Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design

Tyler K. Perrachione
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Jiyeon Lee, Louisa Y. Y. Ha, and Patrick C. M. Wong[a)]
Hugh Knowles Center for Clinical and Basic Science in Hearing and Its Disorders, Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, Illinois 60208

(Received 31 August 2010; revised 2 May 2011; accepted 4 May 2011)

Studies evaluating phonological contrast learning typically investigate either the predictiveness of specific pretraining aptitude measures or the efficacy of different instructional paradigms. However, little research considers how these factors interact—whether different students learn better from different types of instruction—and what the psychological basis for any interaction might be. The present study demonstrates that successfully learning a foreign-language phonological contrast for pitch depends on an interaction between individual differences in perceptual abilities and the design of the training paradigm. Training from stimuli with high acoustic-phonetic variability is generally thought to improve learning; however, we found high-variability training enhanced learning only for individuals with strong perceptual abilities. Learners with weaker perceptual abilities were actually impaired by high-variability training relative to a low-variability condition. A second experiment assessing variations on the high-variability training design determined that the property of this learning environment most detrimental to perceptually weak learners is the amount of trial-by-trial variability. Learners' perceptual limitations can thus override the benefits of high-variability training where trial-by-trial variability in other irrelevant acoustic-phonetic features obfuscates access to the target feature. These results demonstrate the importance of considering individual differences in pretraining aptitudes when evaluating the efficacy of any speech training paradigm.
© *2011 Acoustical Society of America*. [DOI: 10.1121/1.3593366]

PACS number(s): 43.71.Es, 43.66.Hg, 43.71.Ft, 43.71.Hw [MSS] Pages: 461–472

I. INTRODUCTION

Much work has been done to quantify the range and sources of individual differences in learning success in a variety of domains. The principal goal of this research has often been to identify the variables that can be measured prior to training that best predict performance outcome, including academic and occupational success (Kuncel *et al.*, 2004; Kuncell and Hezlett, 2007). Second-language acquisition research, in particular, has seen extensive efforts at quantifying individual variability and its relationship to learning success (e.g., Golestani and Zatorre, 2009). Many studies have shown that cognitive factors such as phonological awareness and phonological working memory predict measures of second-language acquisition such as vocabulary growth (Cheung, 1996; Hu, 2003; Speciale *et al.*, 2004; Majerus *et al.*, 2008). Likewise, there has been extensive work reported in the applied and pedagogical literatures focusing on the role of high-level psychological correlates of learning success, including generalized intelligence, personality, and motivation (Ehrman and Oxford, 1995; Robinson, 2001; Dörnyei, 2006). Investigators have also begun to employ noninvasive neuroimaging technologies to identify potential biological predictors of language-learning outcomes, including electrophysiological (Díaz *et al.*, 2008), neurophysiological (Mei *et al.*, 2008; Wong *et al.*, 2007a), and anatomical (Golestani *et al.*, 2002; Wong *et al.*, 2008) correlates of learning achievement.

The relative benefits of various instructional approaches also have been intensely studied. Success at second-language learning has been compared across a variety of training methodologies, including laboratory-based (Wayland and Li, 2008; Hardison, 2003; Lively *et al.*, 1993) and classroom-based studies (Norris and Ortega, 2001). However, little work in the field of speech or second-language learning has addressed whether learners with different pretraining aptitudes might differentially benefit from these various training paradigm designs. Of the few empirical reports that examined whether students' learning aptitudes constrain instructional efficacy (Cronbach and Snow, 1977; Snow *et al.*, 1980; Taylor *et al.*, 2010), none has been situated in the context of speech or language learning.

One feature of speech and language training paradigms frequently employed to enhance learning is that of high stimulus variability (e.g., Wang *et al.*, 1999; Flege, 1995; Kingston, 2003). A large literature on phonological contrast learning has shown that training environments with high acoustic-phonetic stimulus variability, which expose learners to a wide variety of exemplars of the feature or contrast to be learned, result in more robust representations of the learned features, thus improving generalization to novel stimuli in

[a)]Author to whom correspondence should be addressed. Electronic mail: pwong@northwestern.edu

Downloaded 10 Oct 2011 to 18.93.5.211. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

addition to enhancing learning of the trained stimulus set (Lively *et al.*, 1993; Clopper and Pisoni, 2004; Barcroft and Sommers, 2005). Similar results have been observed in morphosyntactic learning (Brooks *et al.*, 2006).

While variability in certain acoustic properties not affecting the phonetics of speech, such as overall amplitude, does not appear to influence listeners' speech perception performance or learning, variability in features that do affect phonetics, such as speaking rate, style, or talker variability, has considerable impact on listeners' performance on speech perception tasks (Sommers and Barcroft, 2007; Bradlow *et al.*, 1999; Sommers *et al.*, 1994). Training environments with high or low acoustic-phonetic stimulus variability due to, e.g., differences between talkers differ substantially in the demands they make on listeners' speech perception processes. When individuals listen to speech in an environment with low acoustic-phonetic variability (i.e., a single talker), they are able to take advantage of processes that adapt to the consistent, predictable features of a talker's phonetics (Johnson and Mullennix, 1997), thus speeding recognition and improving accuracy relative to multi-talker environments (Mullennix and Pisoni, 1990; Magnuson and Nusbaum, 2007). In an environment with high acoustic-phonetic variability (i.e., multiple talkers), there is no trial-by-trial consistency or predictability in the stimuli's phonetic features, so additional processing resources must be deployed for recognizing speech sounds, slowing response times and reducing accuracy, especially in adverse listening conditions (Ben-Artzi and Marks, 1999; Green *et al.*, 1997; Nygaard *et al.*, 1994). Indeed, the increased processing cost incurred when listening in unpredictable, high phonetic variability environments has been demonstrated across numerous behavioral and physiological studies (Mullennix and Pisoni, 1990; Wong *et al.*, 2004; Kaganovich *et al.*, 2006; Creel *et al.*, 2008). Because of these increased processing costs, there is reason to suppose that some listeners may be overwhelmed by such high-variability training environments, impairing their ability to attend the target features or contrasts to be learned. Beyond precluding the benefit typically associated with exposure to multiple exemplars, further obfuscating the target contrast in a high-variability environment may even impair learning relative to low-variability training.

In this study, we examined how individual differences in pretraining predictors of language-learning aptitude interact with the design of various training paradigms to determine achievement in perceptual learning of a novel phonological contrast. Participants in our study learned to recognize a vocabulary of simulated foreign-language words. In order to master the vocabulary, participants had to learn to use a novel phonological contrast called "lexical tone" to distinguish the words. Such a phonological contrast involves pitch contours and is common to many languages (including notably the Chinese languages), but is not present in English or most other Indo-European languages. Previous research has indicated that the major predictor of successfully learning to use lexical tones is the ability to perceive pitch contours in non-lexical contexts (Wong and Perrachione, 2007; Moreno *et al.*, 2009). Participants in our study exhibited a wide range of individual differences in pitch-perception abilities, which suggested a wide range of learning outcomes. We assessed whether the learning outcomes of individuals with either high or low pretraining aptitudes would be differentially affected by the amount of stimulus variability in the training paradigm. The following experiments demonstrate (1) that a high-variability training environment differentially impairs or enhances learning depending on listeners' pretraining perceptual abilities, and (2) that the nature of this impairment is related to accommodating trial-by-trial stimulus variability. We conclude by discussing the implications of these results both for theoretical models of phonological contrast learning and for classroom educational practices in general.

II. EXPERIMENT 1: INDIVIDUAL DIFFERENCES DETERMINE THE BENEFIT OR DETRIMENT OF HIGH STIMULUS VARIABILITY

A. Method

1. Participants

Young adult native speakers of American English ($N = 64$) gave informed written consent overseen by the Northwestern University Institutional Review Board to participate in this study. All participants reported normal speech and hearing, were free from psychological or neurological disorders, and had no prior experience with tone languages.

2. Pretraining assessment

Prior to training, participants indicated their music and foreign language expertise in a brief, self-report questionnaire. Participants' phonological awareness and verbal working memory were assessed using the Sound Blending (SB), Auditory Working Memory (AWM), and Numbers Reversed (NR) subtests of the Woodcock-Johnson III tests of achievement (Woodcock *et al.*, 2007). Participants' basic perceptual abilities for pitch were assessed through a Pitch-Contour Perception Test (PCPT). This test consisted of 180 tokens of isolated vowels, each superimposed with a level, rising, or falling pitch contour. The pitch contours of natural vowel recordings were resynthesized with level, rising, and falling pitch patterns interpolated linearly using the pitch–synchronous overlap and add (PSOLA) method implemented in the software PRAAT (Boersma and Weenink, 2005). This resulted in perceptually natural and highly distinguishable stimuli as judged by native Mandarin speakers (see Wong and Perrachione, 2007, for further acoustic descriptions of these assessment stimuli). Participants identified the pitch contours they heard by matching the auditory stimuli to representative arrows on the computer screen (→, ↗, ↘). Previous work suggests that participants who perform above a 70% criterion on the PCPT are likely to successfully learn a novel lexical tone contrast based on pitch contour (Wong and Perrachione, 2007; Song *et al.*, 2008; Chandrasekaran *et al.*, 2010). As such, we divided participants into two groups based on whether they were likely to successfully master the vocabulary [High-Aptitude Learners (HAL), $N = 31$, PCPT $> 70\%$], or attain only a lower level of achievement [Low-Aptitude Learners (LAL), $N = 33$, PCPT $\leq 70\%$]. Half the participants in each group were randomly assigned to

Downloaded 10 Oct 2011 to 18.93.5.211. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

either the High-Variability (HV) or Low-Variability (LV) training conditions.

3. Stimuli

Participants learned a vocabulary of 18 pseudowords, consisting of six syllables (*dree*: /dɹi/, *nuck*: /nʌk/, *fute*: /fjut/, *pesh*: /pɛʃ/, *nare*: /neɹ/, and *vess*: /vɛs/) superimposed with three pitch contours (level, rising, and falling). Each token (syllable-pitch pairing) was associated with 1 of 18 common objects (e.g., table, bus, phone, etc.) Eight native speakers of American English (four male, four female) produced a single token of each syllable, and the three pitch contours were synthesized using the PSOLA method implemented in the software PRAAT. Pitch resynthesis was applied across the sonorous segments of each token. The mean value of each talker's fundamental frequency (F0) across all productions was used as the baseline value for resynthesis. The level pitch contour began and ended at a talker's baseline pitch. The rising pitch contour began at 74% of the baseline pitch and ended at the baseline pitch. The falling pitch contour began at 110% of the baseline pitch, and fell by 45%. To produce talker-specific variability in the target contrast, the onset and offset values of each linear pitch contour were varied by ±3% across talkers. The variability realized in this stimulus set is illustrated in Fig. 1. These pitch contours were modeled on the values obtained by Shih (1988). This method of synthesis has been shown to produce perceptually natural and highly identifiable pitch contours (Wong and Perrachione, 2007).

Variability in the stimuli occurred due to natural differences between talkers (voice quality, syllable duration, average pitch, formant spacing, etc.), as well as synthetic differences in the pitch contours. Tokens from four talkers (two male, two female) were designated for use in only the training conditions, and those from the remaining four talkers were used in only a post-training Test of Learning Achievement (TLA), assessing mastery of the vocabulary and phonological contrast, as well as the ability to generalize learning to novel talkers.

4. Low-variability training

Participants in the LV training condition learned to recognize the pseudoword vocabulary as spoken by only one of the four training talkers. The specific talker used was counterbalanced across participants. Vocabulary items were trained in groups of three that differed minimally by pitch contour (e.g., *pesh* with level, rising, and falling pitch), and participants were given corrective feedback to help them learn the associations between auditory words and line drawings of objects shown on a computer screen. Feedback indicated whether each response was correct and, for errors, the correct response. Each set of three words was trained four times, resulting in 72 training trials per day. After finishing all training trials, participants completed a daily Word Identification Test (WIT), in which they heard the auditory words and selected the corresponding objects from the full set of 18, spoken by the single training talker, randomized, and repeated four times. The WIT was used to track learning progress across training sessions, but the index of ultimate learning outcome was performance on the post-training TLA (see the following). All participants completed 8 days of training sessions, each of which lasted about 1 h.

5. High-variability training

Participants in the HV training condition learned to recognize the pseudoword vocabulary as spoken by all four training talkers. The structure of the training session was the same as the LV condition, except that the productions of the four training talkers were intermixed within and between sets of training stimuli, rather than repeating a single talker's productions four times. Thus, the overall exposure to vocabulary items was the same across these two training conditions (72 training trials per day), but the HV condition contained four times more acoustic-phonetic variability than the LV condition. The WIT of the HV condition likewise consisted of all 18 vocabulary items produced once by each of the four training talkers. The differences between the various training designs are depicted graphically in Fig. 2.

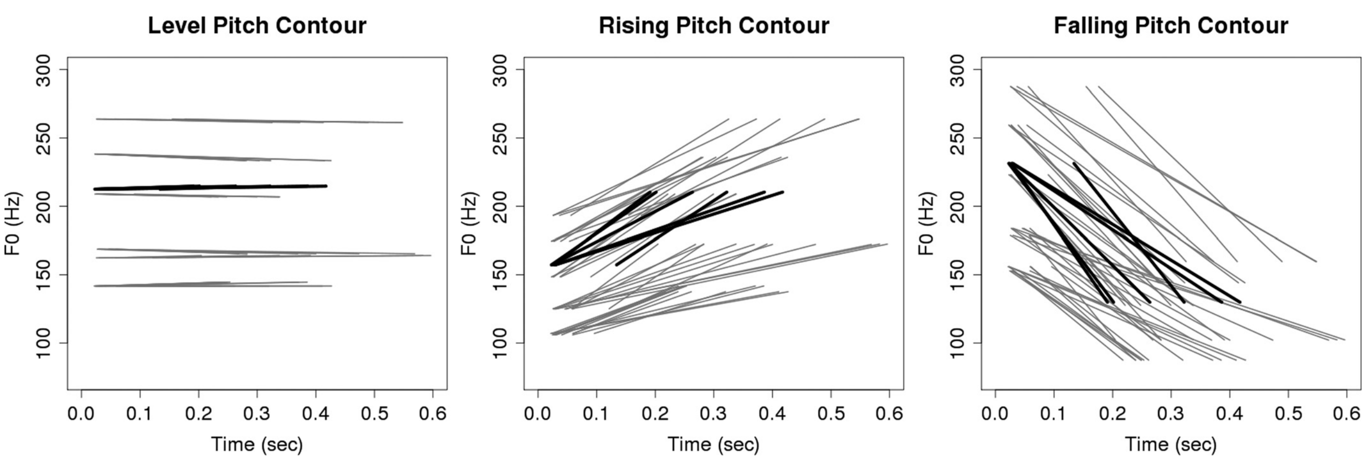

FIG. 1. Variability in synthesized pitch contours. Synthesized pitch contours were based on talkers' own mean fundamental frequency, and variability across stimulus tokens occurred due to natural differences in F0 across talkers, natural differences in vowel duration across tokens, and manipulated differences in synthesized pitch contour range across talkers. Each line represents a single training token; tokens from a given talker have consistent onset and offset pitch values. Bold black lines indicate tokens from one example talker.

Downloaded 10 Oct 2011 to 18.93.5.211. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

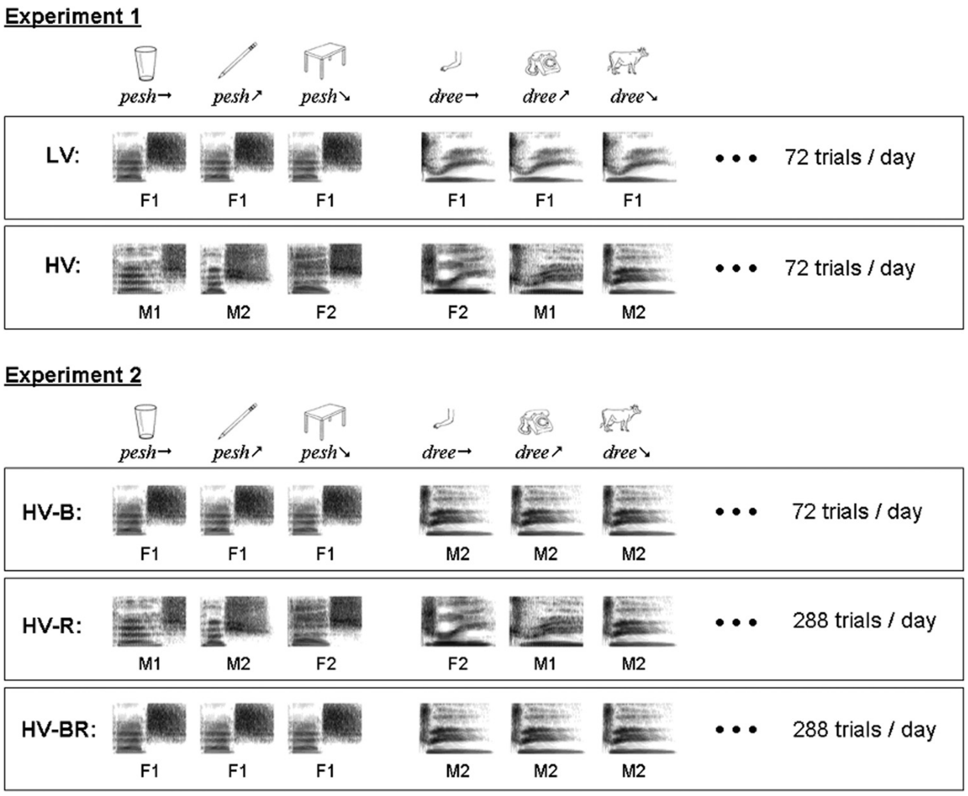

FIG. 2. Training conditions in this study manipulated trial-by-trial acoustic-phonetic variability and overall exposure to the stimuli. Vocabulary items in training trials were organized into triads minimally contrastive by pitch contour. Line drawings and syllables with pitch contours (→, ↗, ↘) reflect the stimulus associations to be learned. Rows indicate the training design manipulations. Spectrograms of the syllables as spoken by each talker (F1, M1, etc.) depict graphically the magnitude of trial-by-trial phonetic variability facing listeners. Some conditions involved no extraneous phonetic variability within training triad (LV, HV-B, HV-BR), while others had substantial extraneous variability (HV, HV-R).

6. *Test of Learning Achievement and generalization*

Participants undertook a TLA, which was also compared against their performance on the final day of training as an index of generalization. After completing all eight training sessions, participants returned to the lab on a separate day to take the TLA. The TLA was similar in structure to the WIT from the HV condition. Each pseudoword was produced once by each of the four (previously unheard) generalization talkers, and participants indicated the appropriate object from the complete list of 18 items. The TLA was presented as a single, uninterrupted test, in which participants heard all the test stimuli produced by a single generalization talker before hearing those produced by the remaining talkers. Participants' performance on the TLA was used as the measure of their ultimate learning attainment to allow unbiased comparisons across experimental conditions in which the daily WIT differed slightly depending on the paradigm design and which talkers were used in training. It bears noting that both the training paradigms and the TLA were strictly perceptual; at no point were participants explicitly required, or assessed for their ability, to produce the target contrasts.

B. Results

1. *Cognitive and perceptual assessments*

Participants' scores on the pretraining cognitive/perceptual assessments were submitted to four 2×2 univariate analyses of variance to investigate differences between groups (HAL vs LAL) and training types (LV vs HV) and interactions. The HAL and LAL groups did not differ in their scores on the SB cognitive test [$F(1,60) = 2.567$, $p = 0.114$], on the AWM cognitive test [$F(1,60) = 0.770$, $p = 0.384$], or on the NR cognitive test [$F(1,60) = 0.622$, $p = 0.434$]. Likewise, participants in LV training did not differ from those in HV training on the SB cognitive test [$F(1,60) = 2.913$, $p = 0.093$], on the AWM cognitive test [$F(1,60) = 0.119$, $p = 0.732$], on the NR cognitive test [$F(1,60) = 0.012$, $p = 0.918$], or on the PCPT [$F(1,60) = 0.418$, $p = 0.521$]. There were no Group × Condition interactions (all $p > 0.290$). Participants' performance on these pretraining assessments are delineated by group and training condition in Table I.

2. *Learning progress*

Participants' progress at learning the vocabulary is illustrated in Fig. 3. To assess the rate at which participants mastered the pseudoword vocabulary during training, we calculated the linear slope of each individual's learning curve between chance (pretraining) and session 4 (the midpoint of training). These values were submitted to a 2×2 univariate analysis of variance (ANOVA) for effects of group (HAL vs LAL) or training type (LV vs HV) or interactions. The HAL group demonstrated significantly faster learning than the LAL group [$F(1,60) = 28.514$, $p < 1.51 \times 10^{-6}$] across training conditions, and both groups demonstrated significantly faster learning during the LV condition than HV [$F(1,60) = 21.090$, $p < 2.3 \times 10^{-5}$]. There was no Group × Condition interaction ($p = 0.117$).

3. *Learning achievement*

Figure 4 shows the two groups' learning achievement in each training condition. Prior to analysis, all proportional (percent) measures, including scores on the TLA, were

Downloaded 10 Oct 2011 to 18.93.5.211. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

TABLE I. Mean (± standard deviation) performance on pretraining assessments of cognition and perception by group and training condition.

Group	Condition	Cognitive/perceptual assessment SB[a]		NR		AWM		PCPT		Musical experience[b]		Language experience[c]
HAL	LV	0.80	(0.11)	0.65	(0.29)	0.80	(0.15)	0.84	(0.08)	5.06	(2.86)	2 / 10 / 4
	HV	0.75	(0.16)	0.74	(0.25)	0.80	(0.20)	0.85	(0.08)	5.27	(3.03)	4 / 9 / 2
	HV-B	0.85	(0.11)	0.77	(0.22)	0.93	(0.06)	0.91	(0.03)	6.50	(3.17)	0 / 9 / 1
	HV-R	0.87	(0.05)	0.77	(0.13)	0.89	(0.10)	0.89	(0.08)	7.00	(3.83)	1 / 4 / 5
	HV-BR	0.76	(0.21)	0.78	(0.16)	0.80	(0.15)	0.86	(0.06)	5.60	(4.55)	2 / 6 / 2
LAL	LV	0.76	(0.19)	0.67	(0.27)	0.74	(0.25)	0.59	(0.08)	2.88	(2.99)	3 / 8 / 5
	HV	0.66	(0.22)	0.60	(0.32)	0.77	(0.24)	0.56	(0.07)	2.47	(2.58)	6 / 9 / 2
	HV-B	0.69	(0.23)	0.74	(0.24)	0.79	(0.19)	0.59	(0.09)	5.50	(4.38)	1 / 8 / 1
	HV-R	0.71	(0.17)	0.60	(0.26)	0.69	(0.21)	0.60	(0.05)	3.60	(3.81)	1 / 6 / 3
	HV-BR	0.62	(0.24)	0.55	(0.28)	0.77	(0.18)	0.58	(0.08)	1.09	(1.51)	1 / 7 / 3

[a]Range on cognitive batteries SB, NR, AWM, and PCPT is 0.0–1.0.
[b]Musical experience (in years) of longest played instrument.
[c]Participants' familiarity with a foreign language: N = (monolingual English/some foreign language familiarity/fluent in another language).

arcsine-transformed (Studebaker, 1985) to meet the assumptions of inferential statistical tests. To assess the extent of successful vocabulary learning, these scores were compared in a 2×2 univariate ANOVA across Group (HAL vs LAL) and Training Condition (LV vs HV). As shown in Fig. 4, learning performance differed significantly between the two learner groups [$F(1,60) = 90.592$, $p < 1.34 \times 10^{-13}$], with the HAL group outperforming the LAL group in both training types. There was a trend toward learning differences based on training condition [$F(1,60) = 3.790$, $p = 0.056$], with LV training resulting in slightly improved learning compared to HV training. This surprising result can be explained in terms of a significant Group × Condition interaction effect [$F(1,60) = 16.314$, $p < 0.00016$]—the HAL group exhibited predictably greater learning achievement following HV training [$t(29) = 1.963$, $p < 0.03$, Cohen's $d = 0.73$]; however, participants in the LAL group demonstrated a surprising and significant impairment in learning from HV training [$t(31) = -3.615$, $p < 0.0011$, $d = 1.30$].

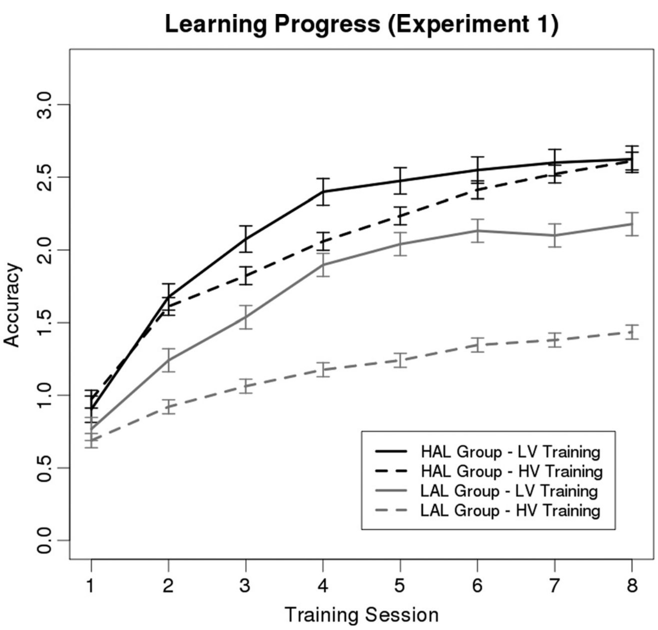

FIG. 3. Learning progress in experiment 1. Participants in the HAL group demonstrated more rapid learning than those in the LAL group. Both groups of participants, particularly LAL, demonstrated more rapid learning from the low stimulus-variability training paradigm. Ordinate values have been arcsine transformed. Error bars indicate the standard error of the mean.

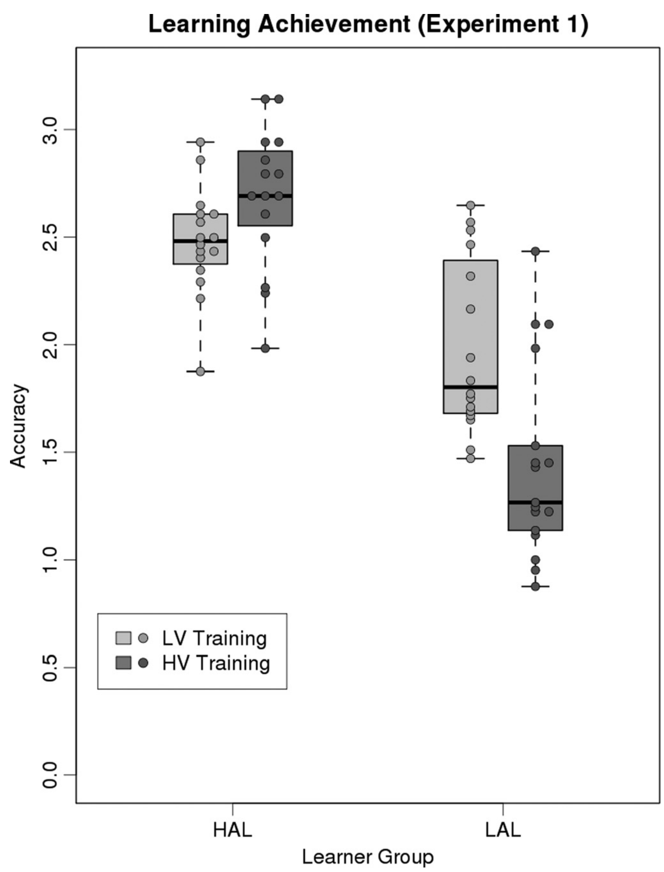

FIG. 4. Instructional paradigms interact with individual differences to determine learning achievement. HV training significantly enhanced learning for the HAL group, whereas the LAL group was significantly impaired by increased stimulus variability. Ordinate values have been arcsine transformed. *Boxplots*: Shaded region indicates interquartile range; whiskers extend to extreme values; solid bar indicates median; points indicate individual participant values, and those of equal value spread along the abscissa to avoid overlap.

Downloaded 10 Oct 2011 to 18.93.5.211. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

We also considered whether pretraining behavioral measures were predictive of learning achievement. A multiple linear regression model, including the four cognitive scores, as well as the training condition as a covariate, revealed the PCPT to be by far the best predictor of learning success, with some additional variance also explained by SB [$\beta_{\text{PCPT}} = 1.112$, $p < 3.28 \times 10^{-12}$; $\beta_{\text{SB}} = 0.880$, $p < 0.002$; $R^2 = 0.686$], (Fig. 5). Scores on NR or AWM were not additionally predictive of learning success (both $p > 0.240$).

4. Generalization

In addition to being a measure of participants' learning achievement, the TLA also provided a means to assess whether there were any differences in participants' ability to generalize to novel talkers compared to their performance on the WIT on the last day of training. We computed participants' generalization index "G" as the ratio of their accuracy on the novel talkers in the TLA to their accuracy on the trained talkers on the WIT from the last day of training. These values were submitted to a 2×2 univariate ANOVA for effects of Group (HAL vs LAL) and Training Condition (LV vs HV) and interactions. There was no effect of Group [$F(1,60) = 1.635$, $p = 0.206$], indicating no difference in HAL and LAL groups' ability to generalize learning to novel talkers. A main effect of Condition [$F(1,60) = 13.351$, $p < 0.00055$] revealed the HV condition to result in greater generalization to novel talkers (mean $G_{\text{HV}} = 1.02$) than the LV condition (mean $G_{\text{LV}} = 0.93$), [$t(62) = 3.614$, $p < 0.00061$, $d = 0.92$]. There was no Group $\times$ Condition interaction, indicating both HAL and LAL groups exhibited better generalization to unheard voices following HV training than LV training.

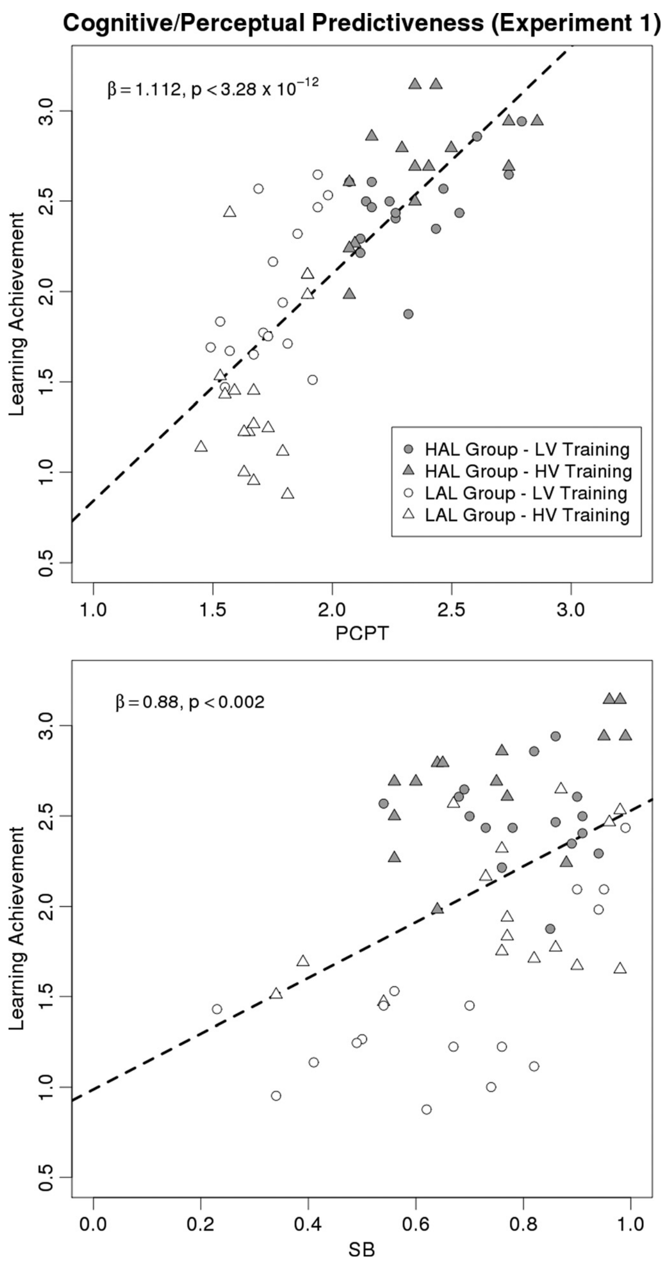

FIG. 5. Pretraining assessment measures differentially predict learning success. Perceptual ability (measured by PCPT) accounts for most of the variance in learning success, with some variance also explained by phonological awareness (SB). Measures of attention and working memory (NR, AWM) did not contribute additional predictiveness.

C. Discussion

Although an extensive literature in both linguistic (Lively *et al.*, 1993; Barcroft and Sommers, 2005; Brooks *et al.*, 2006) and nonlinguistic (Paas and Van Merriënboer, 1994) domains describes enhanced learning following high-variability training, here we see that whether such training is beneficial depends on individual differences in the learners who undergo it. Consistent with previous literature in this domain, all participants exhibited better generalization to novel stimuli following HV than LV training. Critically, however, the different training paradigms produced markedly different levels of learning achievement depending on learners' pretraining aptitudes. Rather than the classically expected improvement in learning achievement, the learning outcomes of participants with low pretraining aptitude measures (LAL) were further impaired in the high-variability environment, unlike high-aptitude learners whose learning outcome was even better following HV training. This result demonstrates an individual–instructional interaction, raising specific pedagogical implications: Instructional approaches designed to maximize learning for students in general may end up being even more deleterious to those students who would already have found the material challenging. (It is worth noting that the increased stimulus variability in the HV training was not without cost even for the HAL group, who, despite significant improvement to their learning achievement scores, learned more slowly in this condition compared to LV training.)

Why was the high-variability training environment disproportionately detrimental to the LAL group's learning achievement? The LV and HV training paradigms differed from one another on two key features: (1) the amount of trial-by-trial acoustic-phonetic variability and (2) the amount of exposure to each individual stimulus token.

First, the greater degree of trial-by-trial variability in the HV condition may have exceeded low-aptitude learners' ability to normalize task-irrelevant variation in the stimuli and attend the auditory dimension most informative to the

Downloaded 10 Oct 2011 to 18.93.5.211. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

task. If this was the source of detriment to learners in the LAL group, then reducing the trial-by-trial variability while retaining global variability in the stimuli should allow those listeners to more successfully identify and attend to the relevant auditory features distinguishing the vocabulary items.

Second, the amount of exposure to individual stimulus tokens differed between the LV and HV conditions. In the LV condition, listeners heard each token repeated four times by a single talker; whereas in the HV condition, each talker–token pairing was heard only once. Low-aptitude learners may require more extensive exposure to any individual stimulus before successfully committing it to memory. If this was the source of detriment to the LAL group, then increasing the number of times each unique stimulus is presented in a high-variability paradigm should facilitate learning for those listeners.

Finally, these two differences may have worked in combination to impair learning in the LAL group, who may require both reduced trial-by-trial variability and increased stimulus exposure for successful learning. To determine which of these features of the high-variability training environment were detrimental to the LAL group's learning achievement, we designed a series of follow-up experiments looking at variations on the HV training paradigm.

III. EXPERIMENT 2: TRIAL-BY-TRIAL VARIABILITY IS THE SOURCE OF DETRIMENT TO LOW-APTITUDE LEARNERS

A. Method

1. Participants

We recruited new participants ($N = 61$) for these experiments who met the same criteria as those in experiment 1. Participants completed the same battery of cognitive/perceptual assessments as in experiment 1 and were assigned to the High-Aptitude Learner (HAL) group ($N = 30$) or Low-Aptitude Learner (LAL) group ($N = 31$) based on their PCPT performance. Participants in each group were assigned randomly among the three HV-training variants.

2. Stimuli

The stimuli were the same as in experiment 1.

3. Blocked high variability (HV-B)

The HV-B condition was identical to the HV condition from experiment 1, except rather than randomly mixing the four training talkers within minimally contrastive sets of training items, sets were presented from only one talker at a time. Thus, all variability within a training block was related to the phonological contrast to be learned and not due to extraneous phonetic differences between talkers. This condition tested the hypothesis that, for the LAL group in the HV condition, learning was impaired by the large amount of trial-by-trial variability in the auditory stimuli, which is largely removed when blocking by talker. Like the HV and LV conditions, the HV-B condition consisted of 72 training tokens per day.

4. Repeated high variability (HV-R)

The HV-R training condition was similar in structure to the HV condition from experiment 1, except rather than hearing the 18 vocabulary items only once from each talker, the training portion of the daily paradigm was repeated four times, resulting in 72 tokens from each of the four talkers. This condition was analogous to undergoing the training portion of the HV condition four times each day. Participants in the HV-R condition had as much exposure to each of the four training talkers as participants in LV condition had with their single training talker, but trial-by-trial variability remained the same as in the original HV condition. This condition tested the hypothesis that the LAL group in the HV condition simply did not have enough experience with any given talker to fully learn the vocabulary. The HV-R condition thus consisted of 288 training tokens per day.

5. Blocked and repeated high variability (HV-BR)

The HV-BR training condition combined the manipulations of the HV-B and HV-R conditions. Here, training stimuli were again blocked by talker (HV-B), as well as repeated four times (HV-R), such that this condition was analogous to undergoing HV-B training four times each day. This condition tested the conjunction of the previous hypotheses: That Group B learners needed *both* to have stimuli blocked by talker to overcome variability *and* to have stimuli repeated a sufficient number of times. The HV-BR condition consisted of 288 training tokens per day. The differences between the various high-variability training design manipulations are depicted graphically in Fig. 2.

B. Results

1. Cognitive and perceptual assessments

Participants' scores on the pretraining cognitive/perceptual assessments were submitted to four 2×2 univariate analyses of variance to investigate differences between group (HAL vs LAL) and training type (HV-B vs HV-R vs HV-BR) and interactions. The HAL and LAL groups differed significantly on all cognitive and perceptual measures, including the SB cognitive test [$F(1,55) = 11.222$, $p < 0.0015$], the AWM cognitive test [$F(1,55) = 9.696$, $p < 0.003$], and the NR cognitive test [$F(1,55) = 6.877$, $p < 0.012$]. Importantly, however, there was no effect of training condition on participants' scores on the SB cognitive test [$F(1,55) = 1.654$, $p = 0.201$], the AWM cognitive test [$F(1,55) = 1.239$, $p = 0.298$], the NR cognitive test [$F(1,55) = 0.897$, $p = 0.414$], or the PCPT [$F(1,55) = 0.985$, $p = 0.380$]; and there were no Group $\times$ Condition interactions (all $p > 0.212$).

2. Learning progress

Participants' progress at learning the vocabulary from the various HV manipulations is illustrated in Fig. 6. The rate of vocabulary acquisition was determined as in experiment 1, and assessed in a $2 \times 2 \times 2$ univariate ANOVA for Group (HAL vs LAL), Blocking (Blocked vs Not), and

Downloaded 10 Oct 2011 to 18.93.5.211. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

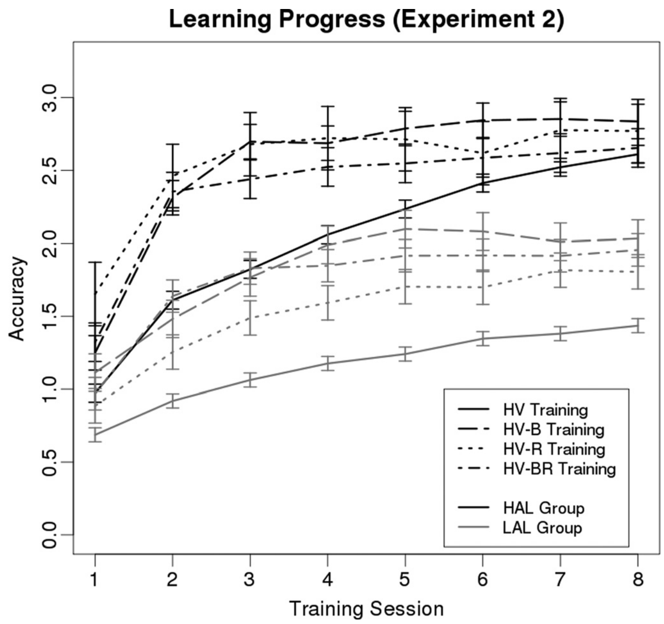

FIG. 6. Learning progress in experiment 2. The HAL group consistently demonstrated more rapid learning than the LAL group. Both blocking and repeating manipulations to the training paradigms increased the rate of learning for both groups compared to the original high-variability design. Error bars indicate the standard error of the mean.

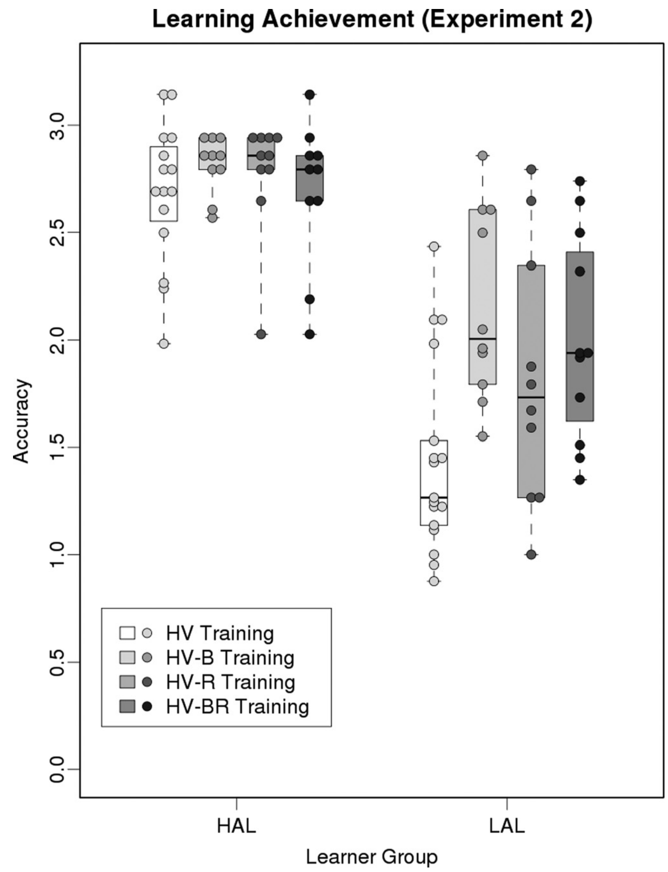

FIG. 7. Tailoring training paradigms to individuals' perceptual abilities helps ameliorate learning achievement. Training paradigms that reduced trial-by-trial perceptual load (HV-B, HV-BR) significantly enhanced learning for participants with weak perceptual abilities. *Boxplots*: Conventions as in Fig. 4.

Repeating (Repeated vs Not) on participants' learning rate from the three training conditions in experiment 2 (HV-B, HV-R, HV-BR), as well as the HV condition from experiment 1, which, being neither blocked nor repeated, fully parametrized the Blocked × Repeated contrast. Like experiment 1, the HAL group demonstrated more rapid mastery of the vocabulary than the LAL group [$F(1,85) = 61.953$, $p < 1.03 \times 10^{-11}$] across conditions. Blocking the stimuli (HV-B, HV-BR), resulted in a significant increase in learning rate [$F(1,85) = 20.153$, $p < 2.23 \times 10^{-5}$] for both groups. Repeating the stimuli also resulted in more rapid learning across groups [$F(1,85) = 5.498$, $p < 0.022$], There were no two- or three-way interactions on learning speed between Group and either the blocking or repeating manipulation (all $p > 0.111$); however, there was a significant Blocked × Repeated interaction [$F(1,85) = 9.244$, $p < 0.0032$], due to the fact that combining the Blocking and Repeating manipulations did not produce more rapid learning than either of these manipulations by itself.

3. Learning achievement

In experiment 2, we were particularly interested in the effect of manipulating the design of the high-variability training condition on the LAL group's learning achievement. To determine the relative benefit to these learners of either the Blocked or Repeated manipulations, we performed a $2 \times 2 \times 2$ univariate ANOVA for Group (HAL vs LAL), Blocking (Blocked vs Not), and Repeating (Repeated vs Not) on the TLA scores from the three training conditions in experiment 2 (HV-B, HV-R, HV-BR) as well as the HV condition from experiment 1. The results of this analysis are illustrated in Fig. 7. In this analysis, there was a significant effect of Group [$F(1,85) = 106.442$, $p < 2.2 \times 10^{-16}$], with the HAL group outperforming the LAL group in all training conditions. There was also a main effect of Blocking [$F(1,85) = 10.042$, $p < 0.0022$], indicating more successful learning in HV-B and HV-BR than HV and HV-R training conditions. There was not a significant effect of Repeating [$F(1,85) = 0.730$, $p = 0.395$], indicating that this manipulation did not improve learning achievement scores. Like the learning-rate results, there was also a significant interaction between Blocking and Repeating [$F(1,85) = 4.858$, $p < 0.031$], due to there being no additional gain in learning performance in the HV-BR condition compared to the HV-B condition, despite the addition of the Repeating treatment. Most important, a significant Group × Blocking interaction [$F(1,85) = 7.213$, $p < 0.009$] revealed the beneficial effects of reducing cognitive load were confined to the LAL group [$t(46) = 3.354$, $p < 0.002$, $d = 1.0$], whereas the HAL group saw no additional gain from this manipulation [$t(43) = 0.373$, $p = 0.711$]. There were no Group × Repeating [$F(1,85) = 0.746$ $p = 0.390$] or three-way [$F(1,85) = 0.868$, $p = 0.354$] interactions.

We confirmed that participants in the LAL group demonstrated greater learning success, as assessed by the TLA, after receiving training with stimuli blocked by talker (HV-B, HV-BR) compared to the original high-variability training design (HV) through a series of independent-sample, Bonferroni-corrected t-tests. There were significant learning outcome gains over HV training in both the HV-B [$t(25) = 3.990$,

Downloaded 10 Oct 2011 to 18.93.5.211. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

$p < 0.002$, $d = 1.65$] and HV-BR [$t(25) = 3.124$, $p < 0.018$, $d = 1.25$] conditions, but not in the HV-R condition [$t(25) = 1.882$, $p = 0.286$]. No such improvement in learning outcome was seen for the HAL group (HV-B vs HV; [$t(25) = 1.190$, $p = 0.246$]).

The predictive relationship between participants' cognitive measures and their learning success was again assessed via multiple linear regression, including the four cognitive tests with training condition as a covariate. As in experiment 1, PCPT was by far most predictive of learning success, with some variance also explained by SB [$\beta_{PCPT} = 1.073$, $p < 4.04 \times 10^{-14}$; $\beta_{SB} = 0.558$, $p < 0.002$; $R^2 = 0.660$] (Fig. 8). Scores on NR or AMW were not additionally predictive of learning success (both $p > 0.459$).

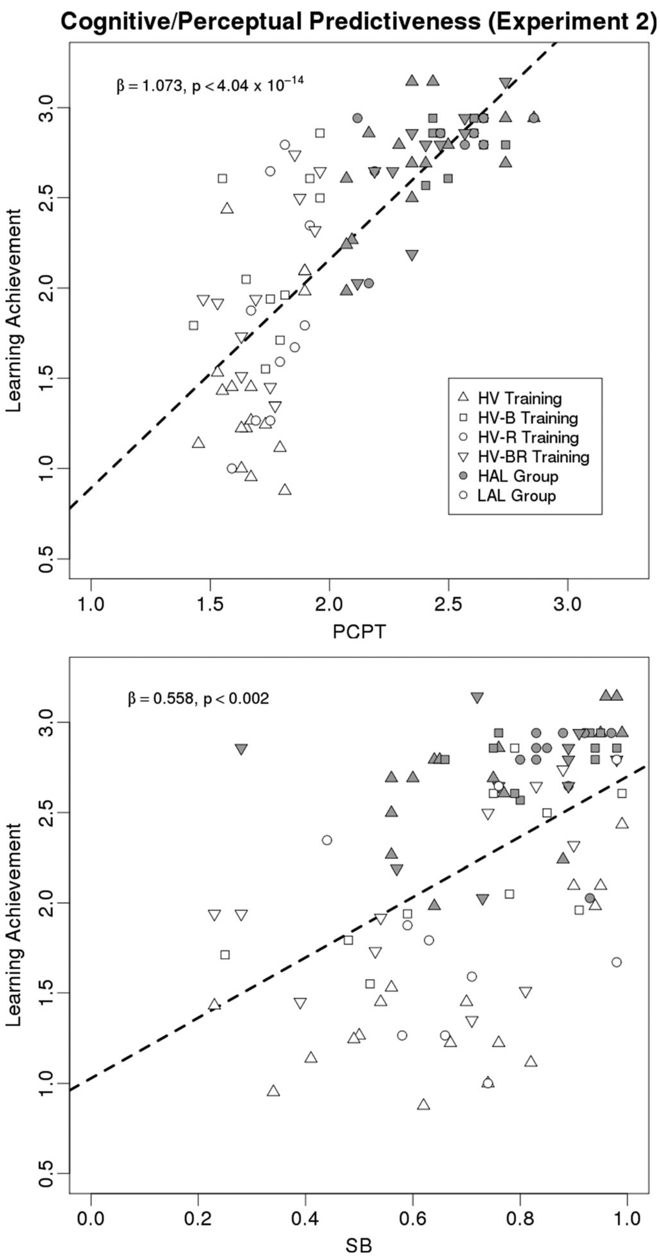

FIG. 8. Pretraining assessment measures differentially predict learning success. Perceptual ability (measures by PCPT) accounts for most of the variance in learning success, with some variance also explained by phonological awareness (SB). Measures of working memory (NR, AWM) did not contribute additional predictiveness.

4. Generalization

As in experiment 1, in each of the four conditions in experiment 2 we computed participants' generalization index "G" as the ratio of their accuracy on the novel talkers in the TLA to their accuracy on the trained talkers on the WIT from the last day of training. These values were submitted to a $2 \times 2 \times 2$ univariate ANOVA for Group (HAL vs LAL), Blocking (Blocked vs Not), Repeating (Repeated vs Not), and interactions. In experiment 2, we found no effects of Group ($p = 0.746$), Blocking ($p = 0.576$), Repeating ($p = 0.681$), or any two-way (all $p > 0.170$) or three-way ($p = 0.222$) interactions. These results indicate that all high-variability training designs (mixed, blocked, repeated, or blocked and repeated) were equally effective in facilitating generalization to stimuli produced by novel talkers ($G_{HV} = 1.02$; $G_{HV\text{-}B} = 1.03$; $G_{HV\text{-}R} = 1.01$; $G_{HV\text{-}BR} = 1.02$) for both the HAL and LAL groups. All four high-variability manipulations were also more effective at facilitating generalization than the LV design from experiment 1 [all $t > 3.6$, all Bonferroni-corrected $p < 0.0025$].

C. Discussion

The results of experiment 2 demonstrated an additional individual–instructional interaction: The stimulus-blocking revision to the training paradigm hypothesized to improve learning outcome did so only for participants with a certain learning-aptitude profile. Unlike in experiment 1, this difference in design did not work to the relative detriment of one group's learning. Indeed, if only one training paradigm from our study could be chosen for all learners to undergo, HV-B would result in the greatest overall learning, because perceptually strong learners benefit from globally high stimulus variability, while perceptually weak learners require low trial-by-trial variability. It is worth noting that HV-B training involved exactly the same amount of training (days of training, duration of paradigm, and number of tokens practiced) as the LV and HV conditions in experiment 1. Increased learning achievement following this training was thus specifically related to the presence of high variability globally (beneficial to the HAL group) and low variability locally (beneficial to the LAL group). These results evince the importance of considering individual–instructional interactions in designing training paradigms to maximize language-learning achievement. In experiment 1, we observed how high stimulus variability, a training-paradigm property typically thought to enhance learning achievement, was actually detrimental to learners with a specific pretraining aptitude profile. Here, we adapted the training paradigm to reduce the amount of trial-by-trial variability, removing the detrimental, extraneous processing demands of high stimulus variability locally while retaining the benefits of high variability globally, thus facilitating maximally successful learning outcomes across all participants. We likewise observed that, in all training paradigms involving high stimulus variability globally, participants of both high and low aptitudes exhibited similarly improved ability to generalize learning to novel stimuli was facilitated compared to LV training in experiment 1, suggesting the principal impediment to

Downloaded 10 Oct 2011 to 18.93.5.211. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

low-aptitude learners is the challenge of handling high trial-by-trial stimulus variability during initial learning.

Unlike experiment 1, the participant groups in experiment 2 did differ in their cognitive/perceptual assessment scores, with the HAL group outscoring the LAL group on SB, NR, and AWM. However, critically, there were no within-group differences across the various task manipulations or interactions between group and task design. Although the LAL group's lower cognitive and perceptual abilities may have contributed to their reduced learning relative to the HAL group, differences in the LAL group's learning achievement across the various learning conditions cannot be explained by differences in these cognitive/perceptual measures. Learners in the LAL group did not differ from one another in any cognitive or perceptual measure across training conditions (all $F < 1.4$, all $p > 0.265$). Thus, the enhanced learning achievement under the blocked designs is accounted for specifically by this experimental manipulation in task design rather than differences in learners' aptitudes.

IV. GENERAL DISCUSSION

The question of how individual learners' aptitudes interact with instructional paradigms has received only scant empirical investigation (Cronbach and Snow, 1977; Snow *et al.*, 1980) despite its being axiomatic among educators that different students learn best from different types of instruction. Here we have shown that there exist individual–instructional interactions in language-learning achievement, a result not previously described empirically. Individuals with weaker pitch perception abilities are disproportionately impaired in a high-variability training environment in which their perceptually stronger peers excel. Using preinstructional assessments to gain an indication of learning aptitude allows students to be assigned an optimal instructional regimen for their cognitive abilities. Moreover, this study shows that effectively translating ideas of individual differences into instructional practice is incumbent on the use of task-appropriate assessment metrics. Although constructs such as "phonological working memory," measured by tests like AWM and NR, have frequently been shown to predict achievement in other language-learning tasks (Cheung, 1996; Hu, 2003; Speciale *et al.*, 2004; Page and Norris, 2009), here we saw that a more domain-specific assessment (PCPT) was a better predictor of learning achievement for the new phonological contrast based on pitch. It remains the question of future work to identify what assessments may best predict the acquisition of unfamiliar phonological contrasts based on other features, such as voicing, duration, or manner of articulation.

These patterns of results are well described in the context of contemporary theories of perceptual learning. For example, Reverse-Hierarchy Theory (RHT; Ahissar and Hochstein, 2004; Ahissar *et al.*, 2009) stipulates that effective perceptual learning occurs when listeners gain access to the most informative perceptual level. Successfully identifying the correct perceptual level (here, pitch contour) requires consistent exposure to features that vary meaningfully with the appropriate behavioral response. High-variability environments such as HV or HV-R training obfuscate the most informative level because of considerable trial-by-trial variability in other, uninformative cues. Conversely, in low-variability environments (LV) and high-variability environments with low trial-by-trial variability (HV-B), features that do not vary across trials can be rejected, and attention can be directed to the features that covary predictably and meaningfully with correct behavioral responses. Moreover, RHT specifically predicts that improved access to the most informative level will result in superior generalization—precisely the effect we have seen here for individuals of both high and low perceptual abilities. While RHT provides parsimonious explanations of variability, learning, and generalization, understanding how its parameters accommodate individual differences requires further formal specification of this model. For example, individuals in the HAL group may already have had access to the most informative perceptual level [a proposition supported by recent neuroimaging studies of individual differences in speech learning (Wong *et al.*, 2007a)], thus reducing differences in performance between training paradigms seen for this group.

Previous work on differences in human learning has frequently appealed to the notion of different cognitive styles (e.g., Sternberg and Grigorenko, 1997). Such constructs are often predictive of achievement, but they have yet to be associated with underlying variability in the biological mechanisms that support learning. On the other hand, the present study focused on low-level domain-specific factors that constrained learning achievement, allowing us to better specify the mechanistic bases of individual variability in learning this phonological contrast. We identified pitch contour perception as the major predictor of learning achievement in our task; the biological mechanisms of this ability are increasingly well understood (Bendor and Wang, 2006; Ye *et al.*, 2010; Loui *et al.*, 2009; Plack *et al.*, 2005), including also its genetic (Drayna *et al.*, 2001) and environmental influences (Pantev *et al.*, 1998; Wong *et al.*, 2007b). By understanding the underlying biological sources of variability that result in differences in learning, we will be in a better position to design instructional programs that address those differences and maximize each individual's learning potential. This opportunity is even more pronounced given rapidly growing literature looking at nonbehavioral predictors of learning success, including functional and anatomical human neuroimaging (Díaz *et al.*, 2008; Mei *et al.*, 2008; Wong *et al.*, 2007a; Golestani *et al.*, 2002; Wong *et al.*, 2008) and genetics (Klein *et al.*, 2007). In particular, the *ASPM* and *MCPH1* genes identified by Dediu and Ladd (2007) to be associated with tone language could be good gene candidates for examining lexical tone perception and learning.

When learning a new skill, learning outcome is the result of both individual differences in underlying abilities and the design of the training paradigm. We have shown here that these factors are not independent in influencing language-learning success—there can be significant interactions between individual abilities and training design. These results have important bearing on educational policy,

Downloaded 10 Oct 2011 to 18.93.5.211. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

illustrating, in particular, the need for instructional paradigms to be empirically evaluated for their effects on students of a variety of abilities. For example, implementing an instructional paradigm based on classical perspectives from the second-language acquisition literature extolling the benefits of high stimulus-variability training (e.g., Lively *et al.*, 1993) would actually be detrimental to some students if the amount of trial-by-trial variability is also high. The idea of personalized medicine is revolutionizing that field by helping physicians determine which patients are likely to benefit from a given treatment (Eichelbaum *et al.*, 2006). Our results suggest the tractability of a similar "personalized" approach to second-language instruction, and education in general. Identifying the relevant cognitive, perceptual, and behavioral indices that can be measured before training begins will allow for the selection of an optimized course of instruction—either at the level of the individual, or by taking into account the relative needs of an entire classroom.

In sum, this study demonstrates how individual differences in pretraining measurements of speech- and language-learning aptitudes interact with the design of training paradigms to the benefit or detriment of learning outcome. In particular, while high stimulus variability training environments may typically promote learning among individuals with high learning aptitudes, such designs might actually be detrimental to those with low learning aptitudes. Having a strong and specific predictor of learning aptitude (here, pitch contour perception abilities) facilitates identifying the features of the training design that are detrimental to some learners (for example, the amount of trial-by-trial variability) and prescribing instructional paradigms that account for them. Taking individual differences in speech- or language-learning aptitude into consideration allows for the development of one or more training paradigm designs that will maximally benefit all learners.

ACKNOWLEDGMENTS

We thank Tasha Dees, Yoomin Ahn, and Geshri Gunasekera for their assistance. This work was supported by grants from the National Science Foundation (BCS-0719666) and National Institutes of Health (R01DC008333, R21DC007468, R03HD051827, and R21DC009652) awarded to P.C.M.W. T.K.P. is supported by a NSF Graduate Research Fellowship.

Ahissar, M., and Hochstein, S. (**2004**). "The reverse hierarchy theory of visual perceptual learning," Trends Cogn. Sci. **8**, 457–464.

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.

Barcroft, J., and Sommers, M. S. (**2005**). "Effects of acoustic variability on second language vocabulary learning," Stud. Second Lang. Acquis. **27**, 387–414.

Ben-Artzi, E., and Marks, L. E. (**1999**). "Processing linguistic and perceptual dimensions of speech: Interactions in speeded classification," J. Exp. Psychol. Hum. Percept. Perform. **25**, 579–595.

Bendor, D., and Wang, X. Q. (**2006**). "Cortical representations of pitch in monkeys and humans," Curr. Opin. Neurobiol. **16**, 391–399.

Boersma, P., and Weenink, D. (**2005**). "Praat: doing phonetics by computer" [Computer program]. http://www.praat.org/.

Bradlow, A. R., Nygaard, L. C., and Pisoni, D. B. (**1999**). "Effects of talker, rate, and amplitude variation on recognition memory for spoken words," Percept. Psychophys. **61**, 206–219.

Brooks, P. J., Kempe, V., and Sionov, A. (**2006**). "The role of learner and input variables in learning inflectional morphology," Appl. Psycholinguist. **27**, 185–209.

Chandrasekaran, B., Sampath, P., and Wong, P. C. M. (**2010**). "Individual variability in cue-weighting and lexical tone learning," J. Acoust. Soc. Am. **128**, 456–465.

Cheung, H. (**1996**). "Nonword span as a unique predictor of second-language vocabulary learning," Dev. Psychol. **32**, 869–873.

Clopper, C. G., and Pisoni, D. B. (**2004**). "Effects of talker variability on perceptual learning of dialects," Lang. Speech **47**, 207–239.

Creel, S. C., Aslin, R. N., and Tanenhaus, M. K. (**2008**). "Heeding the voice of experience: The role of talker variation in lexical access," Cognition **106**, 633–664.

Cronbach, L. J., and Snow, R. E. (**1977**). *Aptitudes and Instructional Methods* (Irvington, New York), pp. 1–592.

Dediu, D., and Ladd, D. R. (**2007**). "Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, *ASPM and Microcephalin*," Proc. Natl. Acad. Sci. U.S.A. **104**, 10944–10949.

Díaz, B., Baus, C., Escera, C., Costa, A., and Sebastián-Gallés, N. (**2008**). "Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language," Proc. Natl. Acad. Sci. U.S.A. **105**, 16083–16088.

Dörnyei, Z. (**2006**). "Individual differences in second language acquisition," AILA Rev. **19**, 42–68.

Drayna, D., Manichaikul, A., de Lange, M., Snieder, H., and Spector, T. (**2001**). "Genetic correlates of musical pitch recognition in humans," Science **291**, 1969–1972.

Ehrman, M. E., and Oxford, R. L. (**1995**). "Cognition plus: Correlates of language learning success," Mod. Lang. J. **79**, 67–89.

Eichelbaum, M., Ingelman-Sundberg, M., and Evans, W. E. (**2006**). "Pharmacogenomics and individualized drug therapy," Annu. Rev. Med. **54**, 119–137.

Flege, J. E. (**1995**). "Two procedures for training a novel second language phonetic contrast," Appl. Psycholinguist. **16**, 425–442.

Golestani, N., Paus, T., and Zatorre, R. J. (**2002**). "Anatomical correlates of learning novel speech sounds," Neuron **35**, 997–1010.

Golestani, N., and Zatorre, R. J. (**2009**). "Individual differences in the acquisition of second language phonology," Brain Lang. **109**, 55–67.

Green, K. P., Tomiak, G. R., and Kuhl, P. K. (**1997**). "The encoding of rate and talker information during phonetic perception," Percept. Psychophys. **59**, 675–692.

Hardison, D. M. (**2003**). "Acquisition of second-language speech: Effects of visual cues, context, and talker variability," Appl. Psycholinguist. **24**, 495–522.

Hu, C. F. (**2003**). "Phonological memory, phonological awareness, and foreign language word learning," Lang. Learn. **53**, 429–462.

Johnson, K., and Mullennix, J. W. (**1997**). *Talker Variability in Speech Processing* (Academic, San Diego), pp. 1–237.

Kaganovich, N., Francis, A. L., and Melara, R. D. (**2006**). "Electrophysiological evidence for early interaction between talker and linguistic information during speech perception," Brain Res. **1114**, 161–172.

Kingston, J. (**2003**). "Learning foreign vowels," Lang. Speech **46**, 295–349.

Klein, T. A., Neumann, J., Reuter, M., Hennig, J., von Cramon, D. Y., and Ullsperger, M. (**2007**). "Genetically determined differences in learning from errors," Science **318**, 1642–1645.

Kuncel, N. R., and Hezlett, S. A. (**2007**). "Standardized tests predict graduate students' success," Science **315**, 1080–1081.

Kuncel, N. R., Hezlett, S. A., and Ones, D. S. (**2004**). "Academic performance, career potential, creativity, and job performance: Can one construct predict them all?," J. Pers. Soc. Psychol. **86**, 148–161.

Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). "Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories," J. Acoust. Soc. Am. **94**, 1242–1255.

Loui, P., Alsop, D., and Schlaug, G. (**2009**). "Tone deafness: A new disconnection syndrome?," J. Neurosci. **29**, 10215–10220.

Magnuson, J. S., and Nusbaum, H. C. (**2007**). "Acoustic differences, listener expectations, and the perceptual accommodation of talker variability," J. Exp. Psychol. Hum. Percept. Perform. **33**, 391–409.

Majerus, S., Poncelet, M., Van der Linden, M., and Weekes, B. S. (**2008**). "Lexical learning in bilingual adults: The relative importance of short-term memory for serial order and phonological knowledge," Cognition **107**, 395–419.

Downloaded 10 Oct 2011 to 18.93.5.211. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

Mei, L., Chen, C., Xue, G., He, Q., Li, T., Xue, F., Yang, Q., and Dong, Q. (**2008**). "Neural predictors of auditory word learning," NeuroReport **19**, 215–219.
Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., and Besson, M. (**2009**). "Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity," Cereb. Cortex **19**, 712–723.
Mullennix, J. W., and Pisoni, D. B. (**1990**). "Stimulus variability and processing dependencies in speech perception," Percept. Psychophys. **47**, 379–390.
Norris, J. M., and Ortega, L. (**2001**). "Does type of instruction make a difference? Substantive findings from a meta-analytic review," Lang. Learn. **51**, 157–213.
Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (**1994**). "Speech perception as a talker-contingent process," Psychol. Sci. **5**, 42–46.
Paas, F. G. W. C., and Van Merriënboer, J. G. (**1994**). "Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach," J. Educ. Psychol. **86**, 122–133.
Page, M. P. A., and Norris, D. (**2009**). "A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms," Philos. Trans. R Soc. London, Ser. B **364**, 3737–3753.
Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., and Hoke, M. (**1998**). "Increased auditory cortical representation in musicians," Nature (London) **392**, 811–814.
Plack, C. J., Oxenham, A. J., Fay, R. R., and Popper, A. N. (**2005**). *Pitch: Neural Coding and Perception* (Springer, New York), pp. 1–364.
Robinson, P. (**2001**). "Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition," Second Lang. Res. **17**, 368–392.
Shih, C. L. (**1988**). "Tone and intonation in Mandarin," Working Papers of the Cornell Phonetics Laboratory **3**, 83–109.
Snow, R. E., Federico, P -A., and Montague, W. E. (**1980**). *Aptitude, Learning, and Instruction* (Erlbaum, Hillsdale, NJ), Vol. 1, pp. 1–368, Vol. 2, pp. 1–352.
Sommers, M. S., and Barcroft, J. (**2007**). "An integrated account of the effects of acoustic variability in first language and second language: Evidence from amplitude, fundamental frequency, and speaking rate variability," Appl. Psycholinguist. **28**, 231–249.
Sommers, M. S., Nygaard, L. C., and Pisoni, D. B. (**1994**). "Stimulus variability and spoken word recognition. I. Effects of speaking rate and overall amplitude," J. Acoust. Soc. Am. **96**, 1314–1324.
Song, J. H., Skoe, E., Wong, P. C. M., and Kraus, N. (**2008**). "Plasticity in the adult human auditory brainstem following short-term linguistic training," J. Cogn. Neurosci. **20**, 1892–1902.
Speciale, G., Ellis, N. C., and Bywater, T. (**2004**). "Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition," Appl. Psycholinguist. **25**, 293–320.
Sternberg, R. J., and Grigorenko, E. L. (**1997**). "Are cognitive styles still in style?," Am. Psychol. **52**, 700–712.
Studebaker, G. A. (**1985**). "A rationalized arcsine transform," J. Speech Hear. Res. **28**, 455–462.
Taylor, J., Roehrig, A. D., Hensler, B. S., Connor, C. M., and Schatschneider, C. (**2010**). "Teacher quality moderates the genetic effects on early reading," Science **328**, 512–514.
Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. (**1999**). "Training American listeners to perceive Mandarin tones," J. Acoust. Soc. Am. **106**, 3649–3658.
Wayland, R. P., and Li, B. (**2008**). "Effects of two training procedures in cross-language perception of tones," J. Phonetics **36**, 256–267.
Wong, P. C. M., Nusbaum, H. C., and Small, S. L. (**2004**). "Neural bases of talker normalization," J. Cogn. Neurosci. **16**, 1173–1184.
Wong, P. C. M., and Perrachione, T. K. (**2007**). "Learning pitch patterns in lexical identification by native English-speaking adults," Appl. Psycholinguist. **28**, 565–585.
Wong, P. C. M., Perrachione, T. K., and Parrish, T. B. (**2007a**). "Neural characteristics of successful and less successful speech and word learning in adults," Hum. Brain Mapp. **28**, 995–1006.
Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (**2007b**). "Musical experience shapes human brainstem encoding of linguistic pitch patterns," Nat. Neurosci. **10**, 420–422.
Wong, P. C. M., Warrier, C. M., Penhune, V. B., Roy, A. K., Sadehh, A., Parrish, T. B., and Zatorre, R. J. (**2008**). "Volume of left Heschl's gyrus and linguistic pitch learning," Cereb. Cortex **18**, 828–836.
Woodcock, R., McGrew, K., Schrank, F., and Mather, N. (**2007**). *Woodcock-Johnson III Normative Update* (Riverside, Rolling Meadows, IL).
Ye, C. Q., Poo, M. M., Dan, Y., and Zhang, X. H. (**2010**). "Synaptic mechanisms of direction selectivity in primary auditory cortex," J. Neurosci. **30**, 1861–1868.

Downloaded 10 Oct 2011 to 18.93.5.211. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp