Enhancing speech learning by combining task practice with periods of stimulus exposure without practice

Beverly A. Wright^{a)}

*Department of Communication Sciences and Disorders and Knowles Hearing Center,
Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208, USA*

Melissa M. Baese-Berk

Department of Linguistics, 217 Agate Hall, 1290, University of Oregon, Eugene, Oregon 97403, USA

Nicole Marrone

*Department of Speech, Language, and Hearing Sciences, University of Arizona, 1131 E. 2nd Street,
Tucson, Arizona 85721, USA*

Ann R. Bradlow

Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208, USA

(Received 13 September 2014; revised 17 June 2015; accepted 10 July 2015; published online 18 August 2015)

Language acquisition typically involves periods when the learner speaks and listens to the new language, and others when the learner is exposed to the language without consciously speaking or listening to it. Adaptation to variants of a native language occurs under similar conditions. Here, speech learning by adults was assessed following a training regimen that mimicked this common situation of language immersion without continuous active language processing. Experiment 1 focused on the acquisition of a novel phonetic category along the voice-onset-time continuum, while Experiment 2 focused on adaptation to foreign-accented speech. The critical training regimens of each experiment involved alternation between periods of practice with the task of phonetic classification (Experiment 1) or sentence recognition (Experiment 2) and periods of stimulus exposure without practice. These practice and exposure periods yielded little to no improvement separately, but alternation between them generated as much or more improvement as did practicing during every period. Practice appears to serve as a catalyst that enables stimulus exposures encountered both during and outside of the practice periods to contribute to quite distinct cases of speech learning. It follows that practice-plus-exposure combinations may tap a general learning mechanism that facilitates language acquisition and speech processing.

© 2015 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4927411>]

[CGC]

Pages: 928–937

I. INTRODUCTION

Language learners are not always focused on the language that they are trying to learn. Rather, adults acquiring a second language—and infants acquiring a first—typically alternate between times when they are actively speaking and listening to the new language and times when they are otherwise occupied but with the new language present in the background. For example, an adult who is attempting to learn a second language might spend a few minutes trying to understand a radio announcer who is speaking that language, but then return to preparing a meal with the radio still on. Similarly, native speakers who are adapting to the foreign accent of non-native speakers of their language are also likely to alternate between periods of trying to understand the accented speech and periods of background exposure to that speech. Here we investigated how a training regimen that mimicked this common situation of language immersion without continuous active language processing influenced

improvement in two well-studied, but quite distinct, cases of speech learning by adults: the acquisition of a novel phonetic category (a challenge for non-native listeners) and adaptation to foreign-accented speech (a challenge for native listeners). The training regimen was based on established procedures for successful learning in each case, but was modified such that it alternated between periods of practice with classification (phonetic-contrast task) or recognition (accent task) of the relevant speech stimuli and periods of stimulus exposure without classification or recognition practice. Thus, by examining two quite distinct cases of speech learning, we aimed to test whether a general mechanism may underlie various speech learning situations. We report that this practice-plus-exposure combination aided both forms of speech learning under controlled laboratory conditions, suggesting that its natural occurrence, or purposeful implementation in speech learning procedures, may facilitate language acquisition and speech processing more broadly.

The present investigation arose in part from evidence from fine-grained discrimination tasks that perceptual learning has at least two requirements. First, perceptual learning appears to require task performance. Learning typically does

^{a)}Electronic mail: b-wright@northwestern.edu

not arise from stimulus exposure alone, but rather requires performance of a perceptual task with relevant stimuli. The need for task performance is indicated by numerous reports that learning did not transfer from one perceptual task to another one, even when the same standard stimulus was used for both tasks (Ahissar and Hochstein, 1993; Crist *et al.*, 1997; Fahle, 1997; Gilbert and Sigman, 2007; Karni and Sagi, 1991; Levi and Polat, 1996; Meinhardt, 2002; Shiu and Pashler, 1992; Wright and Zhang, 2009a). It is also suggested by cases in which physiological changes associated with learning occurred due to training on the target task, but not from training on a different task with the same stimuli or from stimulus exposure alone (Ahissar *et al.*, 1992; Bao *et al.*, 2004; Blake *et al.*, 2006; Crist, *et al.*, 2001; Gao and Suga, 1998; Li *et al.*, 2004; Polley *et al.*, 2006; Recanzone *et al.*, 1992; Recanzone *et al.*, 1993; Stefan *et al.*, 2004). Second, perceptual learning appears to require sufficient practice per day, but no more than that. The retention of learning across days requires an adequate number of daily training trials (Aberg *et al.*, 2009; Wright and Sabin, 2007; Wright *et al.*, 2010). Critically, once this number is reached, additional training seems to be superfluous (Aberg *et al.*, 2009; Hauptmann and Karni, 2002; Hauptmann *et al.*, 2005; Ofen-Noy *et al.*, 2003; Ortiz and Wright, 2010; Wright and Sabin, 2007).

The combination of the needs for task performance and sufficient practice per day implies that perceptual learning requires task practice throughout training. However, another possibility is that learning requires sufficient stimulus exposure, only a portion of which needs to be encountered during task practice. We explored this alternative in two previous investigations in which we asked whether practice on the target task (the task to be learned) was needed throughout training, or whether a portion of the necessary practice could be replaced with stimulus exposure without target-task practice. In the first, we examined learning on a target task of auditory pure-tone frequency discrimination using five different multiple-day training regimens across different groups of adult listeners (Wright *et al.*, 2010). We began by documenting that training on the target frequency-discrimination task led to learning on that task. We then established that training on a non-target temporal-interval discrimination task that employed the same standard stimulus as the frequency task did not yield improvement on the frequency task. This outcome indicated that there was no learning transfer from the temporal-interval to the frequency task and that numerous exposures to the standard frequency from the frequency task were not alone sufficient to generate learning on that task. We next determined an amount of daily practice on the frequency task that did not lead to learning across sessions. Then, in the crucial case, in each training session we combined the amount of daily practice on the frequency task that did not yield learning on that task with practice on the temporal-interval task that also did not yield learning on the frequency task. The training therefore combined two experiences that were each ineffective on their own. Nevertheless, this regimen promoted as much learning as did practicing frequency discrimination throughout the entire training session. The outcome was the same when the practice trials on

the temporal-interval task from the combined regimen were replaced with an equal period of a written symbol-to-number matching task while the stimuli from the temporal-interval task were played in the background. Thus, a subset of the practice trials needed for learning on this frequency task could be replaced with relevant stimulus exposures linked to performance of a different auditory task or to no auditory task.

More recently we reported a similar result for a visual orientation-comparison task (Szpiro *et al.*, 2014). Observers participated in one of four different multiple-day training regimens. This time, we first established an amount of daily training on the target orientation task that did not yield improvement on that task. We then confirmed that the same amount of training on a non-target spatial-frequency comparison task that used the same standard stimulus as the orientation task also did not lead to learning on the orientation task. However, the combination of these two unsuccessful regimens yielded orientation learning. Further, practicing for the same total number of trials as in the combined regimen, but all on the orientation task did not result in improvement on that task, indicating that the learning in the combined regimen did not arise simply from the greater total number of trials in that regimen compared with the other regimens. Rather, in this instance, the combined regimen actually augmented learning on the target orientation task over training on that task alone. These results show that, as in the case of frequency-discrimination learning (Wright *et al.*, 2010), a subset of the practice trials needed for learning on this visual-orientation task could be replaced with relevant stimulus exposures encountered while performing a different visual task.

Here we asked whether the effectiveness of the practice-plus-exposure training regimen extends beyond the discrimination of simple sounds and visual stimuli to speech learning. To begin to address this question, we chose to examine two forms of speech learning as test cases: categorization of a non-native phonetic contrast and adaptation to foreign-accented speech. The theoretical motivation for examining these two cases was that they provided an opportunity to assess the effectiveness of the practice-plus-exposure training regimen on speech-learning tasks faced by different listener groups in different speech-learning situations. The selection of these test cases also yielded practical advantages. There already were training protocols with demonstrated effectiveness for each task that could be used as baseline regimens, the stimulus sets were readily created (phonetic-contrast task) or already existed (accent task), and baseline data were already available for one of the tasks (accent task). We assessed learning on the two tasks in separate experiments with different listeners in each trained group, thereby further evaluating the general efficacy of practice-plus-exposure training.

Demonstrations that the practice-plus-exposure regimen aids learning on two quite disparate cases would support the idea that multiple forms of speech learning and non-speech discrimination learning tap a common learning mechanism. Acquisition of non-native phonetic contrasts requires that listeners learn to associate a particular phonetic distinction that

does not carry linguistic information in the native language with contrasting category (phoneme) labels that are relevant for the non-native language. This is a challenge for non-native listeners, particularly when the speech-sound contrasts of the non-native language do not map transparently onto the contrasts of the native language. For example, native speakers of Japanese have trouble distinguishing between the English sounds /r/ and /l/ because the Japanese inventory of speech sounds does not include this contrasting pair. In contrast, adaptation to foreign-accented speech—understanding a non-native speaker of one’s native language—is a challenge for native listeners because it requires perceptual adjustment to speech production patterns that differ systematically along multiple acoustic dimensions from the native-accented norm. It is well established that, with practice, non-native listeners can acquire new phonetic categories (e.g., Bradlow *et al.*, 1999; Iverson *et al.*, 2005; McClaskey *et al.*, 1983; Wang *et al.*, 1999; and many others) and native listeners can become better at understanding foreign-accented speech (e.g., Adank *et al.*, 2010; Baese-Berk *et al.*, 2013; Bradlow and Bent, 2008; Clarke and Garrett, 2004; Sidaras *et al.*, 2009). These studies have advanced our understanding of the stimulus-, listener-, and talker-related conditions that promote speech learning, including the extent and limits of generalization of learning across items and talkers [for a review see Samuel and Kraljic (2009)]. Of interest in the present study is whether speech learning in non-native phonetic contrast categorization and adaptation to foreign-accented speech requires practice throughout the entire training period or whether a portion of the practice can be replaced by stimulus exposures without practice.

II. ACQUISITION OF A NOVEL PHONETIC CATEGORY

A. Methods

1. General protocol

Four separate groups of adult native speakers of English participated in this investigation. All listeners completed two 1-h sessions on consecutive days. Each session included a pre-training test, a training phase, and a post-training test. The pre- and post-training tests were identical both to each other and across the four listener groups, but the intervening training phase differed across the groups. Prior to the pre-training test in the first session, all listeners took part in a brief stimulus-familiarization phase.

2. Listeners

Thirty-one adults served as listeners (four groups of seven to eight listeners per group, three to eight females per group, and mean ages of 19–22 yr per group; see below for details). All listeners reported a monolingual, American-English language background, normal hearing, no history of a language or learning disability, and no previous experience with psychoacoustic tasks. Listeners received either an hourly wage or course credit for their participation.

3. Task and stimuli

The listener’s task was to classify the initial consonant of individual consonant-vowel syllables that varied in voice-onset-time (VOT) as belonging to one of three phonetic categories: less than ~ -25 ms (negative VOT) vs ~ -25 ms to $\sim +25$ ms (near-zero VOT) vs greater than $\sim +25$ ms (positive VOT). This three-way phonetic contrast in VOT is present in many languages, including Thai and Hindi, but English has only a two-way contrast between $\leq \sim +25$ ms (near-zero VOT) and greater than $\sim +25$ ms (positive VOT) in utterance initial CV syllables. However, a three-way contrast (by inclusion of the less than ~ -25 ms negative-VOT category into the system) can be acquired by native English speakers with practice (McClaskey *et al.*, 1983; Pisoni *et al.*, 1982).

On each trial, the listener was asked to categorize a single syllable selected randomly from a 14-step VOT continuum as having a negative, near-zero, or positive VOT. Following Tremblay *et al.* (1997), we labeled the negative VOT (non-native) as “mba,” the near-zero VOT (native) as “ba,” and the positive VOT (native) as “pa.” The label of “mba” is inaccurate phonetically, in that pre-voicing does not involve nasal air flow. Nevertheless, we adopted it because it provided a label for a non-native VOT that listeners could interpret easily. The three category labels were displayed in separate boxes on a computer screen and the listener used a computer mouse to select the box corresponding to the chosen category. The task was self-paced. The VOT continuum ranged from -60 ms to $+70$ ms in 10-ms steps. Feedback, when provided, was based on the division of the 14-step continuum into three categories as follows: four steps for “mba” (-60 to -30 ms), five steps for “ba” (-20 to $+20$ ms), and five steps for “pa” ($+30$ to $+70$ ms).

The stimuli were modified from tokens of “ba” and “pa” spoken by a female native speaker of American English, using Praat software (Boersma, 2001). They were presented to the left ear (by lab convention) via Sennheiser HD265 headphones at a comfortable listening level using a custom interface programmed in MATLAB. All listeners were tested in a double-walled, sound-attenuating booth.

4. Familiarization

At the beginning of the first session, listeners were given verbal instructions about the phonetic categorization task and samples of prototypical tokens for the three categories. These tokens had VOTs of -50 ms for “mba,” 0 ms for “ba,” and $+50$ ms for “pa” (see below). Each token was presented five times, in random order, with the button corresponding to the correct response for the current token indicated by a color change on the visual display. Listeners were instructed to attend to the relationship between the token and the corresponding button, but had no formal task requiring a response.

5. Pre- and post-training tests

Listeners were tested on their categorization ability before and after training each day to monitor performance over time (pre-test day 1, post-test day 1, pre-test day 2,

post-test day 2). In these pre- and post-training tests, listeners completed 120 trials of the categorization task. Each step from the VOT continuum was presented eight times except for the 0-ms VOT which was presented 16 times due to a stimulus-coding error. The syllables were presented in random order. No feedback on response accuracy was provided.

6. Training regimens

The training phase on each of two days was divided into four ~6 min blocks. During each block listeners practiced the categorization task for 60 trials (practice), performed a written symbol-to-number matching task while 60 syllables were presented in the background (exposure), or performed the written matching task in quiet (silence). For the matching task, listeners were given a symbol-to-number decoding key and asked to convert a string of symbols into a sequence of numbers by matching the symbols in the string to the numbers in the key. The task was self-paced. During the exposure blocks, a silent period of ~5 s followed each syllable presentation, mirroring the average time taken for listeners to provide a response on the categorization task. Feedback on response accuracy for the categorization task was given after each trial via a visual display (“correct” or “wrong”) during the practice periods.

The four training regimens provided different balances of practice and exposure. To document the amount of learning obtained from continuous practice, in the **Practice-Only** regimen, the four blocks in each session were all practice blocks (240 practice trials per session) ($n=7$ listeners, 3 females, mean age 22 yr, standard deviation: 3.8). To ascertain the influence of exposure alone, in the **Exposure-Only** regimen, the four blocks in each session were all exposure blocks (240 exposures per session) ($n=8$ listeners, 4 females, mean age 21 yr, s.d.: 1.3). To determine the effect of replacing half of the practice trials with exposure alone, in the **Practice + Exposure** regimen, the four blocks in each session alternated between practice and exposure blocks, beginning with practice (120 practice trials + 120 exposures per session) ($n=8$ listeners, 8 females, mean age 19 yr, s.d.: 1). Finally, to establish how much learning in the Practice + Exposure regimen could be attributed to the practice portions of that regimen, in the **Practice + Silence** regimen, the four blocks in each session alternated between practice and silence blocks, beginning with practice (120 practice trials per session) ($n=8$ listeners, 6 females, mean age 20 yr, s.d.: 1.1).

Throughout the training phase (during both practice and exposure blocks), the syllables to be presented were drawn randomly from a trimodal distribution along the VOT continuum. The distribution peaks were centered on prototypical VOTs for each category: -50 ms for “mba,” 0 ms for “ba,” and +50 ms for “pa.” Thus, the presentation probability was higher for prototypical than for atypical tokens [see Hayes-Harb (2007) and Maye *et al.* (2002) for a similar approach to stimulus distribution].

7. Analyses

The measure of interest was the slope of the non-native category boundary between “mba” and “ba.” To compute

this slope, we fit the categorization data with a probit function, separately for each listener and each test. The slope of the function was scaled to range from a theoretical minimum of 0.0, representing an abrupt shift from 100% to 0% “mba” identifications as the VOT became less negative, to a maximum of 1.0, representing an equally abrupt but phonetically inappropriate shift from 0% to 100% “mba” responses. Values at 0.5 therefore represent a flat slope (no true boundary) (MacKain *et al.*, 1981). There was a positive correlation between the slopes of these functions at the pre-training test on day 1 and the post-training test on day 2 amongst the entire data set ($r^2=0.36$, $p=0.0003$). Therefore, prior to statistical analysis, each slope was adjusted to take into account differences in performance at pre-training test 1, using the procedure that underlies analysis of covariance [see Sabin *et al.* (2013) for a detailed example from perceptual learning]. Improvement within each of the four groups was assessed using four separate one-way repeated-measures ANOVAs on the adjusted slopes across the pre- and post-training tests on each of the two days. Performance across the four groups was evaluated using pairwise comparisons on the adjusted slopes of post-training test 2, taking into account multiple comparisons (Hochberg, 1988). These comparisons followed a significant one-way analysis of variance (ANOVA) across groups on those adjusted slopes ($F_{3,27}=11.07$, $p<0.0001$, $\eta^2=0.55$).

III. RESULTS

At the end of training [Fig. 1(A), day 2 post], the slope of the non-native category boundary between “mba” and “ba,” adjusted for individual differences in initial slope, was steeper for the Practice + Exposure group (filled circles; $n=8$) than for each of the other three groups: Exposure-Only (open triangles; $n=8$) ($p<0.0001$, $d=2.83$), Practice + Silence (open squares; $n=8$) ($p=0.013$, $d=1.51$), and Practice-Only (open diamonds; $n=7$) ($p=0.019$, $d=1.49$). This category-boundary slope did not differ between the Practice + Silence and Practice-Only groups ($p=0.913$, $d=0.06$), but was steeper for these two groups than for the Exposure-Only group (both $p=0.044$, both $d\geq 1.31$).¹ The final post-training adjusted category-boundary slopes mark the end point of significant four-test learning curves for all but the Exposure-Only group (Practice + Exposure: $F_{1,29}=51.24$, $p<0.0001$, $\eta^2_p=0.64$; Practice + Silence: $F_{1,29}=17.59$, $p=0.0002$, $\eta^2_p=0.38$; Practice-Only: $F_{1,25}=6.00$, $p=0.021$, $\eta^2_p=0.19$; Exposure-Only: $F_{1,29}=0.023$, $p=0.879$, $\eta^2_p=0.01$). The learning emerged during the first session for the Practice-Only group [paired t -test; $t(6)=3.52$, $p=0.013$, $d_z=1.33$], but not until the second session for the Practice + Silence group [first session: $t(7)=1.12$, $p=0.301$, $d_z=0.40$].

The same key patterns hold at the individual level across a wide range of raw starting values. For any given initial raw slope of the “mba”-“ba” category boundary [Fig. 1(B), day 1 pre-training slope, x axis], the final slope (day 2 post-training slope, y axis) was nearly always steeper (points are farther below the diagonal) for the Practice + Exposure group (filled circles) than for each of the other groups (open symbols). The

FIG. 1. Acquisition of a non-native category boundary. (A) Mean performance across tests. Mean slope of the non-native category boundary between “mba” and “ba” for native speakers of English across daily pre- and post-training tests for each of four two-session training regimens: **Exposure-Only** (open triangles; $n=8$), **Practice + Silence** (open squares; $n=8$), **Practice-Only** (open diamonds; $n=7$), and **Practice + Exposure** (filled circles; $n=8$). Each box in the schematics represents 60 trials of practice (black boxes), 60 trials of stimulus exposure without practice (gray boxes), or the time span of 60 trials spent in silence (white boxes). The slope of the function is scaled such that the closer the value is to zero, the sharper the category boundary. Error bars indicate $\pm$ one standard error of the mean. Slopes are adjusted to account for individual differences in the slope at pre-training test 1. (B) Individual performance. Raw scaled category boundary slopes between “mba” and “ba” from the day 1 pre-training test (abscissa) and day 2 post-training test (ordinate) for each individual listener. Performance of the Practice + Exposure group (filled circles) is shown with that of the Exposure-Only (left panel; open triangles), Practice + Silence (middle panel; open squares), and Practice-Only (right panel; open diamonds) groups. The key result is that a combination of practice categorizing speech tokens with additional exposures to those tokens without practice (Practice + Exposure) facilitated the acquisition of a non-native category boundary. Learning with this combination was enhanced relative to when the training involved only brief (Practice + Silence) or longer (Practice-Only) periods of categorization practice, even though mere exposure to the speech tokens (Exposure-Only) yielded no learning.

final raw slope was also frequently steeper than the initial slope for the Practice + Silence and Practice-Only groups (points are below the diagonal), but was similar at both tests for the Exposure-Only group (points are near the diagonal). Thus, alternating between brief periods of practice categorizing speech tokens along a VOT continuum and periods of additional exposures to those tokens without active categorization practice (Practice + Exposure) enhanced the acquisition of a non-native category boundary. This enhancement arose from a beneficial interaction between the practice and exposure trials because (1) the combination yielded more learning than only the practice portions of the combination (Practice + Silence) and more learning than the same total number of trials of practice alone (Practice-Only), and (2) mere exposure to the same total number of tokens (Exposure-Only) yielded no learning at all.

IV. ADAPTATION TO FOREIGN-ACCENTED SPEECH

A. Methods

1. General protocol

The results from five separate groups of adult native speakers of English are reported here. All listeners

completed two 1-h training sessions on consecutive days followed by a post-training test that was administered on the second day immediately following the second training session. The training sessions differed across the trained groups.

2. Listeners

The data from a total of 50 listeners are described (five groups of 10 listeners per group, six to nine females per group, and mean ages of 19–26 yr per group; see below for details). The results from the Practice-Only ($n=10$) and No-Foreign-Accent-Practice-Only-Control ($n=10$) groups (see below) were reported previously (Baese-Berk *et al.*, 2013; Bradlow and Bent, 2008; their multi-talker and task-control groups) and are included as points of comparison for the data from the newer groups. All listeners were native speakers of American English, recruited from the Northwestern University community, who reported normal speech and hearing abilities and limited experience with foreign-accented speech. None had parents, grandparents, or roommates who were non-native English speakers. Listeners received either an hourly wage or course credit for their participation.

3. Task and stimuli

Except where noted, the listener's task was to transcribe English sentences spoken by multiple Mandarin-accented talkers into standard English orthography. On each trial, the listener was presented with the production of a single, grammatical, Mandarin-accented English sentence, once, and asked to write it down on a specially prepared answer sheet. The sentence presentation rate was self-paced.

The foreign-accented sentence recordings came from the Northwestern University Foreign-Accented English Speech Database (NUFAESD) (Bent and Bradlow, 2003; Bradlow and Bent, 2008), which includes four Bamford-Kowal-Bench (BKB) (Bamford and Wilson, 1979; Bench *et al.*, 1979) sentence lists recorded by a total of 32 non-native speakers of English from various native language backgrounds. Each list contains 16 simple declarative sentences and each sentence contains three or four keywords, for a total of 50 keywords per list. NUFAESD recordings from a total of six male Mandarin-accented English talkers were used in the present study, five as training talkers and one as a test talker. The five Mandarin-accented training talkers had overall sentence intelligibility scores that were in the mid range as assessed by native English listeners during the development of the NUFAESD. The test talker (a sixth Mandarin-accented talker from the NUFAESD) was not included in any training set. The talkers and sentences used during training and in the post-training test were the same as those used by Baese-Berk *et al.* (2013) and Bradlow and Bent (2008). All sentences were presented in white noise at a signal-to-noise ratio of +5 dB. The noise began 500 ms before and ended 500 ms after each sentence presentation.

Sentence presentation was controlled by experiment running software (Superlab Pro 2.01). The sentences were played through a computer soundcard (SoundBlaster Live) and were delivered through both earpieces of Sennheiser HD 580 headphones at a comfortable listening level. All testing occurred in a sound-attenuating booth.

4. Training regimens

The training phase on each of two days was divided into ten ~1.5 min blocks. During each block listeners either performed the sentence-transcription task for eight trials (practice), performed the written symbol-to-number matching task described above while eight sentences were presented in the background (exposure), or performed the written matching task in silence (silence). During the exposure blocks, a silent period of 7 s followed each audio sentence presentation, mirroring the average time taken for listeners to transcribe the sentences. No feedback on response accuracy was provided during the practice periods.

Four of the five training regimens provided different balances of practice and exposure to Mandarin-accented English sentences. These regimens followed the same rationale and design as in the phonetic-contrast experiment. To document the amount of learning obtained from continuous practice, in the **Practice-Only** regimen, the 10 blocks in each session were all practice blocks (80 practice trials per

session) ($n = 10$ listeners, 8 females, mean age 19 yr, standard deviation: 1). To ascertain the influence of exposure alone, in the **Exposure-Only** regimen, the 10 blocks in each session were all exposure blocks (80 exposures per session) ($n = 10$ listeners, 6 females, mean age 20 yr, s.d.: 1.1). To determine the effect of replacing half of the practice trials with exposure alone, in the **Practice + Exposure** regimen, the 10 blocks in each session alternated between practice and exposure blocks, beginning with practice (40 practice trials + 40 exposures per session) ($n = 10$ listeners, 9 females, mean age 19 yr, s.d.: 0.5). Finally, to establish how much learning in the Practice + Exposure regimen could be attributed to the practice portions of that regimen, in the **Practice + Silence** regimen, the 10 blocks in each session alternated between practice and silence blocks, beginning with practice (40 practice trials per session) ($n = 10$ listeners, 8 females, mean age 19 yr, s.d.: 0.9). The fifth regimen served as a task-control baseline. In that **No-Foreign-Accent-Practice-Only-Control** regimen, the 10 blocks in each session were all practice blocks, but the sentences to be transcribed were spoken by native American-English talkers rather than Mandarin-accented talkers (80 trials per session) ($n = 10$ listeners, 6 females, mean age 26 yr, s.d.: 9). Bradlow and Bent (2008) reported that this task-control regimen yielded better performance at the post-training test than did no training at all, making the task-control regimen the stricter baseline measure of the two.

The audio track used for each session for the Exposure-Only, Practice-Only, and Practice + Exposure regimens consisted of five repetitions of the same set of 16 English sentences, with each repetition produced by a different Mandarin-accented English talker (five total talkers). The sentences were grouped by talker, and the order of sentences within each group was the same. Different sets of 16 sentences, spoken by the same five talkers, were presented in the first and second training sessions, yielding a grand total of 160 sentence presentations [5 repetitions (1 for each talker) $\times$ 16 sentences (2 consecutive blocks of 8 sentences each) $\times$ 2 sessions]. Only the first eight of the 16 sentences of each repetition set were presented for the Practice + Silence regimen, and each repetition was produced by a different native American-English talker (five total talkers who were not part of the original NUFAESD) for the No-Foreign-Accent-Practice-Only-Control regimen.

5. Post-training test

The post-training test was administered at the end of the second session and was identical for all groups. All listeners transcribed a novel set of 16 English sentences produced by a novel Mandarin-accented talker. No feedback on response accuracy was provided. The three or four keywords in each sentence were scored as either correct or incorrect, yielding a recognition-accuracy score (proportion of keywords correctly recognized) for each sentence. We did not include a pre-training test in this experiment, following prior work. Adaptation to foreign accents can occur quite quickly and so could take place largely during the trials required for a pre-training test. This design also allowed us to compare the

present results directly with those of Bradlow and Bent (2008) and Baese-Berk *et al.* (2013).

6. Analyses

Performance across the five groups was evaluated using pairwise comparisons on the recognition-accuracy score on the post-training test as the dependent variable, taking into account multiple comparisons (Hochberg, 1988). These comparisons followed a significant one-way ANOVA across groups on those recognition-accuracy scores ($F_{4,45} = 8.11$, $p < 0.0001$, $\eta^2 = 0.42$).

V. RESULTS

The Practice + Exposure group ($n = 10$) performed better on the post-training test than the Exposure-Only ($n = 10$; $p = 0.005$, $d = 2.14$), Practice + Silence ($n = 10$; $p = 0.011$, $d = 2.04$), and No-Foreign-Accent-Practice-Only-Control ($n = 10$; $p = 0.001$, $d = 1.87$) groups (Fig. 2). Post-training performance did not differ between the Exposure-Only and Practice + Silence groups ($p = 1.00$, $d = -0.18$) or between those groups and the No-Foreign-Accent-Practice-Only-Control group (both $p \geq 0.985$, $d \leq 0.05$). Further, the Practice + Exposure group performed as well as the Practice-Only group ($n = 10$; $p = 1.00$, $d = -0.22$). The Practice-Only group performed better than the other three groups (Exposure-Only: $p = 0.022$, $d = 1.36$; Practice + Silence: $p = 0.051$, $d = 1.25$; No-Foreign-Accent-Practice-Only-Control: $p = 0.004$, $d = 1.36$), matching the pattern of the Practice + Exposure group.² Thus, neither mere exposure to foreign-accented sentences in noise nor brief periods of transcription of those sentences aided speech recognition of sentences with that accent, but the combination of these elements helped (Practice + Exposure), and did so as much as transcribing all of the sentences (Practice-Only).

VI. DISCUSSION

A. Effectiveness of the practice-plus-exposure regimen

The present results demonstrate that speech learning can be induced by combining periods of practice with a speech-processing task and periods of stimulus exposure without task practice. This combination helped different groups of adult native speakers of English to (1) acquire a non-native phonetic contrast (go from a two-way to a three-way voicing distinction) and (2) adapt to foreign-accented speech (Mandarin-accented English). Neither the practice portions of the practice-plus-exposure regimen, nor stimulus exposure without practice, yielded much improvement separately. Yet, alternation between brief periods of practice and stimulus-exposure alone generated as much or more improvement as did continuous practice. Thus, the practice-plus-exposure regimen markedly increased the benefit of task practice due to a helpful interaction between two experiences that were otherwise largely ineffective.

There is now evidence that the practice-plus-exposure regimen can enhance learning on auditory frequency discrimination (Wright *et al.*, 2010), visual orientation

FIG. 2. Adaptation to foreign-accented speech. Percent-correct performance for key words transcribed from spoken Mandarin-accented English sentences by native speakers of English following each of five two-session training regimens: **Practice-Only-No-Foreign-Accent-Control**, **Exposure-Only**, **Practice + Silence**, **Practice + Exposure**, and **Practice-Only** ($n = 10$ per group). Results from the sole post-training test on day 2 are shown for each individual listener (open circles) as well as for the mean of each group (horizontal lines). Error bars indicate $\pm$ one standard error of the mean. The two groups marked with asterisks performed similarly to each other, but significantly better than the other three groups (* $p \leq 0.04$; ** $p \leq 0.01$). The key result is that a combination of practice transcribing foreign-accented sentences combined with additional exposures to those sentences (Practice + Exposure) aided speech recognition of sentences with that accent. Learning with this combination was enhanced relative to when the training involved only brief (Practice + Silence) periods of sentence transcription or mere exposure to foreign-accented sentences (Exposure-Only), and yielded as much improvement as did transcribing all of the sentences (Practice-Only).

comparison (Szpiro *et al.*, 2014), non-native phonetic categorization (here), and foreign-accented sentence recognition (here). These cases cover a wide range of task and stimulus complexity. Improvement on pure-tone frequency discrimination requires that increasingly fine-grained distinctions be made between simple sounds that vary along a single acoustic dimension. The learning of specific non-native phonetic contrasts instead entails adjustment of existing phonetic categories for complex stimuli along a well-defined and limited set of acoustic dimensions. Foreign accents result from a systematic and complex interaction between the entire phonetic and phonological systems of the first (L1) and second (L2) languages of bilingual talkers. Therefore, listener adaptation to foreign-accented speech involves simultaneous re-mapping along many acoustic dimensions, especially in the case of sentence-length stimuli, as used here, where L1–L2 interactions at many time-scales (sub-phonemic, segmental, and supra-segmental) have an opportunity to manifest themselves in the speech of the foreign-accented talker. These examples thus encompass learning on a basic perceptual skill as well as on two speech tasks that involve fundamentally different types of change in speech processing. The acquisition of a non-native phonetic contrast requires changes in the perception of a non-native language and hence modifications in L2 listening. Adaptation to foreign-accented speech requires a broadening of acceptable pronunciations in a native language, and hence modifications in L1 listening. These instances further comprise manipulated and natural speech, single and multiple-talker training, the presence and

absence of feedback, and testing before and after as well as only after training. The effectiveness of the practice-plus-exposure regimen across these varied cases implies that this regimen taps a general learning mechanism, and therefore that it will aid learning on other tasks as well.

B. Learning framework

The present investigation arose from indications that learning on fine-grained discrimination tasks requires task practice for a sufficient number of trials per day, a portion of which can be replaced with stimulus exposure alone (see Introduction). These behavioral requirements suggest corresponding neural requirements for learning. The need for task practice leads to the idea that task performance releases internal permissive signals, such as attention and reward, which place the neural circuitry engaged by the task in a sensitized state in which it can be modified. (Ahissar and Hochstein, 2004; Gilbert and Sigman, 2007; Seitz and Dinse, 2007). The requirement for sufficient practice suggests that the sensitized neural circuitry must be stimulated sufficiently while it is in that sensitized state (Wright and Sabin, 2007). The implication is that changes to the sensitized neural circuitry become more permanent (in these cases, last across days) only if some physiological threshold is reached through adequate stimulus exposure (Seitz and Dinse, 2007; Wright and Zhang, 2009b). Finally, the capacity to replace a portion of the practice trials with stimulus exposure alone implies that learning can be supported by stimulation that occurs outside of the time of the direct activation of the internal permissive signals (Wright *et al.*, 2010).

The current results fit within this framework. First, there was improvement on both speech tasks with continuous practice on those tasks (Practice-Only), but not without it (Exposure-Only). This outcome is consistent with the idea that practice provided the internal permissive signals needed to sensitize the neural processes to be modified. The need for practice observed here (also see Tremblay *et al.*, 2010) separates these results from previous reports of improvement on speech-perception skills arising from stimulus exposure alone (Maye *et al.*, 2002; Saffran *et al.*, 1996). Note, though, that those improvements in the absence of task performance may still have depended on the posited internal permissive signals, given evidence that such improvements appear to be contingent on attention (Toro *et al.*, 2005).

Second, there was learning when the relevant task was practiced throughout each training session (Practice-Only) but not when it was practiced for half of each session (Practice + Silence). This pattern was evident at the single post-training test for the accent task and during the first session for the phonetic-contrast task. These data are consistent with the idea that there must be sufficient stimulation of the sensitized neural process for behaviorally observable learning to arise. Note that the proposed requirement is only that sufficient stimulation must occur in connection with the sensitized state. There are no restrictions on how that state is induced for any given stimulus exposure, whether directly by task performance (during practice) or by a non-simultaneous influence of task performance (potentially

encountered during stimulus-exposure alone), as described below. In this overall context, we acknowledge that by the end of the second session the two regimens (Practice-Only and Practice + Silence) yielded similar improvement for the phonetic-contrast task. This result on its own raises the possibility that each practice trial contributed equally to improvement on this task until learning saturated (at 240 total trials). However, we have additional preliminary data for this task that are consistent with the current proposal, and that suggest that the breaks that were interleaved with the shorter practice periods affected the learning time course.

The relationship between the amount of daily training and the magnitude of learning on speech-perception tasks has clear practical and theoretical implications. For both auditory and visual fine-grained discrimination tasks, there is evidence that improvement across days requires sufficient (and sometimes extensive) training per day (Aberg *et al.*, 2009; Wright and Sabin, 2007; Wright *et al.*, 2010), and that additional daily training beyond the sufficient amount provides no further benefit (Aberg *et al.*, 2009; Hauptmann and Karni, 2002; Hauptmann *et al.*, 2005; Wright and Sabin, 2007). In combination, these results suggest that the consolidation of perceptual learning may function as an all-or-none process, a characteristic that could be exploited to help optimize perceptual training regimens (Wright and Sabin, 2007). With the present data also indicating an apparent requirement for sufficient daily practice for speech learning, it could prove fruitful to test the corresponding prediction that there will be an amount of daily practice beyond which additional practice provides no further benefit.

Finally, there was more learning when periods of stimulus exposure without practice were interleaved with periods of practice (Practice + Exposure) than when either element was provided alone (Exposure-Only or Practice + Silence). This result is consistent with the idea that stimulation can aid improvement even when it does not occur simultaneously with activation of the internal permissive signals. Thus, given sufficient stimulus exposures, only a subset of those exposures need to be encountered during task-relevant practice to aid at least some speech-perception skills. For the phonetic-contrast task, the Practice + Exposure regimen actually ultimately generated greater improvement than did practice throughout each session (Practice-Only). This outcome indicates that the additional stimulus exposures were even more effective than extra practice trials on this task. We previously reported a similar result for learning on a visual orientation-comparison task (Szpiro *et al.*, 2014).

C. Connection to natural language acquisition

The present demonstration that the practice-plus-exposure regimen can enhance learning on speech-perception tasks could help account for natural language acquisition in adults and infants. For instance, learning a second language is thought to be facilitated by spending time in environments in which that language dominates. This idea is implied by the proliferation of language-immersion programs (for a review, see Norris and Ortega, 2000), and recently has received experimental support through a direct comparison

between immersed and non-immersed learners (Reuland *et al.*, 2012). Many factors are likely to contribute to this benefit, including, for example, the possibility that language immersion leads to greater inhibition of the native language while speaking the non-native one (Linck *et al.*, 2009). The present results suggest that another such factor may be that during language immersion the learner alternates between periods of language practice and periods of language exposure without practice and thus may take advantage of the learning processes that thrive in such situations.

Similarly, speech learning in infants appears to be aided by social interactions (Goldstein *et al.*, 2003; Kuhl *et al.*, 2003). Assuming that these interactions “gate” learning, as has been suggested (Goldstein *et al.*, 2010; Kuhl, 2007), they could serve the same function as direct practice in adults, providing the internal permissive signals for learning. If so, and if as in adults, stimulus exposures that occur outside of the time of the activation of these internal permissive signals can contribute to learning in infants, then social interaction with an infant that is focused on speech processing, combined with speech exposure separate from that interaction, would facilitate infant language acquisition.

D. Implications

On a practical level, implementations of the practice-plus-exposure regimen could increase the effectiveness of training while reducing the amount of direct practice required for speech learning in applied settings. The regimen could be applied in formal speech training to aid aspects of language acquisition (a change in the perception of L2) as well as to broaden the array of acceptable pronunciations of a native language (a change in the perception of L1). The learning processes the regimen invokes also potentially could be tapped naturally through language immersion. If so, immersion in a non-native language environment should aid the acquisition of that non-native language, as appears to be the case (Reuland *et al.*, 2012), and immersion in a multilingual environment with one common language should aid the ability of native speakers of the common language to understand foreign-accented speech.

More broadly, the effectiveness of the practice-plus-exposure regimen suggests one potential solution to the stability/plasticity dilemma (Grossberg, 1987; Abraham and Robins, 2005; Samuel and Kraljic, 2009). The nervous system must have rules for when it maintains its current state (stability) and when it implements changes to that state in response to demands (plasticity). The current results imply that intentional practice in adults serves as a catalyst that enables change and that also marks the type of stimulation that can contribute to that change both during and outside of the catalytic periods. This combination would control what is and is not learned, while simultaneously reducing the amount of intentional practice required to learn the vetted material.

ACKNOWLEDGMENTS

We thank Jessica Conderman, David Little, Matt Goldrick, Andy Sabin, and Yuxuan Zhang for technical assistance, help with data analysis, and useful discussions.

This work was supported by NIH/NIDCD grants R01-DC004453 (BAW) and R01-DC005794 (A.R.B.) as well as a Northwestern University Cognitive Science Graduate Fellowships for Interdisciplinary Research Projects (M.M.B.B.).

¹We also analyzed the results using the Bonferroni, rather than the Hochberg, correction for multiple comparisons. Both corrections help protect against false positives (type I errors), but the possibility of a false negative (type II error) is higher with the Bonferroni than the Hochberg method (e.g., Farcomeni, 2008). Use of the Bonferroni correction did not affect any of the statistical conclusions involving the Practice + Exposure group. However, at the end of training, the non-native category-boundary slope was steeper for the Practice + Silence and Practice-Only groups than for the Exposure-Only group with the Hochberg correction (both $p = 0.044$), but not with the Bonferroni correction (both $p \geq 0.137$).

²Use of the Bonferroni rather than the Hochberg correction did not affect any of the statistical conclusions.

- Aberg, K. C., Tartaglia, E. M., and Herzog, M. H. (2009). “Perceptual learning with chevrons requires a minimal number of trials, transfers to untrained directions, but does not require sleep,” *Vision Res.* **49**(16), 2087–2094.
- Abraham, W. C., and Robins, A. (2005). “Memory retention—the synaptic stability versus plasticity dilemma,” *Trends Neurosci.* **28**(2), 73–78.
- Adank, P., Hagoort, P., and Bekkering, H. (2010). “Imitation improves language comprehension,” *Psychol. Sci.* **21**(12), 1903–1909.
- Ahissar, E., Vaadia, E., Ahissar, M., Bergman, H., Arieli, A., and Abeles, M. (1992). “Dependence of cortical plasticity on correlated activity of single neurons and on behavioral context,” *Science* **257**(5075), 1412–1415.
- Ahissar, M., and Hochstein, S. (1993). “Attentional control of early perceptual learning,” *Proc. Natl. Acad. Sci. U.S.A.* **90**(12), 5718–5722.
- Ahissar, M., and Hochstein, S. (2004). “The reverse hierarchy theory of visual perceptual learning,” *Trends Cogn. Sci.* **8**(10), 457–464.
- Baese-Berk, M. M., Bradlow, A. R., and Wright, B. A. (2013). “Accent-independent adaptation to foreign accented speech,” *J. Acoust. Soc. Am.* **133**(3), EL174–EL180.
- Bamford, J., and Wilson, I. (1979). “Methodological considerations and practical aspects of the BKB sentence lists,” in *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children* (Academic, London), pp. 148–187.
- Bao, S., Chang, E. F., Woods, J., and Merzenich, M. M. (2004). “Temporal plasticity in the primary auditory cortex induced by operant perceptual learning,” *Nat. Neurosci.* **7**(9), 974–981.
- Bench, J., Kowal, Å., and Bamford, J. (1979). “The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children,” *Br. J. Audiol.* **13**(3), 108–112.
- Bent, T., and Bradlow, A. R. (2003). “The interlanguage speech intelligibility benefit,” *J. Acoust. Soc. Am.* **114**(3), 1600–1610.
- Blake, D. T., Heiser, M. A., Caywood, M., and Merzenich, M. M. (2006). “Experience-dependent adult cortical plasticity requires cognitive association between sensation and reward,” *Neuron* **52**(2), 371–381.
- Boersma, P. (2001). “Praat. A system for doing phonetics by computer,” *Glott Int.* **5**(9/10), 341–345.
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., and Tohkura, Y. (1999). “Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production,” *Percept. Psychophys.* **61**(5), 977–985.
- Bradlow, A. R., and Bent, T. (2008). “Perceptual adaptation to non-native speech,” *Cognition* **106**(2), 707–729.
- Clarke, C. M., and Garrett, M. F. (2004). “Rapid adaptation to foreign-accented English,” *J. Acoust. Soc. Am.* **116**(6), 3647–3658.
- Crist, R. E., Kapadia, M. K., Westheimer, G., and Gilbert, C. D. (1997). “Perceptual learning of spatial localization: Specificity for orientation, position, and context,” *J. Neurophysiol.* **78**(6), 2889–2894.
- Crist, R. E., Li, W., and Gilbert, C. D. (2001). “Learning to see: Experience and attention in primary visual cortex,” *Nat. Neurosci.* **4**(5), 519–525.
- Fahle, M. (1997). “Specificity of learning curvature, orientation, and vernier discriminations,” *Vision Res.* **37**(14), 1885–1895.
- Farcomeni, A. (2008). “A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion,” *Stat. Methods Med. Res.* **17**(4), 347–388.

- Gao, E., and Suga, N. (1998). "Experience-dependent corticofugal adjustment of midbrain frequency map in bat auditory system," *Proc. Natl. Acad. Sci. U.S.A.* **95**(21), 12663–12670.
- Gilbert, C. D., and Sigman, M. (2007). "Brain states: Top-down influences in sensory processing," *Neuron* **54**(5), 677–696.
- Goldstein, M. H., King, A. P., and West, M. J. (2003). "Social interaction shapes babbling: Testing parallels between birdsong and speech," *Proc. Natl. Acad. Sci. U.S.A.* **100**(13), 8030–8035.
- Goldstein, M. H., Waterfall, H. R., Lotem, A., Halpern, J. Y., Schwade, J. A., Onnis, L., and Edelman, S. (2010). "General cognitive principles for learning structure in time and space," *Trends Cogn. Sci.* **14**(6), 249–258.
- Grossberg, S. (1987). "Competitive learning: From interactive activation to adaptive resonance," *Cognit. Sci.* **11**(1), 23–63.
- Hauptmann, B., and Karni, A. (2002). "From primed to learn: The saturation of repetition priming and the induction of long-term memory," *Brain Res. Cogn. Brain Res.* **13**(3), 313–322.
- Hauptmann, B., Reinhart, E., Brandt, S. A., and Karni, A. (2005). "The predictive value of the leveling off of within session performance for procedural memory consolidation," *Brain Res. Cogn. Brain Res.* **24**(2), 181–189.
- Hayes-Harb, R. (2007). "Lexical and statistical evidence in the acquisition of second language phonemes," *Second Lang. Res.* **23**(1), 65–94.
- Hochberg, Y. (1988). "A sharper Bonferroni procedure for multiple tests of significance," *Biometrika* **75**(4), 800–802.
- Iverson, P., Hazan, V., and Bannister, K. (2005). "Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults," *J. Acoust. Soc. Am.* **118**(5), 3267–3278.
- Karni, A., and Sagi, D. (1991). "Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity," *Proc. Natl. Acad. Sci. U.S.A.* **88**(11), 4966–4970.
- Kuhl, P. K. (2003). "Human speech and birdsong: Communication and the social brain," *Proc. Natl. Acad. Sci. U.S.A.* **100**(17), 9645–9646.
- Kuhl, P. K. (2007). "Is speech learning 'gated' by the social brain?," *Dev. Sci.* **10**(1), 110–120.
- Levi, D. M., and Polat, U. (1996). "Neural plasticity in adults with amblyopia," *Proc. Natl. Acad. Sci. U.S.A.* **93**(13), 6830–6834.
- Li, W., Piech, V., and Gilbert, C. D. (2004). "Perceptual learning and top-down influences in primary visual cortex," *Nat. Neurosci.* **7**(6), 651–657.
- Linck, J. A., Kroll, J. F., and Sunderman, G. (2009). "Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning," *Psychol. Sci.* **20**(12), 1507–1515.
- MacKain, K. S., Best, C. T., and Strange, W. (1981). "Categorical perception of English /r/ and /l/ by Japanese bilinguals," *Appl. Psycholinguist.* **2**(04), 369–390.
- Maye, J., Werker, J. F., and Gerken, L. (2002). "Infant sensitivity to distributional information can affect phonetic discrimination," *Cognition* **82**(3), B101–B111.
- McClaskey, C. L., Pisoni, D. B., and Carrell, T. D. (1983). "Transfer of training of a new linguistic contrast in voicing," *Percept. Psychophys.* **34**(4), 323–330.
- Meinhardt, G. (2002). "Learning to discriminate simple sinusoidal gratings is task specific," *Psychol. Res.* **66**(2), 143–156.
- Norris, J. M., and Ortega, L. (2000). "Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis," *Lang. Learn.* **50**(3), 417–528.
- Ofen-Noy, N., Dudai, Y., and Karni, A. (2003). "Skill learning in mirror reading: How repetition determines acquisition," *Brain Res. Cogn. Brain Res.* **17**(2), 507–521.
- Ortiz, J. A., and Wright, B. A. (2010). "Differential rates of consolidation of conceptual and stimulus learning following training on an auditory skill," *Exp. Brain Res.* **201**(3), 441–451.
- Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (1982). "Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants," *J. Exp. Psychol. Hum. Percept. Perform.* **8**(2), 297–314.
- Polley, D. B., Steinberg, E. E., and Merzenich, M. M. (2006). "Perceptual learning directs auditory cortical map reorganization through top-down influences," *J. Neurosci.* **26**(18), 4970–4982.
- Recanzone, G. H., Merzenich, M. M., Jenkins, W. M., Grajski, K. A., and Dinse, H. R. (1992). "Topographic reorganization of the hand representation in cortical area 3b owl monkeys trained in a frequency-discrimination task," *J. Neurophysiol.* **67**(5), 1031–1056.
- Recanzone, G. H., Schreiner, C. E., and Merzenich, M. M. (1993). "Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys," *J. Neurosci.* **13**(1), 87–103.
- Reuland, D. S., Slatt, L. M., Aleman, M. A., Fernandez, A., and Dewalt, D. (2012). "Effect of Spanish language immersion rotations on medical student Spanish fluency," *Fam. Med.* **44**(2), 110–116.
- Sabin, A. T., Clark, C. A., Eddins, D. A., and Wright, B. A. (2013). "Different patterns of perceptual learning on spectral modulation detection between older hearing-impaired and younger normal-hearing adults," *J. Assoc. Res. Otolaryngol.* **14**(2), 283–294.
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). "Statistical learning by 8-month-old infants," *Science* **274**(5294), 1926–1928.
- Samuel, A. G., and Kraljic, T. (2009). "Perceptual learning for speech," *Atten. Percept. Psychophys.* **71**(6), 1207–1218.
- Seitz, A. R., and Dinse, H. R. (2007). "A common framework for perceptual learning," *Curr. Opin. Neurobiol.* **17**(2), 148–153.
- Shiu, L. P., and Pashler, H. (1992). "Improvement in line orientation discrimination is retinally local but dependent on cognitive set," *Percept. Psychophys.* **52**(5), 582–588.
- Sidas, S. K., Alexander, J. E., and Nygaard, L. C. (2009). "Perceptual learning of systematic variation in Spanish-accented speech," *J. Acoust. Soc. Am.* **125**(5), 3306–3316.
- Stefan, K., Wycislo, M., and Classen, J. (2004). "Modulation of associative human motor cortical plasticity by attention," *J. Neurophysiol.* **92**(1), 66–72.
- Szpiro, S. F., Wright, B. A., and Carrasco, M. (2014). "Learning one task by interleaving practice with another task," *Vision Res.* **101**, 118–124.
- Toro, J. M., Sinnett, S., and Soto-Faraco, S. (2005). "Speech segmentation by statistical learning depends on attention," *Cognition* **97**(2), B25–B34.
- Tremblay, K., Kraus, N., Carrell, T. D., and McGee, T. (1997). "Central auditory system plasticity: Generalization to novel stimuli following listening training," *J. Acoust. Soc. Am.* **102**(6), 3762–3773.
- Tremblay, K. L., Inoue, K., McClannahan, K., and Ross, B. (2010). "Repeated stimulus exposure alters the way sound is encoded in the human brain," *PLoS One* **5**(4), e10283.
- Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. (1999). "Training American listeners to perceive Mandarin tones," *J. Acoust. Soc. Am.* **106**(6), 3649–3658.
- Wright, B. A., and Sabin, A. T. (2007). "Perceptual learning: How much daily training is enough?," *Exp. Brain Res.* **180**(4), 727–736.
- Wright, B. A., Sabin, A. T., Zhang, Y., Marrone, N., and Fitzgerald, M. B. (2010). "Enhancing perceptual learning by combining practice with periods of additional sensory stimulation," *J. Neurosci.* **30**(38), 12868–12877.
- Wright, B. A., and Zhang, Y. (2009a). "A review of the generalization of auditory learning," *Philos. Trans. R. Soc. London, B* **364**(1515), 301–311.
- Wright, B. A., and Zhang, Y. (2009b). "Insights into sound processing gained from perceptual learning," in *The Cognitive Neurosciences*, edited by M. Gazzaniga (MIT Press, Cambridge, MA), 4th ed., pp. 353–366.