Restorative Neurology and Neuroscience 25 (2007) 263–272
IOS Press

Generalization of non-linguistic auditory perceptual training to syllable discrimination

Kala Lakshminarayanan* and Paula Tallal
Center for Molecular and Behavioral Neuroscience, Rutgers University, Newark, NJ, USA

Abstract. *Purpose:* The generalization of non-linguistic auditory perceptual training to syllable discrimination was investigated in two experiments.
Methods: Participants were divided into a control and training group. Both groups came for pre and post-testing sessions spaced ten days apart. Following pre- testing, the training group also participated for five consecutive days in non-linguistic auditory perceptual training. Training was adaptive and involved active sequencing of rising and falling frequency modulated sweeps for 30 minutes per day. Sweeps were passively varied in onset frequency, duration and rate of presentation. A syllable discrimination threshold (SDT) task was used as the pre and post-test measure. In experiment 1, a /ba/-/da/ syllable continuum was used. In experiment 2, the pre-test battery was expanded to include /ba/-/da/, /ba/-/wa/, and /sa/-/sta/ syllable continua and a tone sequencing task that mimicked other parameters of training.
Results: Results of experiment 1 revealed that the training group had a significantly lowered (better) SDT following training as compared to the control group. The extent of training-driven perceptual gain was significantly correlated with pre-training performance. In experiment 2, training resulted in a significantly lowered SDT for /ba/-/da/, but not for the other syllables or the tone sequencing task.
Conclusions: Results showed that task-specific attention drives generalization of auditory perceptual training from non-linguistic to linguistic contexts. Furthermore, individual differences in initial perceptual performance affect the degree of generalization following training.

Keywords: Discrimination training, generalization, perceptual learning, auditory plasticity, syllable discrimination

1. Introduction

Practice with psychophysical tasks is known to improve performance and has often been the reason that researchers prefer to use ‘trained’ participants to avoid confounding effects of practice. Practice or experience driven improvement in performance as measured by an individual’s response to a particular stimulus is termed ‘perceptual learning’. Recently, research in perceptual learning has gained considerable interest owing to its potential in clinical rehabilitation of neurological disorders. In the motor domain, patients with strokes leading to paralysis have regained significant motor function following constraint induced motor training (Taub & Uswatte, 2003). For dyslexic children with deficits in eye movement control, visual training has helped improve perception and also voluntary saccade control (Fischer & Hartnegg, 2000). In amblyopic individuals, who have reductions in contrast sensitivity, vernier acuity, contour detection and spatial resolution, low level visual training has been shown to improve both contrast sensitivity and performance in letter recognition tasks (Polat, Ma-Naim, Belkin, & Sagi, 2004). Intervention programs based on auditory perceptual learning also have been shown to be beneficial for individuals with central auditory processing disorders and dyslexia (Kujala et al., 2001) as well as language learning impairments (Merzenich et al., 1996; Tallal et al., 1996). Importantly, these studies have not only demonstrated perceptual performance gains as a result of training, but also shown correlations between underlying neural rep-

---

*Corresponding author: Kala Lakshminarayanan, Center for Molecular and Behavioral Neuroscience, 197 University Avenue, Newark, NJ 07102, USA. E-mail: kala@pegasus.rutgers.edu.

resentations and performance following training (Temple et al., 2003; Tremblay, Kraus, Carrell, & McGee, 1997). However, one of the difficulties with clinical studies is that perceptual training given for rehabilitation often includes, by design, simultaneous training of multiple perceptual, cognitive and linguistic abilities. For example, in interventions designed to ameliorate spectro-temporal auditory processing deficits in individuals with reading and language learning disabilities (Merzenich et al., 1996; Tallal et al., 1996), training exercises were designed to also include specific components of linguistic processing as well as attention and memory skills. Thus, despite promising outcomes these clinical studies do not allow strong conclusions to be drawn about the specificity of the relation between the training stimuli and performance following training. Better understanding of these relationships will not only increase our theoretical understanding of the mechanisms underlying specific disorders, but also may help improve the design of training programs and lead to more effective outcomes.

In the auditory modality, non-clinical research studies on the transfer of learning and generalization have primarily focused on frequency discrimination (Amitay, Hawkey, & Moore, 2005; Ari-Even Roth, Amir, Alaluf, Buchsenspanner, & Kishon-Rabin, 2003; Delhommeau, Micheyl, & Jouvent, 2005; Delhommeau, Micheyl, Jouvent, & Collet, 2002; Demany, 1985; Demany & Semal, 2002; Grimault, Micheyl, Carlyon, Bacon, & Collet, 2003; Irvine, Martin, Klimkeit, & Smith, 2000), and only a few have been concerned with duration (Nagarajan, Blake, Wright, Byl, & Merzenich, 1998; Wright, Buonomano, Mahncke, & Merzenich, 1997) or speech perception (McClelland, Fiez, & McCandliss, 2002; Tremblay, Kraus, Carrell, & McGee, 1997). Even in stringently controlled studies on the effects of perceptual training, it is difficult to separate task learning from 'perceptual' learning. This is primarily because in most studies the tasks used in training and testing are the same. For example, listeners may be trained on discrimination at one particular frequency and their performance evaluated at various other frequency ranges (Delhommeau, Micheyl, Jouvent, & Collet, 2002; Demany, 1985) or they may be trained at discrimination of one particular time interval and their performance evaluated at other intervals (Wright, Buonomano, Mahncke, & Merzenich, 1997). In such a case it is difficult to rule out the possibility that the improvement observed in performance is not the result of greater efficiency in performing the task itself, rather than true sharpening of perceptual thresholds.

One of the primary goals of perceptual training is generalization beyond the explicitly trained stimuli. In the reported studies, we address the question of generalization as relevant for auditory training using psychophysical methods. We specifically examine whether training in a purely non-linguistic context will generalize to a linguistic context. We further examine how factors related to design of training and individual differences may impact the extent of generalization and transfer. These questions are of considerable clinical and theoretical interest as they are likely to better inform rehabilitation efforts.

Other studies have previously demonstrated that performance on standardized speech, language and reading tests improve following training using non-linguistic acoustic stimuli as well as acoustically modified speech (Kujala et al., 2001; Merzenich et al., 1996; Tallal et al., 1996; Temple et al., 2003). It is, however, difficult to examine specifically what aspects of training affect generalization when both training and outcome measures include stimuli ranging from tones to syllables to sentences to paragraphs. Processing of sentences and paragraphs involves increasing demands not only on linguistic comprehension, but also on attention and memory, and results could be confounded by individual differences in any combination of these variables. Therefore, in order to be able to quantify the specific effects of non-linguistic acoustic training on speech, we have restricted our focus to the perception of syllables and used syllable discrimination as our target task.

Components of speech can be conceived of in a number of ways: 1) as a combination of sinusoids (Remez, Rubin, Pisoni, & Carrell, 1981), 2) as a combination of amplitude-modulated (AM) bands and frequency modulated (FM) bands (Greenberg, Ainsworth, Popper, & Fay, 2004; Rosen, 1992; Zeng et al., 2005), or 3) as being comprised of rapid transients (FM sweep like formant transitions) that characterize the consonant part of the syllable and steady state portions that characterize the vowel portion of the syllable. For the purposes of this study, we have focused on the latter.

The goal of experiment 1 (E1) was to see if training individuals in identifying a specific non-linguistic acoustic feature (upward and downward FM sweeps) would transfer to their ability to discriminate between speech syllables that differ primarily in that same acoustic feature (the direction of frequency change within formant transitions). Furthermore, the non-linguistic stimuli used for training were FM sweeps varying on multiple acoustic parameters such as onset frequency, duration, and rate. Thus, in Experiment 2 (E2) each of these variables was examined independently. Each of the speech syllable pairs selected for evaluation in E2 was designed to approximately mimic a specific acoustic parameter of training. The design of acoustic training was also such that it allowed for an examination of how the nature of the training task itself may affect generalization. Specifically, during training, while actively identifying up-sweeps from down-sweeps, participants were also listening passively to variations in stimulus duration and rate of presentation.

2. Methods

All details of stimuli and task were exactly the same for both Experiment 1 (E1) and Experiment 2 (E2) except for the difference in number of subjects and various evaluation tasks included for pre and post-testing. Nineteen subjects (8 control and 11 training group) participated in E1 while twenty-four subjects (12 each in control and training group) participated in E2. None of the participants were common to both experiments. For E1, pre-testing and post-testing sessions consisted of only /ba/-/da/ syllable discrimination, while for E2 in addition to /ba/-/da/ syllable discrimination, three other test conditions were included: /ba/-/wa/ syllable discrimination, /sa/-/sta/ syllable discrimination and tone sequencing rate.

Participants: All participants were healthy college students ranging in age from 18–25 yrs and were given a screening test consisting of audiometric testing, and a 2-subtest version of Wechsler's Abbreviated Scale of Intelligence (WASI) to evaluate verbal and non-verbal IQ. The verbal subtest required participants to provide definitions of words that gradually increased in difficulty. In the non-verbal subtest, participants were presented with a grid of images in which one of the images was missing. The individual was required to identify from among 5 alternatives, the image that would fill the missing piece in the grid. Participants had normal hearing levels ($\leqslant$ 25 dB) in the range of 250 Hz to 8000 Hz. They were divided into one of two groups: control (CG) and training (TG). None of the participants had any history of hearing or language disabilities and all were above the 50^{th} percentile for the population on IQ scores as estimated by the WASI. All subjects were paid and given gifts as incentive for participation. The Rutgers University institutional review board for human subject research approved the study protocol.

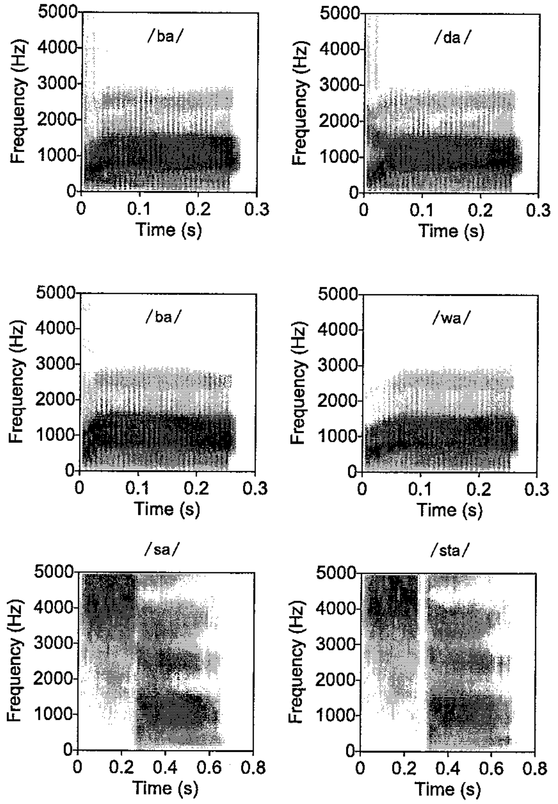

Fig. 1. Spectrograms of the endpoints of the syllable continua. Note that in the top syllable pair the second formant is rising for /ba/ and falling for /da/; in the middle pair the duration of transition for /ba/ is short while for /wa/ it is extended, in the lower pair there is a gap between the frication and steady state part for /sta/.

Study Design: The study included three sessions: Pre-Testing, Training and Post-Testing. The control group (CG) participated only in the Pre- and Post-Testing sessions. The training group (TG) participated in all three sessions. Pre-Testing was conducted in two 45-minute sessions on Thursday and Friday of the first week. Training was conducted in week two for 30 minutes per day on 5 consecutive days. Post-Testing was conducted in week three, two days following completion of training.

Pre and Post testing sessions consisted of a battery of 4 psychoacoustic tasks that included a 3-down-1-up adaptive, 2 interval-2 alternative forced choice (2I-2AFC) paradigm to determine syllable discrimination thresholds for /ba/-/da/, /ba/-/wa/, and /sa/-/sta/ syllable continua (Fig. 1) and a 3-down 1-up adaptive tone sequencing task to determine rate thresholds for tone sequencing.

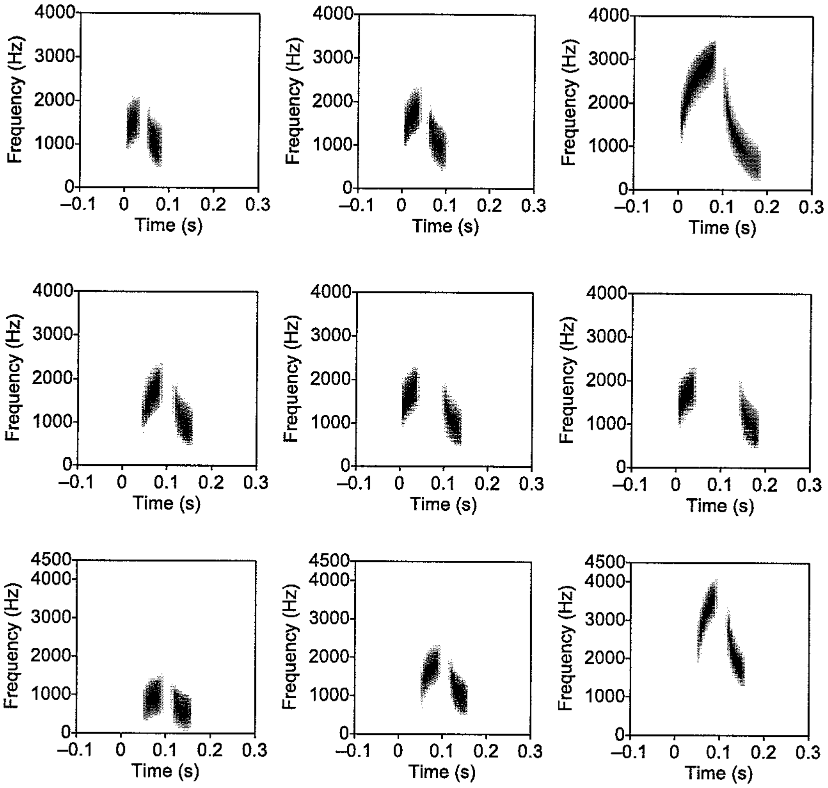

Fig. 2. Spectrograms of FM sweeps used as training stimuli illustrating the dimensions involved in training. During training 6 variations of duration, 45 variations of ISI and 3 variations of onset frequency pitch were presented resulting in a total of 810 different stimuli. Top panel illustrates durations of 80 ms, 40 ms, and 30 ms; middle panel illustrates ISI variations of 100 ms, 60 ms and 30 ms; bottom panel illustrates onset frequency variations of 2 KHz, 1 KHz, and 500 Hz.

Training consisted of an adaptive task in which participants were required to replicate by an appropriate key press or mouse response the order in which they heard an upward and downward frequency modulated (FM) sweep.

Training stimuli (Fig. 2) included variations in 1) direction of FM sweep, 2) duration of FM sweep, and 3) inter-stimulus-interval (ISI) between the FM sweeps. The pre- and post-testing speech syllable stimuli were chosen so as to correspond to the non-linguistic acoustic training parameters. This led to the selection of the /ba/-/da/ syllable continuum that varied in the *direction* of the formant transition, the /ba/-/wa/ syllable continuum varying in the *duration* of formant transition, and the /sa/-/sta/ syllable continuum varying in the *duration of silence* between the offset of frication and onset of the steady state portion of the syllable. In addition, in E2 a temporal tone sequencing task was used to evaluate if there was generalization of the rate of presentation or temporal order judgment aspect of training.

2.1. Test sessions

Stimuli: The three sets of syllable stimulus continua /ba/-/da/, /ba/-/wa/ and /sa/- /sta/. were digitally synthesized using Sensyn, a formant based Klatt synthesizer and Goldwave, a wave editing software.- Stimulus pairs were created for all three of the continua in increasing order of difficulty, with stimuli 1 and 21 forming the first (most acoustically distinct) pair, followed by 1 and 20, 1 and 19 and so on. The inter-stimulus interval (ISI) between syllables in a pair was 700 ms.

/ba/-/da/: The /ba/-/da/ continuum varies in the onset of the second formant (F2) and hence in the direction of the formant transition. The first formant (F1) and third formant (F3) onsets were the same for both endpoints of the continuum at 400 Hz and 2100 Hz, respectively. F2 onset was 800 Hz for /ba/ and 1600 Hz for /da/. The steady state vowel frequencies of 800 Hz, 1200 Hz and 2500 Hz were reached over a period of 30 ms. Overall duration of the syllable was 270 ms. F2 onset was varied using a step size of 40 Hz to create the 21-stimulus continuum.

/ba/-wa/: The /ba/-/wa/ continuum varies in the duration of the formant transition. Overall duration of the syllable was 270 ms. All parameters for /ba/ were the same as in the /ba/-/da/ continuum. Formant transition duration was changed from 30 ms to 70 ms for /wa/ following a step size of 2 ms in order to create the 21-stimulus continuum.

/sa/-/sta/: The syllable /sa/ was modified from the CUNY NST stimulus set developed by the digital sensory aids laboratory. Silence of various durations was inserted at the appropriate zero-crossing of the waveform to create the /sta/ token. Silence durations of 0, 2, 4, 6 and so on, up to 40 ms duration were inserted creating a total of 21 tokens in the continuum. The frication was 262 ms long and the vowel was 410 ms long resulting in overall duration of 672 ms to 712 ms.

Tone sequencing rate (TSR): Two five-harmonic complexes of a 500 Hz sinusoid (low tone: L) and 1000 Hz sinusoid (high tone: H) were used. Three tone sequences (HLH; LHL etc.) were created wherein the ISI was varied from 500 ms to 0 ms. Step size was varied in 50 ms steps in going from 500 ms to 100 ms ISI; then in 20 ms steps down to 20 ms ISI and then in 2 ms steps down to 0 ms ISI. Three tone sequences were used to avoid ceiling effects with young healthy adults that were seen with two tone sequences in pilot studies.

Procedure: Participants were asked to discriminate pairs of stimuli as same or different. Trials began with the easiest pair and increased in difficulty adaptively following a 3-down 1-up procedure. After three consecutive correct responses difficulty was increased by two step-sizes and for every incorrect response difficulty was decreased by one step-size. Trials were terminated after six reversals. Average of the four best discrimination limens (among the six reversals) was taken as the individual's threshold. Trials were also to be terminated if an individual got five incorrect responses to the first stimulus pair (easiest to discriminate, largest acoustic distance). However, none of the subjects had incorrect responses for the first pair.

Calculation of threshold gain following training: Threshold limens were calculated both at pre-testing and post-testing sessions. For /ba/-/da/, the limen was calculated as the difference between the second formant onset for /da/ and that for /ba/. Hence for the easiest pair consisting of the end-points, it would be 800 Hz. For /ba/-/wa/, the limen was calculated as the transition duration for /wa/ minus the transition duration for /ba/. For the first pair it would be 40ms. Similarly the limen for /sa/-/sta/ corresponded to the duration of the silence in the syllable /sta/. For the first pair the limen is 40ms. Gain was calculated as the pre-test limen minus post-test limen and normalized with respect to each individual's initial threshold (i.e. gain was represented as a percentage of the individual's own initial threshold) in order to account for the variable nature of speech thresholds and to control for any difference in initial limens of the control group and training group. For the tone sequencing rate task, the difference between the pre-test threshold and post-test threshold was taken as the gain in performance.

2.2. *Training sessions (Common for both E1 and E2):*

Procedure: For each trial, participants were required to indicate the order in which they heard the rising and falling sweep by pressing in the same order the appropriate icons (up and down arrow) on the screen. The duration and ISI of the training stimuli was varied adaptively according to individual accuracy. Trials began with the 80 ms long sweeps with 500 ms ISI and 1000 Hz onset frequency and subsequently the difficulty level was increased or decreased in accordance to individual accuracy. On subsequent days the onset frequency was randomly varied and counterbalanced to one of the 3 possibilities (500 Hz, 1 KHz or 2 KHz). For each trial feedback was presented with cartoon animations and points. A total of 810 levels of training were possible (3 onset levels of frequency X 6 levels of duration X 45 levels of ISI). At the conclusion of training each participant had completed a different number of levels based of their accuracy and speed.

Training Stimuli: Pairs of frequency-modulated sweeps rising or falling in pitch were presented during training. The rate of modulation of the sweeps was 16 octaves per second, the average rate seen in speech. The sweeps varied in onset frequency, duration, and rate of presentation. Three different onset frequencies were used: 500 Hz, 1000 Hz or 2000 Hz. Duration varied between 80 ms, 60 ms, 40 ms, 35 ms, 30 ms and 25 ms. The ISI between the two sweeps ranged from 500 ms to 0 ms in 45 steps. From 500 ms to 300 ms, the step size was 100 ms and then it was 50 ms until the ISI reduced to 200 ms after which the step size was further reduced to 5 ms. The task and stimuli used in the training section were derived from a subsection of the commercially available software, Fast ForWord® (©Scientific Learning Corporation).

3. Results

Changes in discrimination limens for syllable continua (/ba/-/da/, /ba/-/wa/ and /sa/-/sta/) were normalized with respect to each individual's own pre-test threshold prior to analysis as described in the methods section. Alpha was set to 0.05 to determine significance for all analyses.

3.1. Experiment 1

In this experiment our primary goal was to test if purely non-linguistic acoustic training designed to mimic acoustic features of linguistic stimuli would generalize to perception of linguistic stimuli. Specifically, the question was to see if perceptual training in discrimination of FM sweep direction would improve (or lower) the discrimination threshold of /ba/-/da/ syllable continua that differ in direction of formant transition. There was no difference between the control group and training group in terms of composite IQ [$t(17) = -0.2$, $p = 0.8$], or specific verbal [$t(17) = -0.04,$]$p = 0.7$] or non-verbal abilities [$t(17) = -0.09$, $p = 0.9$] based on standardized WASI assessment. T-test analysis of the effect of training on /ba/-/da/ discrimination threshold revealed a significant effect of training [$t(17) = -2.38$, $p = 0.015$] on improvement in performance following training (Fig. 3). The training group showed a significantly greater threshold gain at post-test (Mean = 8.7; S.D = 8.1) compared to the control group (Mean = 1.3; S.D = 3.99). A second look at the results shows that the training group had wider variability compared to the control group. We hypothesized that the individuals who had the greatest difficulty in discriminating the /ba/-/da/ syllable pair initially at pre-testing would be the ones likely to benefit most from the training. A correlation was run to examine if the initial (pre-training) discrimination thresholds were related to the threshold gain following training. Our prediction was that greater the initial threshold, the greater would be the gain following training. The percentage of training completed was computed [(individual level reached/total number of levels) X 100] for each individual. As predicted, correlation analysis demonstrated that there was a significant relationship between initial (pre-test) discrimination threshold and threshold gain following training [$r(10) = 0.58$, $p = 0.03$].

Based on this result we further hypothesized that those individuals who had greater difficulty discriminating the /ba/-/da/ syllable pair initially would also likely be more challenged by the training task, and so may benefit more from it. Hence, we predicted that individuals who were unable to make as much progress on the training task (made more errors and so progressed more slowly) were likely to benefit more from training. Correlation analysis between percentage of training completed and change in syllable discrimination threshold following training supported this inverse relationship that approached, but did not reach significance [$r(10) = -0.48$, $p = 0.065$].

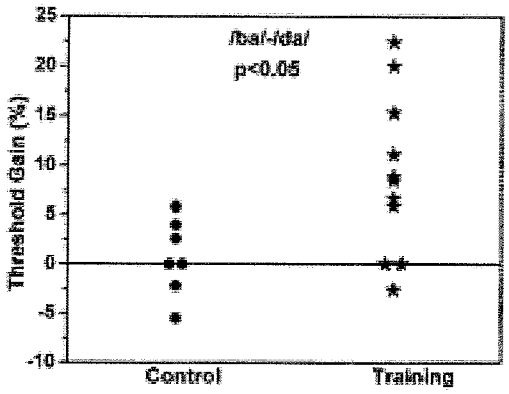

Fig. 3. Distribution of threshold gain (pre-test discrimination threshold limen minus post-test discrimination threshold limen), normalized with respect to initial performance, is plotted for the two groups. Zero percent indicates no change between pre- and post-test. Positive change indicates lowered threshold (or improvement) at post-test and negative percentages indicate higher thresholds at post-test.

3.2. Experiment 2

This experiment was designed to examine how the physical characteristics of the training stimuli and specific task components might affect generalization of training from one context to another. Thresholds for discriminating three different pairs of syllables (/ba/-/da/, /ba/-/wa/ and /sa/-/sta/) were determined at pre and post-test phases. Recall that the syllable contrasts were chosen specifically to reflect the varying acoustic parameters within the non-linguistic training stimuli (sweep direction, sweep duration, inter-stimulus interval). However, in order to progress successfully in training, participants were only required to actively attend to the variation in sweep direction. The overall duration of the sweeps and the rate at which the sweeps were presented represented passive components of the task and changed adaptively based on the subject's performance on the active component of the task. Based on neurophysiological studies of auditory cortical plasticity in animals (Fritz, Elhilali, & Shamma, 2005; Fritz, Shamma, Elhilali, & Klein, 2003; Polley, Steinberg, & Merzenich, 2006), we hypothesized that only the actively attended to aspects of training would result in threshold discrimination gains.

As in experiment 1, no significant differences were seen between the control group and training group in terms of composite IQ [$t(22) = -0.3$, $p = 0.7$] or specific verbal [$t(22) = -0.1$, $p = 0.9$] or non-verbal performance scores [$t(22) = -1.4$, $p = 0.2$]. One-tailed t-test analysis of the effect of non-linguistic training on syllable discrimination threshold change revealed a significant difference only in the case of /ba/-/da/ discrimination [$t(22) = 1.95$; $p = 0.03$], but not for /ba/-/wa/ [$t(22) = 0.34$; $p = 0.38$], /sa/-/sta/ [$t(22) = 1.17$; $p = 0.17$] or tone sequencing rate TSR [$t(22) = 0.04$; $p = 0.48$]. This replicated the results from E1 with regard to generalization of training to /ba/-/da/ discrimination, while also demonstrating that generalization of the passive aspects of training did not occur (Fig. 4a and 4b).

The next question addressed was whether the relation between the extents of training completed and the change in threshold following training found in E1 would replicate in E2. Correlation analysis between percentage of training completed and change in threshold following training showed that change in threshold on /ba/-/da/ was, again, inversely correlated with the percent of training completed [$r(11) = -0.67$, $p = 0.01$]. Thus, subjects who reached higher levels on the non-linguistic acoustic training (were less challenged by the training) showed lesser benefit in terms of a threshold change for discriminating /ba/-/da/ following training. This replicated the results from E1. No significant correlations were found for the other speech contrasts.

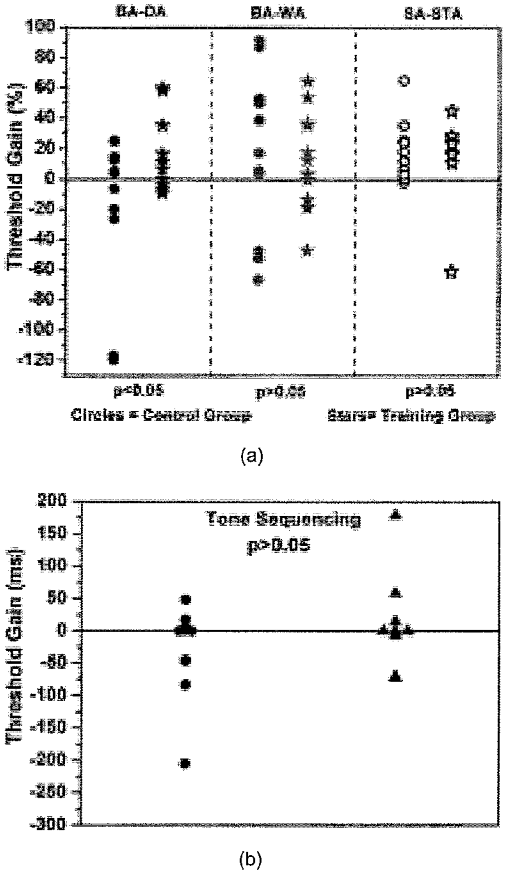

Fig. 4. (a) Distribution of threshold gain (pre-test threshold limen minus post-test threshold limen), normalized with respect to initial performance, is plotted for the two groups. Zero percent indicates no change between pre- and post-test. Positive change indicates lowered threshold (or improvement) at post-test and negative percentages indicate higher thresholds at post-test. The only significant improvement was observed for /ba/-/da/ ($p < 0.05$). (b) Distribution of threshold gain (pre-test threshold limen minus post-test threshold limen) is plotted for the two groups for the tone-sequencing task. Zero milliseconds indicates no change between pre- and post-test. Positive change indicates lowered threshold (or improvement) at post-test and negative values indicate higher thresholds at post-test.

4. Discussion

The experiments reported in this paper were designed to address two main questions: 1) Does auditory perceptual training in a non-linguistic context generalize to discrimination of linguistic stimuli? 2) How do the physical attributes of the stimulus and task used in training impact the extent of generalization?

There were two groups of individuals in each study-of which one group participated in training and the other did not. For both groups syllable discrimination thresholds were obtained at pre-testing and post-testing sessions. Individuals in the training group were trained on identification of the direction of FM sweeps. Although the task used in training involved actively identifying sweep direction, other parameters such as the rate of presentation of the sweeps and duration of sweeps varied passively. In E1, generalization of training was examined specifically for the /ba/-/da/ syllable pair that required discrimination of the direction of formant transition, comparable to the identification of the direction of FM sweeps in training. In E2, generalization was examined for additional syllable pairs that mimicked the acoustic features of training stimuli that were passively changed during training.

Results from both experiments E1 and E2 confirmed that non-linguistic acoustic perceptual training can impact syllable discrimination thresholds as seen from significant improvements in /ba/-/da/ discrimination in the training group as compared to the control group. In E2, a significant effect of training was seen only for the /ba/-/da/ syllable pair suggesting that generalization was limited to the context that reflected the actively attended aspect of the training task, namely, identification of direction of FM sweeps. Although the duration and rate of presentation of the FM sweeps were varied passively during training, these features of training did not show transfer to the linguistic context. The results suggest that task-specific demands of training influence the extent of generalization. Similar results have been reported in the visual modality. It has been shown that cognitive set affects tuning in retinally local orientation channels (Shiu & Pashler, 1992). Specifically, attending to brightness rather than orientation of lines did not improve the ability to discriminate line orientations.

Neural correlates of task-dependent plasticity recently have also been demonstrated in ferrets and rats using single-unit and multiple-unit recordings of the auditory cortex (Fritz, Elhilali, & Shamma, 2005; Fritz, Shamma, Elhilali, & Klein, 2003; Polley, Steinberg, & Merzenich, 2006). Fritz et al. (2003; 2005) examined cortical plasticity as measured by focal changes in the shape of the spectrotemporal response field (STRF) of individual cells in both active and passive behavioral conditions. In the passive condition STRF's were relatively stable across measurements. However, in active conditions (tone-detection task and frequency discrimination task), a facilitation was observed at the target frequency, caused by the enhancement of an excitatory field of the STRF and weakening of inhibitory sidebands. The STRF changes, seen in the majority of A1 neurons (over 70%), were facilitative and reflective of the animal's goal to enhance performance in a tone detection task. In a more recent study, Polley et al. (2006) demonstrated that when rats were trained to attend to either frequency or intensity, depending on task demands, within an identical set of auditory stimuli, the resulting topographic reorganization corresponded only to the specific acoustic feature attended to. Thus, both animal and human studies are consistent in demonstrating that only features that are actively attended to during training show plasticity and generalization.

Another important finding of the current study was that the success of a training program aimed at generalization depends on its ability to challenge the individual's basic abilities. We found that individuals who had higher (poorer) discrimination thresholds to begin with and greater difficulty with the training task showed greater performance gains following training. A similar result has been reported previously (Amitay, Hawkey, & Moore, 2005) in an auditory training study where various stimulus conditions were used in training and improvement in performance was examined on frequency discrimination. Listeners with higher (poorer) initial frequency discrimination limens showed greater improvement following training. This finding was corroborated by both E1 and E2 in our study. Interestingly in the study by Fritz et al. (2003) discussed above, the magnitude of the facilitative changes in STRF was modulated by attentive behavior as measured by improved behavioral performance of the ferrets. Similarly, the degree of topographic map plasticity observed in the Polley et al. (2006) study was correlated with the degree of perceptual learning in the rats.

Previous studies on learning and generalization have shown that neural mechanisms also are important in determining the extent of transfer of learning. Studies on frequency discrimination training have suggested that specific neural mechanisms may play a significant role in determining the extent of transfer of learning and generalization of training (Demany, 1985; Irvine, Martin, Klimkeit, & Smith, 2000). Demany et al. (1985) trained four groups of individuals on frequency discrimination at 0.2 0.36 2.5 and 6Khz and evaluated the discrimination limen at 0.2 KHz. Training at all frequencies, except 6 KHz, resulted in significant generalization to 0.2 Hz. This interesting result led the authors to suggest that the observed generalization may be due specifically to the fact that temporal coding is in play at lower frequencies, but a place mechanism mediates perception of higher frequencies. Therefore, training in frequencies mediated by a place mechanism would not generalize to discrimination of frequencies mediated by temporal mechanisms. It is beyond the scope of the experimental method used in our study to address the neural correlates of the observed results of training.

Studies with both humans and animals have shown that perceptual learning is mediated by discrete experience-driven changes in specific lower level representations that are critical for the performance of the trained task (Fritz, Elhilali, & Shamma, 2005; Fritz, Shamma, Elhilali, & Klein, 2003; Karni & Bertini, 1997; Karni et al., 1998; Karni & Sagi, 1993; Polley, Steinberg, & Merzenich, 2006). For instance, given the extensive evidence that orientation differences give rise to a strong preattentive discriminability of textures, individuals trained on a texture discrimination task were tested for their abilities on detection of orientation and shape. Improvement was seen on detection of specific orientation gradients, but not on shape discrimination. Similarly, in our study, by design, the sweeps used in training changed in frequency at the rate of 16 octaves per second, which is approximately the rate of formant transition used in the /ba/-/da/ continuum, suggesting that the generalization of training is specific to the acoustic parameter trained for. However, as other rates of frequency change were not studied, the question of specificity needs further examination.

Considering that the duration of training used in these studies was minimal, totaling about 2.5 hours over five days, and the participants in the study were normal healthy adults, the results demonstrating generalization of non-linguistic acoustic training of FM sweep direction to the discrimination of speech syllables differing in direction of formant transition are striking and consistent. There are, however, some limitations of the study that require future investigation. Although both experiments showed a significant effect of non linguistic acoustic training of FM sweep direction on a speech contrast (/ba/ vs /da/) that differs in direction of formant transition, it is evident that the results of E1 were more clear-cut, with the control group being within ±5% of their pre-test threshold limens, and training group showing up to 25% improvement from their pre-test threshold limens. In E2, however, the spread of threshold change was much wider in both groups. Further inspection of the data, however, showed that the extreme data-points on each of the test measures used in E2 did not correspond to the same individual, nor did they result from order effects. Therefore, we did not find it warranted discarding any data and the results reported here derive from analysis of the complete dataset. It is important to note, however, that reanalysis of the /ba/-/da/ data in E2, with the value of the two outliers in the control group equated to the lowest scores of the rest of the control group continued to show a statistically significant effect of training. It is likely that the observed variability in E2 was an effect of fatigue resulting from the extended pre- and post- test sessions required to obtain thresholds on three different syllable continua as well as the TSR task in E2 (about an hour) as compared to only one syllable continuum (less than 15 minutes) for E1. Further we are aware that the design did not include explicit evaluation of learning curves for the training stimuli. In order to do so, additional tasks using non-linguistic stimuli comparable to the training stimuli would have to be included. We were concerned, however, that given the already extended duration of testing in E2 the benefits of adding this additional control would have been outweighed by the negative effects of further fatigue.

Another issue that has not been explicitly addressed in the studies reported here pertains to the question of whether the observed improvements are simply due to learning to attend in general or learning to stay motivated. While the study design does not allow us to delineate the effects of general attentional improvements, if the results were simply due to improved attention, one cannot explain why the performance improvement was observed only for the /ba/-/da/ stimulus contrast and not for the other contrasts. Hence, we argue that even if the observed results are due to learning to attend or stay motivated, these were specific to the acoustic parameter that was trained for, and in addition, the parameter that was specifically attended to governed by task design and individual perceptual abilities.

In conclusion, despite some limitations of the study, these findings are an important first step forward towards gaining a better understanding of the factors that affect generalization of non-linguistic auditory perceptual training to speech perception. Based on the results of the experiments reported in this paper, generalization of non-linguistic auditory perceptual training to speech perception does occur, but is affected by many factors including the experimental design, attentional demands of the task and stimuli used for training, neural mechanisms underlying the processing of stimuli under study, and individual differences in perceptual abilities as relevant for the task being studied. These results have important potential clinical implications for the design of intervention programs aimed at helping individuals with developmental language learning disabilities as well as individuals experiencing age related decline in speech perception and language comprehension.

Acknowledgements

We wish to acknowledge Loraine Wolynic, and Carolyn Kloepping for assistance with recruitment and screening of participants and Ian Creese and Teresa Realpe for helpful discussions. We thank NSF Science of Learning Center program for partial funding of this study as well as Scientific Learning Corporation for permission to extract a subsection of their intervention program, Fast ForWord®, for use in the training section of this study.

References

Amitay, S., Hawkey, D. J., & Moore, D. R. (2005). Auditory frequency discrimination learning is affected by stimulus variability. *Percept Psychophys., 67*(4), 691-698.

Ari-Even Roth, D., Amir, O., Alaluf, L., Buchsenspanner, S., & Kishon-Rabin, L. (2003). The effect of training on frequency discrimination: generalization to untrained frequencies and to the untrained ear. *J Basic Clin Physiol Pharmacol., 14*(2), 137-150.

Delhommeau, K., Micheyl, C., & Jouvent, R. (2005). Generalization of frequency discrimination learning across frequencies and ears: implications for underlying neural mechanisms in humans. *J Assoc Res Otolaryngol., 6*(2), 171-179. Epub 2005 Jun 2010.

Delhommeau, K., Micheyl, C., Jouvent, R., & Collet, L. (2002). Transfer of learning across durations and ears in auditory frequency discrimination. *Percept Psychophys., 64*(3), 426-436.

Demany, L. (1985). Perceptual learning in frequency discrimination. *J Acoust Soc Am., 78*(3), 1118-1120.

Demany, L., & Semal, C. (2002). Learning to perceive pitch differences. *J Acoust Soc Am., 111*(3), 1377-1388.

Fischer, B., & Hartnegg, K. (2000). Effects of visual training on saccade control in dyslexia. *Perception., 29*(5), 531-542.

Fritz, J., Elhilali, M., & Shamma, S. (2005). Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hear Res, 206*(1-2), 159-176.

Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci, 6*(11), 1216-1223.

Greenberg, S., Ainsworth, W., Popper, A. N., & Fay, R. R. (Eds.). (2004). *Speech Processing in the Auditory System* (4th ed.). New York: Springer Verlag.

Grimault, N., Micheyl, C., Carlyon, R. P., Bacon, S. P., & Collet, L. (2003). Learning in discrimination of frequency or modulation rate: generalization to fundamental frequency discrimination. *Hear Res., 184*(1-2), 41-50.

Irvine, D. R., Martin, R. L., Klimkeit, E., & Smith, R. (2000). Specificity of perceptual learning in a frequency discrimination task. *J Acoust Soc Am., 108*(6), 2964-2968.

Karni, A., & Bertini, G. (1997). Learning perceptual skills: behavioral probes into adult cortical plasticity. *Curr Opin Neurobiol, 7*(4), 530-535.

Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M. M., Turner, R., et al. (1998). The acquisition of skilled motor performance: fast and slow experience-driven changes in primary motor cortex. *Proc Natl Acad Sci U S A, 95*(3), 861-868.

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature, 365*(6443), 250-252.

Kujala, T., Karma, K., Ceponiene, R., Belitz, S., Turkkila, P., Tervaniemi, M., et al. (2001). Plastic neural changes and reading improvement caused by audiovisual training in readingimpaired children. *Proc Natl Acad Sci U S A., 98*(18), 10509-10514. Epub 12001 Aug 10521.

McClelland, J. L., Fiez, J. A., & McCandliss, B. D. (2002). Teaching the /r/-/l/ discrimination to Japanese adults: behavioral and neural aspects. *Physiol Behav., 77*(4-5), 657-662.

Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., & Tallal, P. (1996). Temporal processing deficits of language-learning impaired children ameliorated by training. *Science, 271*(5245), 77-81.

Nagarajan, S. S., Blake, D. T., Wright, B. A., Byl, N., & Merzenich, M. M. (1998). Practice-related improvements in somatosensory interval discrimination are temporally specific but generalize across skin location, hemisphere, and modality. *J Neurosci., 18*(4), 1559-1570.

Polat, U., Ma-Naim, T., Belkin, M., & Sagi, D. (2004). Improving vision in adult amblyopia by perceptual learning. *Proc Natl Acad Sci U S A., 101*(17), 6692-6697. Epub 2004 Apr 6619.

Polley, D. B., Steinberg, E. E., & Merzenich, M. M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *J Neurosci, 26*(18), 4970-4982.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212*(4497), 947-949.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci, 336*(1278), 367-373.

Shiu, L. P., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Percept Psychophys, 52*(5), 582-588.

Stevens, C., Sanders, L., Neville, H. Mechanisms of selective auditory attention are both deficient and remediable in children with specific language impairment: evidence from event-related brain potentials. Presented at the Society for Neuroscience annual meeting, 2006 in Atlanta, Georgia.

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., et al. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science, 271*(5245), 81-84.

Taub, E., & Uswatte, G. (2003). Constraint-induced movement therapy: bridging from the primate laboratory to the stroke rehabilitation laboratory. *J Rehabil Med, 41*(Suppl), 34-40.

Temple, E., Deutsch, G. K., Poldrack, R. A., Miller, S. L., Tallal, P., Merzenich, M. M., et al. (2003). Neural deficits in children with dyslexia ameliorated by behavioral remediation: evidence from functional MRI. *Proc Natl Acad Sci USA, 100*(5), 2860-2865.

Tremblay, K., Kraus, N., Carrell, T. D., & McGee, T. (1997). Central auditory system plasticity: generalization to novel stimuli following listening training. *J Acoust Soc Am, 102*(6), 3762-3773.

Wright, B. A., Buonomano, D. V., Mahncke, H. W., & Merzenich, M. M. (1997). Learning and generalization of auditory temporal-interval discrimination in humans. *J Neurosci, 17*(10), 3956-3963.

Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., et al. (2005). Speech recognition with amplitude and frequency modulations. *Proc Natl Acad Sci USA, 102*(7), 2293-2298. Epub 2005 Jan 2227.