004). False alarms of zero were replaced by 1/(2 $\times$ Total Trials). Hit rates of 100 were replaced by 1 − [1/(2 $\times$ Total Trials)].

Table 1
Mean Percentage Correct (%) and Discriminability Index (d′) of the Critical Contrast and Novel Contrasts by Learning Presentation and Sequencing Type in Experiment 1

	Training			
Learning contrast	Progressive	Constant	Random	Antiprogressive
		Critical		
Training				
20% (%)	86	79	79	74
20% (d')	2.170	1.536	1.492	1.086
Exposure				
20% (%)	87	75	76	70
20% (d')	1.899	1.344	1.210	0.852
		Novel		
Training				
40% (%)	92	86	84	88
40% (d')	2.562	2.139	2.101	2.342
10% (%)	73	61	57	56
10% (d')	1.113	0.449	0.215	0.136
Exposure				
40% (%)	97	86	89	83
40% (d')	2.315	1.447	1.738	1.197
10% (%)	62	60	57	58
10% (d')	0.459	0.426	0.186	0.200

eses (VH; progressive = random = antiprogressive > constant). The contrast testing the RH trend was significant, $F(1, 88) = 34.045$, $p < .001$, $\eta^2 = .269$.[2] The contrast testing the VH trend was not ($F < 1$). Planned comparisons between conditions showed that the progressive groups performed better than did the constant groups, $t(46) = 3.256$, $p = .002$, Cohen's $d = 0.961$, and the random groups performed significantly better than did the antiprogressive groups, $t(46) = 2.093$, $p = .039$, $d = 0.616$. The constant groups were numerically but not statistically better than the random groups ($t < 1$).

Even though there was no significant interaction between learning and sequencing type ($F < 1$), it was theoretically important to know whether the RH predicted sequence type pattern was significant even when only the exposure condition was examined. To this end, a one-way ANOVA was performed on just the participants in the incidental condition. The main effect of sequence type was significant, $F(3, 44) = 4.934$, $p = .005$, $\eta^2 = .252$. Planned contrasts showed that the RH trend was significant, $F(1, 44) = 14.328$, $p < .001$, $\eta^2 = .243$,[3] and the VH trend was not ($F < 1$). Planned comparisons showed that the progressive group performed marginally better than did the constant group, $t(22) = 2.001$, $p = .051$, Cohen's $d = 0.855$, and the random group performed numerically but not statistically better than did the antiprogressive group, $t(22) = 1.294$, $p = .203$, $d = 0.552$.

To evaluate how learning generalized to novel stimuli, we performed a $2 \times 2 \times 4$ mixed ANOVA on the d's for the 10% and 40% faster stimuli. Analyses revealed significant main effects of sequencing type, $F(3, 88) = 8.519$, $p < .001$, $\eta^2 = .054$, learning presentation, $F(1, 88) = 15.271$, $p < .001$, $\eta^2 = .032$ (higher in the trained groups), and stimuli, $F(1, 88) = 313.815$, $p < .001$, $\eta^2 = .540$ (better with 40% faster stimuli). There were also significant interactions between stimuli and learning presentation, $F(1, 88) = 6.380$, $p = .013$, $\eta^2 = .011$, and stimuli, learning presentation, and sequence type, $F(3, 88) = 3.586$, $p = .017$, $\eta^2 = .019$. All other interactions were nonsignificant ($Fs < 1$). Planned contrasts revealed a significant RH trend, $F(1, 88) = 5.596$, $p = .022$, $\eta^2 = .081$,[4] but no VH trend ($F < 1$). Planned comparisons indicated better generalization in the progressive groups than in the constant groups, $t(46) = 3.555$, $p < .001$, $d = 1.048$. However, the random groups generalized numerically, but not statistically, better than did the antiprogressive groups ($t < 1$).

To better understand the interaction between the stimuli and the learning presentation and the three-way interaction, separate 2×2 between-participant ANOVAs were performed on the 40% and 10% faster stimuli. For the easier stimuli, there were significant main effects of sequencing type, $F(3, 88) = 4.393$, $p = .006$, $\eta^2 = .107$, and learning, $F(1, 88) = 16.838$, $p < .001$, $\eta^2 = .136$ (trained was better), and a nonsignificant interaction, $F(3, 88) = 1.822$, $p = .149$, $\eta^2 = .044$. To further understand the three-way interaction, planned contrasts testing the sequencing hypotheses for each learning condition separately showed a significant RH trend in the exposure, $F(1,$

[2] The residual sequencing type main effect not explained by the RH trend contrast was not significant ($F < 1$), suggesting that the contrast largely explained the main effect.

[3] The residual sequencing type main effect not explained by the RH trend contrast was not significant ($F < 1$), suggesting that the contrast largely explained the main effect.

[4] The residual sequencing type main effect not explained by the RH trend contrast was significant, $F(2, 88) = 9.160$, $p < .001$, $\eta^2 = .115$, suggesting (like the three-way interaction) that not all performance was explained by the RH trend.

44) = 11.772, $p = .001$, $\eta^2 = .206$,[5] but not the trained group ($F <$ 1). This probably reflects a ceiling effect in the better performing trained group with the easy stimulus contrast.

For the more difficult stimuli, there was a significant main effect of sequencing type, $F(3, 88) = 6.131$, $p < .008$, $\eta^2 = .161$. The main effect of learning ($F < 2$) and interaction, $F(3, 88) = 2.077$, $p = .109$, $\eta^2 = .054$, were nonsignificant. Planned contrasts for each learning condition showed a significant RH trend for the trained group, $F(1, 44) = 22.105$, $p < .001$, $\eta^2 = .310$,[6] but not the exposure group ($F < 2$). This may reflect a floor effect in the incidental condition with the most difficult contrast. No groups showed significant VH trends (Fs < 1).

Discussion

The results show that direct associations between relevant stimulus features and a particular response are not necessary for the progressive advantage to occur. Progressively exposed individuals showed a performance advantage even when they had no training about rate discrimination. The results also seem to support the idea that the sequence, and not just the variability of the exposure, is important. However, there is one possible problem with this interpretation of the critical contrast results in the current experiment. There is a potential confound between the sequence of the training and the probability that the last item heard during training was the fast item from the critical contrast. Though the probability of the last item being that item is the same for the progressive and the constant conditions (50%), it is lower for the conditions with equal variability (0% for antiprogressive and 12.5% for random). Though these complex sound stimuli are unlikely to be retained in a short-term memory buffer for more than 10–20 s (e.g., Cowan, 1984) and the delay for the participants to read the screen and press the space bar to begin the test requires more than 20 s, it is possible that some kind of priming for recent information could operate to give the progressive condition an advantage over the other variable conditions with the critical contrast. The operation of the combination of variability and an advantage related to the most recent stimulus could explain the current results with the critical contrast. How one explains the generalization results with this sort of priming from the most recent stimulus is less obvious. However, a less confounded interpretation of the critical contrast is important to understanding the progressive effect.

Previous research has shown that a progressive over constant training advantage is still evident after a 1-day delay (Liu et al., 2008), clearly showing that it operates on long-term learning, not just performance in the short-term. However, this research cannot disentangle the sequencing and variability of training. In order to assess this combined variability and priming hypothesis, we conducted Experiment 2.

Experiment 2

In Experiment 2, we compared equally variable training conditions with three different sequences (progressive, random, antiprogressive), while controlling the probability that the last training item was the stimuli used for the critical contrast.

Method

Participants. Thirty-one introductory psychology students from the University at Buffalo, The State University of New York, participated as partial fulfillment of their course requirements. Ten participants were assigned to each of three between-participants conditions (progressive, random, and antiprogressive training). One participant was dropped and replaced for making one response more than 80% of the time. Final data analyses included 30 participants.

Design. The experiment employed a simple between-participants design. The independent variable was sequencing type (progressive, random, and antiprogressive). The dependent variables were discrimination of the critical difficult contrast and the novel contrasts.

Stimuli and procedures. The stimuli and procedures were the same as in Experiment 1 except that (a) for simplicity of comparison and to maximize statistical power, only three intentionally trained conditions were used (progressive, random, antiprogressive) and (b) in every condition, the last six items during training had a 50% probability of being the 20% faster item. The overall and relative number of times that the particular stimuli were played was identical across conditions and to Experiment 1.

Table 2
Mean Percentage Correct (%) and Discriminability Index (d′) of the Critical Contrast and Novel Contrasts by Sequencing Type in Experiment 2

Training contrast type	Progressive	Random	Antiprogressive
Critical 20%			
%	87	79	73
d′	1.976	1.395	1.240
Novel 40%			
%	94	91	83
d′	2.656	2.517	1.920
Novel 10%			
%	67	60	59
d′	0.716	0.335	0.343

Results and Discussion

The percentage correct and the discriminability index (d') for each group's test performance on the critical contrast and novel contrasts are presented in Table 2. All statistical comparisons used d' as the dependent measure.

A one-way ANOVA on the critical contrast found a significant main effect of sequencing type, $F(2, 27) = 5.796$, $p = .008$, $\eta^2 =$ .300, showing differences in participants' discrimination ability after different sequences of training. A planned contrast revealed that the data fit the RH predicted linear trend (progressive > random >

[5] The residual sequencing type main effect not explained by the RH trend contrast was not significant ($F < 1$), suggesting that the contrast largely explained the main effect.

[6] The residual sequencing type main effect not explained by the RH trend contrast was not significant, $F(2, 44) = 2.613$, $p < .079$, $\eta^2 = .073$, suggesting that the contrast mostly explained the main effect.

antiprogressive), $F(1, 27) = 11.265$, $p = .002$, $\eta^2 = .292$.[7] Planned comparisons between conditions showed that the progressive group performed better than the random group, $t(18) = 2.174$, $p = .039$, Cohen's $d = 1.025$, and the random group performed numerically but not statistically better than the antiprogressive group, $t(18) = 1.18$, $p = .247$, Cohen's $d = 0.557$.

A 2 × 3 mixed ANOVA exploring the novel contrasts found significant main effects of sequence type, $F(2, 27) = 6.470$, $p = .005$, $\eta^2 = .048$, and stimuli, $F(1, 27) = 251.060$, $p < .001$, $\eta^2 = .838$. The interaction was not significant, $F(2, 27) = 2.152$, $p = .139$, $\eta^2 = .014$. The planned contrast revealed a significant RH linear trend, $F(1, 27) = 12.922$, $p = .001$, $\eta^2 = .324$.[8]

These results show that participants who experienced a progressive sequence during training showed better discrimination of the critical contrast even when the probability of hearing it as the last item was equal across conditions. The progressive sequence, and not just whether it was the last item heard during training, seems to be important.

General Discussion

The results support the idea that temporal dynamics (easy-to-hard sequencing) confer an advantage in learning. Among all sequencing types, progressive sequencing showed the highest discrimination of the critical contrast, as well as the best generalization to novel contrasts even when participants were not intentionally trained to discriminate rate. Most importantly, although participants experienced the exact same stimuli with identical stimulus variability and equal probabilities of having the last trained item be the critical contrast, progressive sequencing still produced better learning than either antiprogressive or random sequencing. This makes it unlikely that the advantage of progressive training comes from increased stimulus variability. This was important to establish because previous research has confounded the temporal dynamics and the variability of training (e.g., Liu et al., 2008; McLaren & Suret, 2000), and variability and interleaved learning are both known to positively affect skill learning (e.g., Schmidt & Bjork, 1992; Lively et al., 1994). Not only do random and antiprogressive sequencing lack the advantage of progressive sequencing, antiprogressive sequencing significantly hurts critical contrast discrimination compared with constant, and adding random variability does not increase performance compared with having more training with the hard discrimination. At least, in this context, increasing variability does not seem to be helpful.

Progressive sequencing also generalized best to novel discriminations, confirming past reports that progressive training with complex sounds can enhance discrimination of novel sounds (e.g., Merzenich et al., 1996). Likewise, this advantage cannot be explained by stimulus variability during training because neither the antiprogressive groups nor the random groups showed greater generalization than the constant groups did. The findings highlight the importance of temporal sequencing for perceptual learning.

Finally, these results show that the advantage of progressive training cannot be explained by simple versions of either elemental-associative or selective-attention theories that assume that the advantage is caused by learning task-relevant features. The fact that progressive sequencing improved rate discrimination with incidental exposure shows that associations between the stimulus properties and the relevant response are not necessary for learning.

Where does this leave our understanding of the temporal dynamics of perceptual learning? These results together with previous findings (Liu et al., 2008; Orduña et al., in press) make it clear that the sequence of exposure to the stimuli is important for improving resolution and generalizing distinctions to novel contrasts. These progressive effects do not seem to be explained by confounds with motivational factors (Eisenberger, 1992), attentional foci (Casale & Pashler, 2011), or stimulus variability.

Why is this sequence advantageous to perceptual learning? Elemental-associative and selective-attention theories have been important to our understanding of a variety of phenomena, and their ability to characterize a wide range of findings with relatively simple models has been illuminating (e.g., McLaren et al., 1989; Petrov, Dosher, & Liu, 2005). However, the current experiments suggest that they are not (in their most straightforward form) sufficient to fully explain the progressive advantage.

On the other hand, because reverse hierarchy theory (Ahissar & Hochstein, 1997) allows top-down perceptual influences to play a role on lower level perception without the need for decision–response level processes, it can account for the current results. That does not mean that this theory can fully characterize all of the data relating to progressive learning. Because of the way that reverse hierarchy theory assumes that top-down processes influence perceptual differentiation, the "progressive" effect is really an "anchoring" effect. Unlike elemental-associative theories, all of the advantage of sequencing is carried by the initial experience with the easy discrimination, and the rest of the progression should be relatively unimportant. Both early research with animals (e.g., Lawrence, 1952) and recent research with humans (Radell, Wisniewski, Church, & Mercado, 2012) show that progressive training produces better learning than anchored training. These findings indicate that like elemental-associative and selective-attention theories, reverse hierarchy theory is also not sufficient to fully explain the progressive advantage.

Further theoretical advances are likely needed. Simple elemental representations of perceptual information without combinatorial properties that can be changed by experience may not fully capture the changes that training with complex stimuli generates. At the same time, theories that rely upon quick all-or-none top-down learning, whether from within perception or at a decision level (Ahissar & Hochstein, 1997; Casale & Pashler, 2011), may also be inadequate. We believe it is time to consider models that allow for incremental change (see Saksida, 1999) to the representation of perceptual information at multiple levels (not just input and output). It will likely be important for future models to test different multilevel architectures to fully account for how perception and perceptual representations change with experience.

[7] The residual sequencing type main effect not explained by the RH trend contrast was not significant ($F < 1$), suggesting that the contrast largely explained the main effect.

[8] The residual sequencing type main effect not explained by the RH trend contrast was not significant ($F < 1$), suggesting that the contrast largely explained the main effect.

References

Ahissar, M., & Hochstein, S. (1997). Task difficulty and visual hierarchy: Reverse hierarchies in sensory processing and perceptual learning. *Nature, 387,* 401–406. doi:10.1038/387401a0

Amitay, S., Hawkey, D. J. C., & Moore, D. R. (2005). Auditory frequency discrimination learning is affected by stimulus variability. *Perception & Psychophysics, 67,* 691–698. doi:10.3758/BF03193525

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4,* 167–207. doi:10.1207/s15327809jls0402_2

Barcroft, J., & Sommers, M. S. (2005). Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition, 27,* 387–414. doi:10.1017/S0272263105050175

Casale, M., & Pashler, H. (2011, November). *Enhancing perceptual category learning through attentional control*. Paper presented at the 52nd Annual Meeting of the Psychonomics Society, Seattle, WA.

Cowan, N. (1984). On short and long term auditory stores. *Psychological Bulletin, 96,* 341–370. doi:10.1037/0033-2909.96.2.341

Douvis, S. (2005). Variable practice in learning the forehand drive in tennis. *Perceptual and Motor Skills, 101,* 531–545.

Eisenberger, R. (1992). Learned industriousness. *Psychological Review, 99,* 248–267. doi:10.1037/0033-295X.99.2.248

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computing, 35,* 116–124. doi:10.3758/BF03195503

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49,* 585–612. doi:10.1146/annurev.psych.49.1.585

Hahn, U., Bailey, T. M., & Elvin, L. B. C. (2005). Effects of category diversity on learning, memory, and generalization. *Memory & Cognition, 33,* 289–302. doi:10.3758/BF03195318

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/–/l/ to Japanese adults. *Journal of the Acoustical Society of America, 118,* 3267–3278. doi:10.1121/1.2062307

Jamieson, D. G., & Morosan, D. E. (1989). Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology, 43,* 88–96. doi:10.1037/h0084209

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science, 262,* 1747–1749. doi:10.1126/science.8259522

Lawrence, D. H. (1952). The transfer of a discrimination along a continuum. *Journal of Comparative and Physiological Psychology, 45,* 511–516. doi:10.1037/h0057135

Liu, E. H., Mercado, E., III, Church, B. A., & Orduña, I. (2008). The easy-to-hard effect in human (*Homo sapiens*) and rat (*Rattus norvegicus*) auditory identification. *Journal of Comparative Psychology, 122,* 132–145. doi:10.1037/0735-7036.122.2.132

Lively, S. E., Pisoni, D. B., & Yamada, R. A. (1994). Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96,* 2076–2087. doi:10.1121/1.410149

Logan, F. A. (1966). Transfer of discrimination. *Journal of Experimental Psychology, 71,* 616–618. doi:10.1037/h0023012

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the –[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective & Behavioral Neuroscience, 2,* 89–108. doi:10.3758/CABN.2.2.89

McLaren, I. P. L., Kaye, H., & Mackintosh, N. J. (1989). An associative theory of the representation of stimuli: Applications to perceptual learning and latent inhibition. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (pp. 102–130). Oxford, England: Oxford University Press.

McLaren, I. P. L., & Suret, M. (2000). Transfer along a continuum: Differentiation or association? In *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Erlbaum.

Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., & Tallal, P. (1996). Temporal processing deficits of language-learning impaired children ameliorated by training. *Science, 271,* 77–81. doi:10.1126/science.271.5245.77

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer-appropriate processing. *Journal of Verbal Learning & Verbal Behavior, 16,* 519–533. doi:10.1016/S0022-5371(77)80016-9

Orduña, I., Liu, E. H., Church, B. A., Eddins, A., & Mercado, E., III. (2012). Evoked-potential changes following discrimination learning involving complex sounds. *Clinical Neurophysiology, 123,* 711–719.

Petrov, A. A., Dosher, B. A., & Liu, Z. L. (2005). The dynamics of perceptual learning: An incremental reweighting model. *Psychological Review, 112,* 715–743.

Radell, M. L., Wisniewski, M. G., Church, B. A., & Mercado, E., III. (2012). *Baby steps: Progressive training does more (and less) than redirect attention.* Manuscript submitted for publication.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning & Verbal Behavior, 6,* 855–863. doi:10.1016/S0022-5371(67)80149-X

Saksida, L. M. (1999). Effects of similarity and experience on discrimination learning: A nonassociative connectionist model of perceptual learning. *Journal of Experimental Psychology: Animal Behavior Processes, 25,* 308–323. doi:10.1037/0097-7403.25.3.308

Sanders, R. E., Gonzalez, D. J., Murphy, M. D., Pesta, B. J., & Bucur, B. (2002). Training content variability and the effectiveness of learning: An adult age assessment. *Aging Neuropsychology, and Cognition, 9,* 157–174. doi:10.1076/anec.9.3.157.9614

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207–217. doi:10.1111/j.1467-9280.1992.tb00029.x

Shalev, L., Tsal, Y., & Mevorach, C. (2007). Computerized progressive attentional training (CPAT) program: Effective direct intervention for children with ADHD. *Child Neuropsychology, 13,* 382–388. doi:10.1080/09297040600770787

Spiering, B. J., & Ashby, F. G. (2008). Initial training with difficult items facilitates information integration but not rule-based category learning. *Psychological Science, 19,* 1169–1177. doi:10.1111/j.1467-9280.2008.02219.x

Tremblay, K. L., & Kraus, N. (2002). Auditory training induces asymmetrical changes in cortical neural activity. *Journal of Speech, Language, and Hearing Research, 45,* 564–572. doi:10.1044/1092-4388(2002/045)

Walker, M. M., Lee, Y., & Bitterman, M. E. (1990). Transfer along a continuum in the discriminative learning of honeybees (*Apis mellifera*). *Journal of Comparative Psychology, 104,* 66–70. doi:10.1037/0735-7036.104.1.66

Wixted, J., & Lee, K. (2004). *Signal detection theory*. Retrieved from University of California, San Diego, Department of Psychology website: http://psy.ucsd.edu/~kang/sdt/sdt.htm

Wright, A. A., & Katz, J. S. (2007). Generalization hypothesis of abstract-concept learning: Learning strategies and related issues in *Macaca mulatta, Cebus apella,* and *Columba livia. Journal of Comparative Psychology, 121,* 387–397. doi:10.1037/0735-7036.121.4.387

Received July 28, 2011
Revision received March 9, 2012
Accepted March 19, 2012 ■