Individual differences in the acquisition of second language phonology

Narly Golestani [a,b,c,*], Robert J. Zatorre [a]

[a] *Cognitive Neuroscience Unit/Montreal Neurological Institute, McGill University, Montreal, Que., Canada*
[b] *Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, UK*
[c] *Functional Brain Mapping Laboratory, Neurology Clinic, University Medical School, 24 rue Micheli-du-Crest 1211 Geneva, Switzerland*

Abstract

Perceptual training was employed to characterize individual differences in non-native speech sound learning. Fifty-nine adult English speakers were trained to distinguish the Hindi dental–retroflex contrast, as well as a tonal pitch contrast. Training resulted in overall group improvement in the ability to identify and to discriminate the phonetic and the tonal contrasts, but there were considerable individual differences in performance. A category boundary effect during the post-training discrimination of the Hindi but not of the tonal contrast suggests different learning mechanisms for these two stimulus types. Specifically, our results suggest that successful learning of the speech sounds involves the formation of a long-term memory category representation for the new speech sound.
© 2008 Elsevier Inc. All rights reserved.

1. Introduction

During adulthood, most individuals perceptually assimilate certain non-native speech sounds with similar ones from the native language (Best, McRoberts, & Sithole, 1988). For example, native English speakers typically cannot hear the difference between the English alveolar and the Hindi retroflex stop consonants. So even though native speakers of Indian languages hear the difference as easily as anyone reading this paper can hear the difference between 'b' and 'd', to native English speakers, both are heard as the voiced alveolar stop consonant 'd' (Polka, 1991; Rivera-Gaxiola, Csibra, Johnson, & Karmiloff-Smith, 2000a; Werker & Lalonde, 1988). There is evidence, however, for considerable variability across individuals in the learning of speech sounds that involve rapid spectral change, in particular ones that contrast place of articula-

tion, such as the r/l contrast for native Japanese listeners, or the dental–retroflex one for native English speakers. Specifically, even among individuals with similar language backgrounds, some are able to hear these non-native sounds in adulthood, or to learn them quickly with limited training, while for others, learning to hear these sounds is more slow and effortful (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Golestani & Zatorre, 2004; Polka, 1991; Pruitt, Strange, Polka, & Aguilar, 1990).

We address two principal questions in the present study. First, we wanted to characterize individual differences in how well or how quickly English-speaking adults can learn to hear the non-native retroflex sound. Very few studies have attempted to account for individual differences in non-native speech sound learning (Bradlow et al., 1997; Flege, MacKay, & Meador, 1999). Individual differences were characterized by reporting performance variability, and by examining relationships between measures of pre- and post-training identification and discrimination test performance, as well as between these test measures and measures of performance during training. We also showed that such individual differences predict individual differences in brain structure in two other studies (Golestani, Molko,

---

[*] Corresponding author. Address: Functional Brain Mapping Laboratory, Neurology Clinic, University Hospital and Department of Fundamental Neurosciences, University Medical School, 24 rue Micheli-du-Crest 1211 Geneva, Switzerland. Fax: +44 (0) 20 7916 8517.
E-mail address: n.golestani@ucl.ac.uk (N. Golestani).

Dehaene, LeBihan, & Pallier, 2007; Golestani, Paus, & Zatorre, 2002). Second, in order to elucidate the specificity of the learning, we trained the same individuals to learn to hear a pitch difference between steady-state tones, and compared post-training performance on the speech and tonal stimuli. A different pattern of performance between the two stimulus continua would suggest that learning of speech sounds involves different mechanisms than does the learning of a tonal pitch contrast.

A secondary aim was to show training-related improvement on identification and discrimination measures for this 'difficult' Hindi speech contrast. Retroflex consonants require a relatively complex articulation, and are rare across languages; only 11% of the world's languages include a retroflex consonant (Burnham, 1986). Infants under the age of 6 months raised in an English-speaking environment can hear the difference between the retroflex and alveolar sounds, but sensitivity to this difference diminishes during the first year of life (Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Lalonde, 1988). As a result, adult English listeners assimilate the alveolar and retroflex sounds such that they perceive both sounds as instances of the alveolar consonant (Polka, 1991; Rivera-Gaxiola et al., 2000a; Werker & Lalonde, 1988). Training-related improvement has been previously shown using naturalistic versions of this contrast in English speakers, but mainly under conditions that place relatively fewer demands on working memory, for example with the use of an AX discrimination task with a relatively short inter-stimulus interval (ISI), or with the use of truncated stimuli (Pruitt, Jenkins, & Strange, 2006; Pruitt et al., 1990; Werker & Logan, 1985; Werker & Tees, 1984b). Other studies, however, have not successfully shown training-related improvement in the perception of this contrast (Tees & Werker, 1984; Werker & Tees, 1984a; Werker et al., 1981). The use of synthetic stimuli in this study allowed us to control the physical variability within- and between-phonetic categories in order to more precisely characterize the physical characteristics underlying the learning. Also, given that these sounds are synthetic, they were unfamiliar to the participants and thus optimal for use in a learning study. On the other hand, it is known that training with *naturalistic* stimuli (or with stimuli generated by more sophisticated synthesizers than the one we used) which contain within-category acoustic variability is more generalisable to new stimuli (Lively, Logan, & Pisoni, 1993). Therefore, although we validate the synthetic stimuli with native speakers of Indian languages in a pilot study (see below), the generalisability-related implications of our study are limited. Note also that due to the fact that synthetic stimuli do not contain as much acoustic variability as do naturalistic ones, we expect training to be overall easier with our stimuli than that which we would expect to find had we used naturalistic ones.

We predicted that given our large sample size, even a limited amount of training would result in an overall improvement in the ability to identify and discriminate the dental–retroflex contrast. Given that we employed identification training, an improvement in discrimination performance after training would provide evidence for generalization of training to a new task, that being AX discrimination. We also predicted that there would be a large variability in performance across participants, and that performance on identification tests would predict performance on discrimination tests, and vice versa. Last, we predicted that we might find different patterns of performance on post-training measures of speech sound and tone perception.

2. Pilot study: Dental–retroflex stimulus validation

2.1. Materials and methods

2.1.1. Participants

Participants for the pilot study included eight students (3 female and 5 male), ranging in age from 22 to 31 years. Participants spoke Hindi (5 participants), Urdu (2 participants), and Gujarati (1 participant) as a first language (i.e., natively), and several also spoke other Indian languages that employ the dental–retroflex contrast (e.g., Punjabi, Pashtoo, Saraki, and Bengali). All also spoke English fluently. All participants gave informed written consent to participate in the study, which was approved by the regional ethical committee.

2.1.2. Stimulus synthesis

Only two published studies, to our knowledge, have synthesized a dental–retroflex contrast, that being the voiced, unaspirated one (Stevens & Blumstein, 1975; Werker & Lalonde, 1988; Lisker, 1985, unpublished, cited in Werker & Lalonde, 1988). We therefore also decided to use this particular retroflex stop as our non-native sound, and used the previously published parameters as a starting point for synthesizing our stimuli. We used different speech synthesis software than did Stevens and Blumstein (1975) or Werker and Lalonde (1988), and therefore the acoustic parameters that we used are somewhat different than those reported previously. Four-formant stimuli were constructed with the use of the Mitsyn (Henke, 1990) signal processing language software and the Klatt model synthesizer (Klatt, 1980). The continuum involved seven stimuli varying in equal steps in terms of the acoustic difference between adjacent items. Stimulus 1 corresponded to the dental and stimulus 7 to the retroflex stop consonant prototype, and all of the consonants were followed by the vowel /a/.

The stimuli began with an initial noise burst; this latter had a bandwidth of 8 kHz (low-pass 8 kHz passive Tchebychev filters of 142 dB/octave) and an exponential off-set. The parameters that were manipulated to create the continuum were the frequency glides of the third formant (F3), as well as the center frequency of the burst. The initial noise burst lasted 5 ms, and its centre frequency decreased in equal step sizes of 217 Hz from 4500 Hz (stimulus 1) to 3198 Hz (stimulus 7). The Voicing began 15 ms after the

onset of the burst. Each stimulus lasted 220 ms in total, and formant transitions (FTs) lasted 40 ms. The fundamental frequency (F0) started at 130 Hz and decreased gradually over the first 40 ms to 125 Hz, then remained steady-state at 125 Hz for 75 ms, and then gradually decreased to 90 Hz over the last 90 ms of the sound. The first formant (F1) rose from 400 to 655 Hz during the transition, and then remained steady-state during the remainder of the sound. The second formant (F2) decreased from 1650 to 1185 Hz over the first 40 ms, and then remained steady-state. The fourth formant (F4) remained steady-state at 3600 Hz throughout the entire sound. The starting frequency of the third formant (F3) varied in equal 111-Hz steps from 3080 Hz (stimulus 1) to 2414 Hz (stimulus 7), and the steady-state portion of this formant was always at 2585 Hz. Refer to Fig. 1a for spectrograms of the dental and retroflex consonant–vowel (CV) syllables.

The stimuli can be listened to at the following web location: http://www.zlab.mcgill.ca/home.html, under 'Supplements' and then 'Language–anatomy'.

2.1.3. Procedure

Participants were required to qualitatively describe and also to identify (label) the sounds from the dental–retroflex continuum. It has been suggested that there is a continuum of places of articulation for the retroflex sound across languages of India, in other words, that the acoustic boundary between the dental and retroflex sounds is quite variable across languages of India (Ladefoged & Bhaskararao, 1983). The native Indian-language speakers that we tested had different language backgrounds, and so it can be expected that the dental–retroflex boundary might be quite variable across participants. The cross-language variability in the articulation of the retroflex sound has been described by Ladefoged and Bhaskararao (1983), in which it can be seen that for example in Hindi, the articulation of the retroflex sound is more post-alveolar than truly retroflex, and therefore probably more subtle than in other Indian languages. For this reason, native identification performance was evaluated not only at the endpoints but also along the entire 7-step continuum.

2.2. Results

The pilot study showed that six out of eight native Indian-language speakers initially qualitatively described the endpoints as corresponding to the dental /da/ versus the retroflex /da/ voiced, unaspirated sounds. The remaining two described sound 1 as corresponding to the dental sound, and sound 7 as corresponding to the retroflex /da/
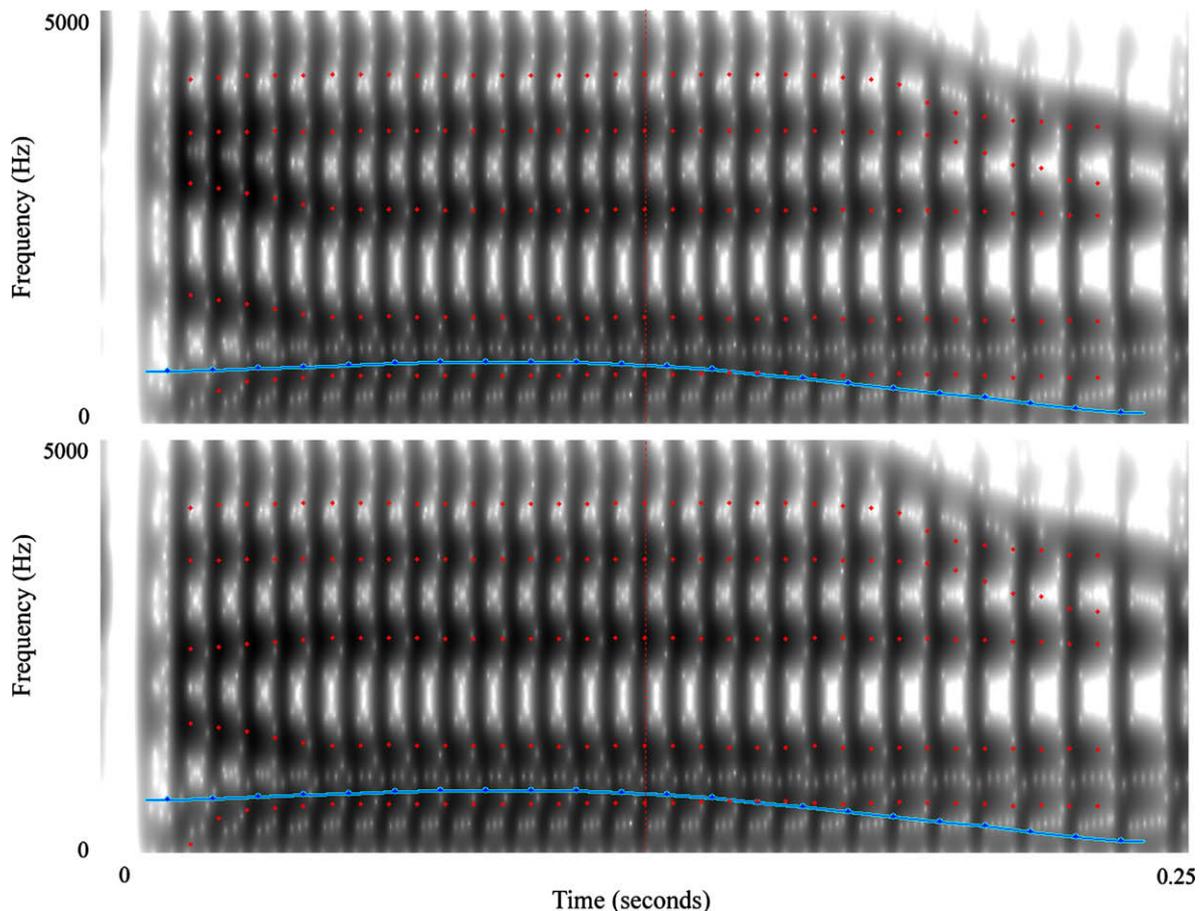

Fig. 1. (a) Spectrograms of the dental and retroflex consonant–vowel (CV) syllables. (b) Schematized frequency–time representations endpoint tonal stimuli.

voiced, aspirated sound. Identification performance was evaluated by requiring participants to label the seven sounds from the synthetic continuum (see Section 3.1.4 for details of testing method).

Results showed that all participants had identification slopes that were in the correct direction (i.e., they labeled sounds towards the dental end of the continuum (e.g., stimuli 1, 2, and 3) as 'A' more frequently, and they labeled sounds towards the retroflex end of the continuum (e.g., stimuli 5, 6, and 7) as 'B' more frequently. The mean identification slope was −0.72, with a standard error of 0.13. However, as expected, there was considerable variability in the location of the dental–retroflex boundary across participants (e.g., for some participants it was between stimuli 4 and 5, whereas for others it was between stimuli 5 and 7, etc.). We therefore examined the percent correct identification performance on a subject-by-subject basis, by examining performance at the stimulus on which each participant most often reported the sound in question as corresponding to the dental sound, and at the stimulus on which he/she least often reported the sound as corresponding to the dental sound. We did this rather than examining performance on the endpoints because performance on the endpoints was not always optimal; in other words, for some participants, it appears that the prototypical 'dental' sound was stimulus 2 rather than 1, and likewise that the prototypical 'retroflex' sound was stimulus 6 rather than 7. Results revealed that on average, the sounds at the dental end of the continuum were correctly labeled as dental 80% of the time, and that sounds at the retroflex end of the continuum were correctly labeled as retroflex 82.5% of the time. The results support the validity of the synthetic stimuli since most native speakers reliably identified sounds from the dental and retroflex ends of the continuum.

3. Training study

3.1. Materials and methods

3.1.1. Participants

Participants for the training study included 59 healthy right-handed adult volunteers (33 male and 26 female), screened to exclude those having learned or had early exposure to languages that employ retroflex sounds contrastively. Many were native English speakers, whereas the others spoke English as a second language, and a subset also spoke one or more additional languages. We performed standard pure-tone audiometry in all participants prior to beginning the experiment, and found that no subject had a significant hearing loss. Linguistic experience was also evaluated before testing began with a detailed questionnaire that took into account number of languages spoken, the level of proficiency with which each language was spoken, and the age at which each language was learned. This information was then used to derive a continuous 'language experience' score, which was based on all of the information provided (i.e., it weighed all three factors).

All participants gave informed written consent to participate in the study, which was approved by the regional ethical committee.

3.1.2. Stimuli
3.1.2.1. Dental–retroflex stimuli. We used the stimuli that were synthesized and validated in the pilot study.

3.1.2.2. Tonal stimuli. The steady-state pure tonal stimuli all lasted 220 ms, and had onset and offset times of 5 and 20 ms, respectively. The acoustic parameter that was manipulated was the frequency of the tones. The continuum consisted of seven tones at frequencies which differed in 7-Hz steps, and ranged from 2000 Hz (stimulus 1) to 2042 Hz (stimulus 7). Refer to Fig. 1b for schematized frequency–time representations of the endpoint tonal stimuli.

3.1.3. Procedure
Testing and training procedures were implemented using MAPLE (Achim, Bregman, & Ahad, 1995). Sounds were presented binaurally through headphones at an intensity level of 70 dB A-weighted SPL in a sound proof room (single-wall industrial acoustic test chamber).

The training study involved testing and training English speakers using identical testing and training procedures with both stimulus types, the order of which was counterbalanced across participants. Testing and training for each stimulus type was *always* done on the same day, however, for reasons of scheduling and availability, some participants were tested and trained with both stimulus types on a single day, whereas others were on different days. For each stimulus type, participants were first familiarized with the sounds by hearing 20 instances of each of the two endpoint stimuli, randomly presented. As each stimulus was presented, participants saw its corresponding label on the computer screen, "A" for one endpoint sound and "B" for the other. In the speech sound condition, "A" corresponded to the dental sound, and "B" to the retroflex one. In this condition, participants were then asked to describe what A and B sounded like to them. Typically, they produced the sound(s) that they heard, and then used words to describe subtle qualitative differences which they may not have been able to accurately produce.

Pre-training identification and discrimination testing (described below) were then conducted. Participants were then trained, using the adaptive training procedure described below, and identification and discrimination tests were re-administered after training.

3.1.3.1. Identification testing. For each stimulus type, participants were presented with one sound at a time from the 7-step stimulus continuum and asked to label each of these sounds by pressing either of two mouse buttons, one corresponding to sound "A", and the other to sound "B". Each of the seven stimuli was presented 10 times in a random sequence, for a total of 70 trials. During pre-training testing, it was difficult for most participants to dis-

tinguish even the endpoint sounds. In the speech sound condition, they were therefore encouraged to think back to the descriptions of A and B that they had provided during familiarization.

3.1.3.2. Discrimination testing. For each stimulus type, participants were tested using an AX discrimination procedure, during which they were to decide whether pairs of sounds were the "same" or "different". The inter-stimulus interval (ISI) was 500 ms. Thirteen possible pairs were presented ten times each, in a randomized order, for a total of 130 trials. The 13 pairs included seven "same" pairs, five 2-step "different" pairs (1 versus 3, 2 versus 4, 3 versus 5, 4 versus 6, and 5 versus 7), and one 6-step "different" pair (1 versus 7).

3.1.3.3. Training. For each stimulus type, training involved 20-trial blocks of identification with feedback. We implemented a "perceptual fading" training method (Jamieson & Morosan, 1986; Morosan & Jamieson, 1986; Terrace, 1963), in which we progressively reduced the acoustic difference between the endpoint sounds as a function of successful performance. This approach allows participants, in cases where they learn to distinguish the endpoint stimuli early during training, to then try to distinguish sounds which are a smaller step-size apart. This training paradigm is likely to help individuals attend to the acoustic differences that critically distinguish the end-point stimuli. For the Hindi stimuli, performance over the course of training itself provided us with a "learning" measure ('*L*-scores', see below), in addition to the pre- and post-training identification and discrimination measures.

Training began with the identification of the endpoint stimuli 1 and 7. Participants heard one stimulus per trial, made a response, and received feedback. If and when an individual achieved criterion performance, defined as at least 16/20 correct responses, on any block, stimuli 2 versus 6 were used for the next training block. Again, if and when criterion was reached with this slightly more difficult pair, stimuli 3 versus 5 were used for the next training block. Training was discontinued either once a participant achieved criterion on this last contrast, or once they had completed a maximum of 200 trials (10 blocks), and lasted twenty to thirty minutes per stimulus type. This is a short training session compared to the amount of training that is typically administered in other phonetic training studies, where it ranges from 1 to 18 training sessions lasting up to an hour each (Carney, Widin, & Viemeister, 1977; Golestani & Zatorre, 2004; Pisoni, Aslin, Perey, & Hennessy, 1982; Strange & Dittmann, 1984).

3.2. Results

3.2.1. Qualitative reports for the dental–retroflex stimuli

After the familiarization with the Hindi stimuli, participants were asked to report what they heard. All reported hearing the alveolar stop phoneme followed by the vowel /a/, or the /da/ syllable, for both endpoints of the dental–retroflex continuum. Some reported that the retroflex sound seemed 'softer' than the dental one.

3.2.2. Identification performance

3.2.2.1. Dental–retroflex contrast. Fig. 2a presents the mean percentage of "A" responses for each of the seven stimuli during pre- and post-training identification performance. As mentioned above, responses "A" and "B" correspond to the dental and retroflex sounds, respectively. For stimulus 1, the range in this value was 10–100% both at pre- and
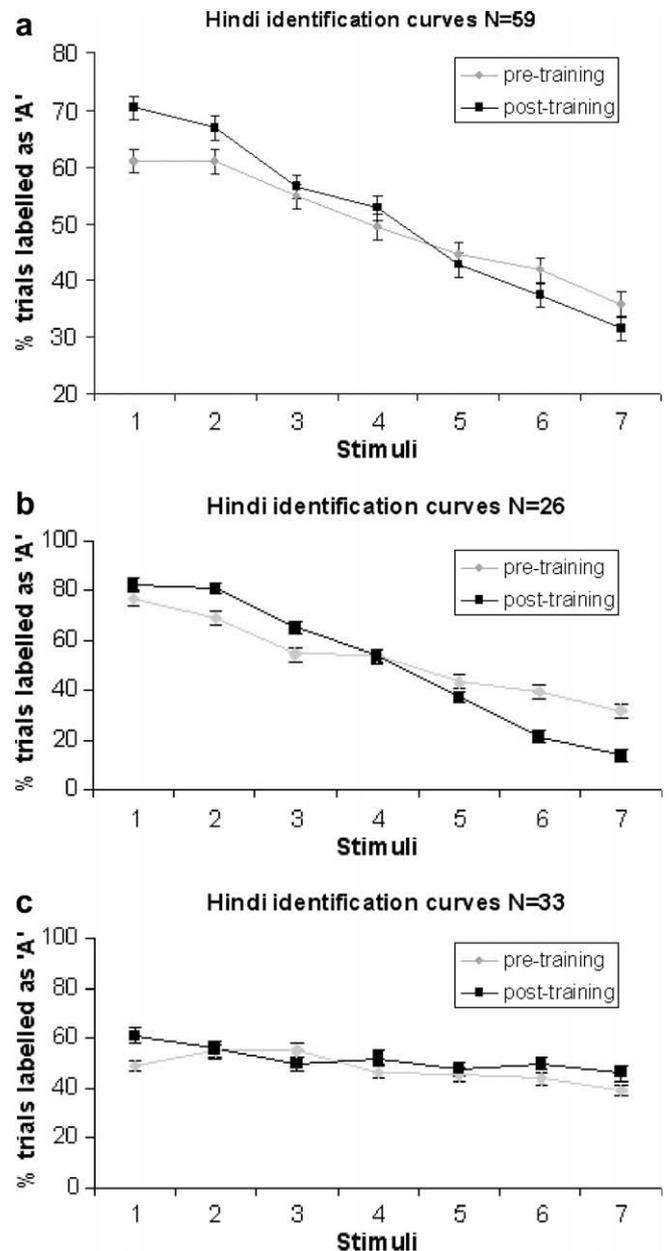

Fig. 2. Pre- and post-training Hindi identification curves of (a) all 59 participants, (b) subgroup of 26 participants who improved in their identification performance after training, and (c) subgroup of 33 participants who did not improve in their identification performance after training.

at post-training, and for stimulus 7, it was 0–80% before training, and 0–90% at post-training.

The slopes of individual participants' identification curves were calculated using a standard regression approach. Relatively steeper negative slopes reflect a better ability to reliably identify the sounds. The mean identification slope before training was –0.44, with a standard deviation of 0.57 and a range of 0.82 to –1.71, and that after training was –0.68, with a standard deviation of 0.65 and a range of 0.57 to –1.89. Note that while probit analyses are sometimes performed on identification data, such a transformation did nothing to clarify the pattern of results, nor did it negate reported results. Correlational analyses revealed that pre- versus post-training slopes were significantly correlated across participants ($r = .50$, $p < .001$), suggesting that individuals who performed well before training were also likely to do so after training. Most importantly, a paired t-test revealed that post-training identification slopes were steeper than pre-training ones ($t(58) = 2.91$, $p < .01$), reflecting improved performance after training. Also, linguistic experience did not predict pre- ($r = .16$, $p > .05$) nor post-training ($r = .12$, $p > .05$) identification performance.

In order to characterize individual differences in identification performance, we evaluated the number of participants who were better than chance at labeling the sounds. This was done by performing regression analyses on post-training identification scores across the stimulus continuum. Out of the 59 participants, 26 showed identification slopes differing significantly from chance (i.e., slope of zero) in the correct direction (negative slope). Refer to Figs. 2b and c for the pre- and post-training identification curves of the sub-groups of 26 participants who performed better than chance after training and of the remaining 33 who did not improve, respectively. It can be seen that qualitatively, in the subgroup of 26 participants who improved (Fig. 2b), pre- and post-training identification curves have negative slopes, and as expected, that this slope is steeper after compared to before training. In contrast, in the subgroup of 33 participants who did not improve (Fig. 2c), identification curves seem relatively flat both before and after training.

3.2.2.2. Tonal contrast. Fig. 3 shows pre- and post-training identification curves for performance on the tonal stimuli. Correlational analyses revealed that pre- versus post-training slopes were significantly correlated across participants ($r = .46$, $p < .001$), suggesting that individuals who performed well before training were also likely to do so after training. A paired t-test revealed that post-training identification slopes were steeper than pre-training ones ($t(58) = 5.19$, $p < .001$), reflecting improved performance after training. Again, linguistic experience did not predict pre- ($r = .11$, $p > .05$) nor post-training ($r = -.01$, $p > .05$) identification performance.

3.2.3. Discrimination performance
3.2.3.1. Dental–retroflex contrast. Each participant's performance was converted to an A' score, an index of sensi-
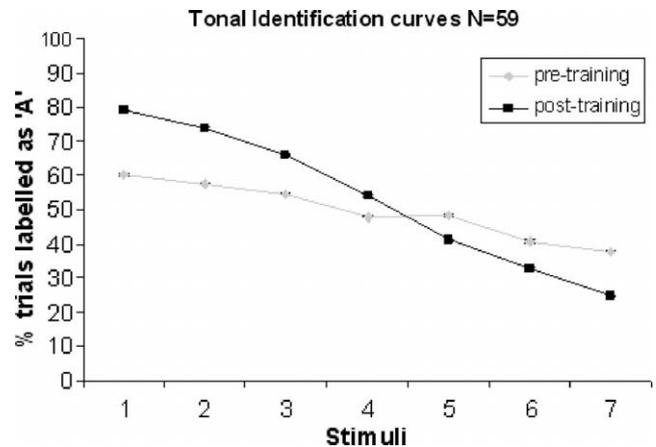

Fig. 3. Pre- and post-training tonal identification curves.

tivity which corrects for individual differences in bias. A' is a non-parametric analog to d', and is used instead of d' when the number of observations is small. Values range from 0 to 1.0, with 0.5 corresponding to chance performance (McNichol, 1972). The mean A' score for performance on the prototypical endpoint stimuli before training was 0.71, with a standard deviation (SD) of 0.19 and a range of 0.06–0.99, and that after training was 0.75, with a standard deviation of 0.19 and a range of 0.15 to 0.99. Correlational analyses revealed that pre- and post-training performance was significantly correlated across participants ($r = .46$, $p < .001$), again suggesting that individuals who performed well before training were likely to do so after training. A paired t-test revealed that there was a significant improvement in performance on the end-point stimuli after compared to before training ($t(58) = 1.85$, $p < .05$, 1-tailed). Linguistic experience did not predict pre- ($r = -0.21$, $p > .05$) nor post-training ($r = -.02$, $p > .05$) discrimination performance.

There was a significant correlation between pre-training identification (i.e., identification slopes) and discrimination (i.e., A' scores on endpoint stimuli) performance ($r = -.39$, $p < .005$), indicating that better performance on the pre-training identification task predicts better performance on the pre-training discrimination task, and vice versa. Likewise, there was a significant correlation between post-training identification and discrimination performance ($r = -.54$, $p < .001$), indicating that better performance on the post-training identification task predicts better performance on the post-training discrimination task, and vice versa. There was also a correlation between post- minus pre-training identification slopes and post-minus pre-training discrimination A' measures ($r = -.51$, $p < .001$), indicating that the bigger the improvement from pre- to post-training on the identification task, the bigger the improvement from pre- to post-training on the discrimination task, and vice versa.

A' scores were also calculated on 2-step pairs from the stimulus continuum (i.e., stimuli 1/3, 2/4, 3/5, 4/6, and 5/7). In order to better characterize individual differences in

performance, we examined discrimination performance for different participant subgroups. Specifically, we examined the discrimination curves (a) in all 59 participants, (b) in the subgroup of 26 individuals who showed evidence of better-than-chance performance after training on an independent measure of performance: identification slopes (see above), and (c) in the subgroup of 33 individuals who did not show perform better than chance on this identification measure. Refer to Fig. 4a for the pre- and post-training 2-step discrimination curves of these subgroups of participants. It can be seen that qualitatively, for all 59 participants taken together (see Fig. 4a), the shape of the post-training curve does not show much change. However, when data from the subgroup of 26 participants who perform better than chance on the post-training identification task is analyzed separately (see Fig. 4b), overall discrimination performance is better (the A' scores are higher), and, there appears to be a boundary effect in the middle of the stimulus continuum after training. Consistent with this, trend analyses showed a significant quadratic trend in the post- ($F(1, 25) = 4.31$, $p < .05$) but not in the pre-training ($F(1, 25) = 2.69$, $p > .05$) results of this latter subgroup of participants.

3.2.3.2. Tonal contrast.
Fig. 5a shows pre- and post-training discrimination curves for performance on the tonal stimuli for all 59 participants. Correlational analyses revealed that pre- and post-training performance was significantly correlated across participants ($r = .50$, $p < .001$), again suggesting that individuals who performed well before training were likely to do so after training. A paired t-test revealed that there was a significant improvement in performance on the endpoint stimuli after compared to before training ($t(58) = 2.38$, $p < .01$). Linguistic experience did not predict pre- ($r = .06$, $p > .05$) nor post-training ($r = .02$, $p > .05$) discrimination performance. Fig. 5b presents the discrimination curves for the 32 participants whose post-training tone identification slopes differed from zero (i.e., who showed evidence of improvement in the identification of the tonal contrast). Similarly, Fig. 5c presents the discrimination curves for the 27 participants whose post-training tone identification slopes did not differ from zero (i.e., who did not show evidence of improvement in the identification of the tonal contrast). Visual inspection reveals that there is no boundary effect in the middle of the stimulus continuum in either of these subgroups after training. Consistent with this, trend analyses showed that there is no significant quadratic trend in the post-training results of either the 32 tonal 'improvers' ($F(1, 31) = 0.64$, $p > .05$) nor in the 27 tonal 'non-improvers' ($F(1, 26) = 0.36$, $p > .05$). Interestingly, the overlap between those individuals who improved on the Hindi contrast and those who improved on the tonal one was not greater than might be expected by chance ($\chi^2 = 2.28$, $p > .05$).

3.2.4. Training results
Participants received an average of 8.8 (SD = 2.1) blocks (i.e., 176 trials) of training with the Hindi stimuli,
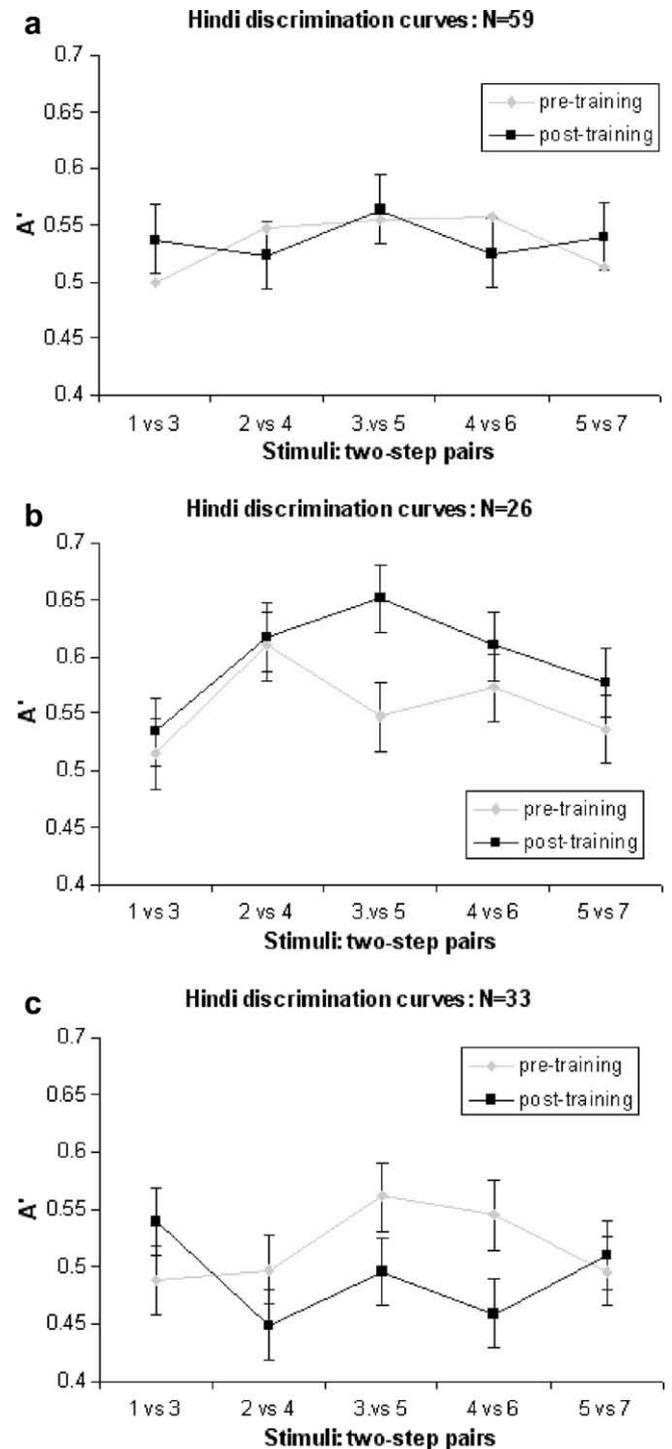

Fig. 4. (a) Pre- and post-training 2-step Hindi discrimination curves of (a) all 59 participants, (b) subgroup of 26 participants who improved in their identification performance, and (c) subgroup of 33 participants who did not improve in their identification performance.

and an average of 7.6 (SD = 2.9) blocks (i.e., 152 trials) of training with the tonal stimuli. For the Hindi stimuli, 21 out of 59 individuals achieved criterion performance on the third, most difficult training level in 200 or fewer trials (i.e., in 10 blocks or less), and eight did so in 5 blocks or less. Given that we discontinued training once criterion was
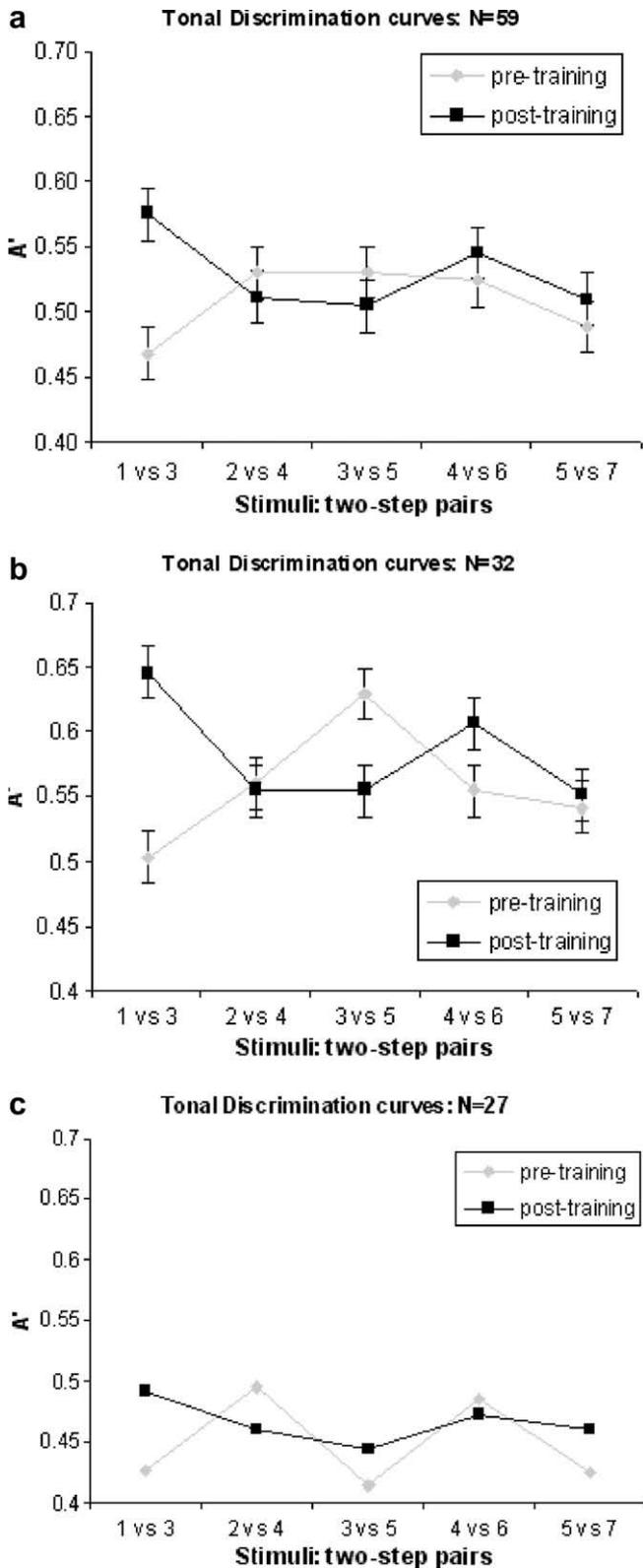
Fig. 5. (a) Pre- and post-training 2-step tonal discrimination curves of (a) all 59 participants, (b) subgroup of 32 participants who improved in their tonal identification performance, and (c) subgroup of 27 participants who did not improve in their tonal identification performance.

achieved on the most difficult training level, participants who received the fewer training blocks were the ones who performed better. The total number of training blocks that

a participant could receive ranged from 3 (since this was the minimum that was administered) to 10 (this was the maximum that was administered). Also, recall that training level 1 was the easiest training level, involving identification of the endpoint stimuli (i.e., 1 versus 7), that level 2 was the intermediate training level, involving identification of stimuli that were one step-size closer (i.e., stimuli 2 versus 6), and that level 3 was the most difficult training level, involving identification of stimuli that were the most similar (i.e., stimuli 3 versus 5). The 'fastest learners' were therefore those who performed well enough on training levels 1 and 2 to be trained on level 3, and the 'slowest learners' remained on training level 1 throughout the training session. Fig. 6 presents the distribution of the participants with respect to the total number of training blocks received. Likewise, Fig. 7 presents the distribution of the participants with respect to the maximum 'training level' achieved. The mean performance accuracy was 14.6/20 ($N = 59$, SD = 2.7), 15.2/20 ($N = 48$, SD = 2.2), and 13.5/20 ($N = 36$, SD = 1.7) on training levels 1, 2, and 3, respectively.

We developed a continuous learning rate variable ("L-score") to quantify participants' performance during training. This variable takes into account both the number of blocks at each level of difficulty and the accuracy of performance on each of the blocks. L-scores were obtained by first weighing the number of incorrect responses in each block by a "difficulty" weight corresponding to that block. The three possible weights are 3, 2, and 1, for the easiest, intermediate, and most difficult blocks, respectively. Resulting values were then subtracted from 600, which was the maximum possible value, yielding an "L-score". The L-scores thus ranged from zero (slowest possible learner) to 600 (fastest possible learner). Fig. 8 shows the distribution of the L-scores, illustrating the degree of variability in performance during training. Linguistic experience did not predict learning rate scores ($r = -0.1$, $p > .05$).

For the Hindi stimuli, we found the following relationships between performance on training and test measures. Please refer to Table 1 for a summary of these analyses
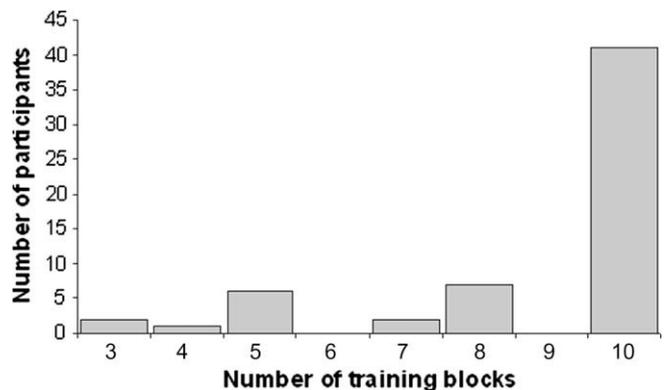

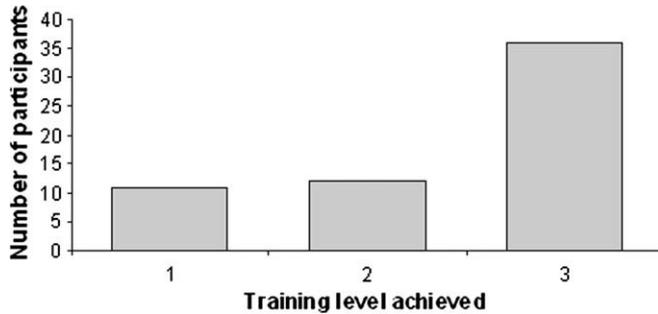Fig. 6. distribution of the participants with respect to the total number of training blocks received.

Fig. 7. Distribution of the participants with respect to the maximum 'training level' achieved.
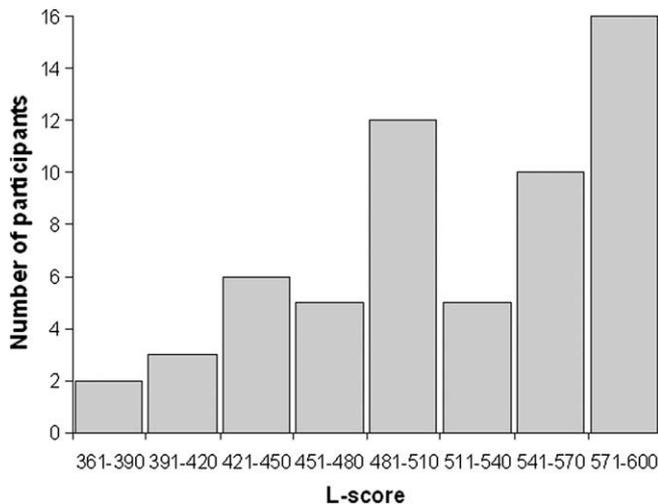

Fig. 8. Distribution of the *L*-scores.

We found a similar but weaker relationship between pre- ($r = -.40$, $p < .005$) and post-training ($r = -.50$, $p < .001$) identification slopes and performance on the intermediate training level (level 2, $N = 36$). We found a relationship between pre-training discrimination performance (A' scores on endpoint stimuli) and the mean performance on training level 1 ($r = .31$, $p < .05$), suggesting that better pre-training discrimination performance better predicts performance on training level 1. There was a similar trend in the same direction between post-training discrimination measures and performance on training level 1 ($r = .24$, $p = .07$). Post-training discrimination performance was, however, significantly related to performance on training level 2 ($r = .31$, $p < .05$), suggesting that better performance on training level 2 predicts better post-training discrimination performance. Pre- ($r = -.41$, $p < .001$) and post-training ($r = -.55$, $p < .001$) identification slopes were correlated with the total number of training levels that individuals achieved, suggesting that individuals who perform better on identification tests are more likely to do well enough during training to achieve training level 3 (the most difficult). Likewise, pre- ($r = .49$, $p < .001$) and post-training ($r = .41$, $p < .001$) identification slopes were correlated with the total number of training blocks performed, suggesting that individuals who perform better on identification tests are more likely to require fewer training blocks (and are also therefore more likely to achieve criterion performance at least on level 1, and maybe also on levels 2 and 3). See Fig. 10 for scatter plot illustrating this relationship at pre-training.

4. Discussion

We show that a limited amount of training results in an overall group improvement in the ability to identify and to discriminate the non-native dental–retroflex contrast. Given that training was done using an identification task, our results show transfer of training to a new task, that being AX discrimination. Also, consistent with this, there was evidence for a relationship between the amount of improvement on identification and discrimination tasks across participants, showing that people who improve on one are likely to improve on the other, and vice versa.

and results. First, we found that pre-training ($r = -.62$, $p < .001$) and post-training ($r = -.61$, $p < .001$) identification slopes were correlated with the mean performance accuracy on training level 1 ($N = 59$), suggesting that individuals who were better at identifying the endpoint stimuli before and after training were more likely to have higher identification scores when tested on the easiest training level. Fig. 9 shows the scatter plot of the relationship between pre-training identification performance and performance on easiest training level.

Table 1
Summary of correlations pre- and post-training test measures and training measures (ns means not significant)

| Test measures	Training measures			
	Mean performance on training level 1	Mean performance on training level 2	Total number of training blocks performed	Total number of training levels achieved
Pre-training identification slope	$r = -0.62$, $p < 0.001$	$r = -0.40$, $p < 0.005$	$r = 0.49$, $p < 0.001$	$r = -0.41$, $p < 0.001$
Post-training identification slope	$r = -0.61$, $p < 0.001$	$r = -0.50$, $p < 0.001$	$r = 0.41$, $p < 0.001$	$r = -0.55$, $p < 0.001$
Pre-training discrimination A' score	$r = 0.31$, $p < 0.05$	ns	ns	ns
Post-training discrimination A' score	$r = 0.24$, $p = 0.07$	$r = 0.31$, $p < 0.05$	ns	ns
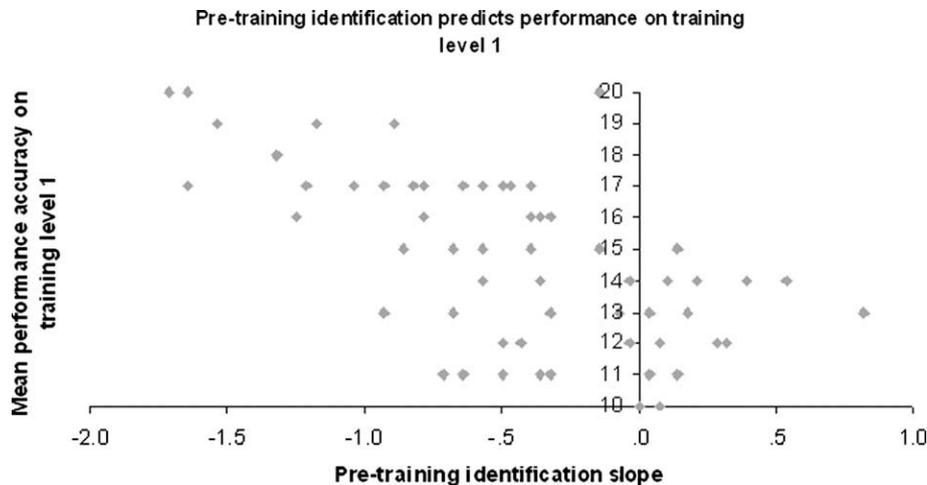
Fig. 9. Relationship between pre-training identification performance and performance on easiest training level.
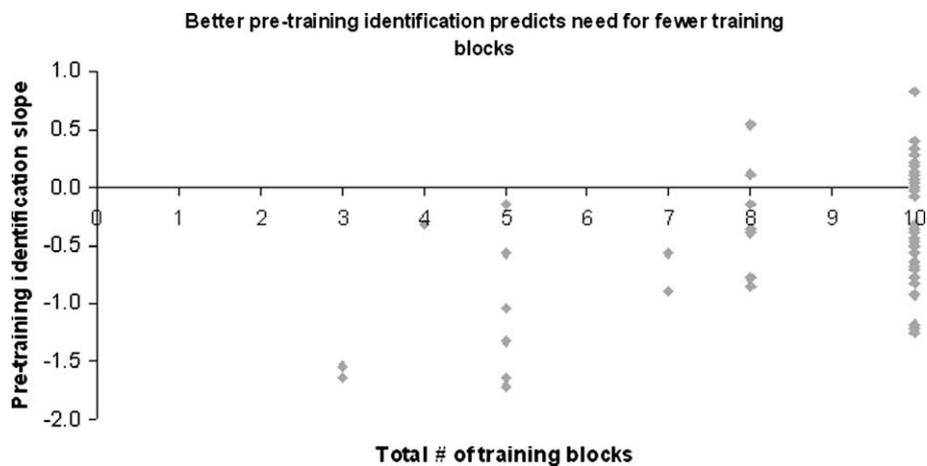

Fig. 10. Relationship between pre-training identification performance and total number of training blocks performed.

Previous laboratory studies have shown that training using naturalistic exemplars is effective under task or stimulus conditions that place relatively lesser demands on working memory (Pruitt et al., 1990; Pruitt et al., 2006; Werker & Tees, 1984b). Other studies, however, have not successfully shown a training-related improvement in the perception of this contrast (Tees & Werker, 1984; Werker & Tees, 1984a; Werker et al., 1981). We show group improvement after only a short training session. We show such improvement even on an identification task with non-truncated stimuli. Labeling (i.e., identification) tasks are known to place greater demands on working memory than discrimination ones. During speech sound classification, participants only hear one sound at a time and therefore need to be able to maintain and have access to a long-term memory representation of the prototypical sounds for accurate performance, whereas during a discrimination task, they likely compare the two sounds presented on each trial to each other. Our large sample size may have contributed to our finding an overall group improvement on both discrimination and identification tasks. In addition, the perceptual fading

training procedure as well as the use of synthetic stimuli, which contain little within-category acoustic variability, may have facilitated learning by helping participants to attend to the relevant acoustic cues. The implications of our results with respect to generalization to natural speech sounds are limited since these latter contain more complex acoustic information than do our synthetic ones, and in processing these, one must therefore be able to disregard within-category variability when hearing different instances of the same phoneme while being able to discriminate between categories when hearing different phonemes.

As expected, we also found considerable variability in both pre- and post-training performance as well as in performance during training across participants. Indeed, only about half of the participants showed significant learning. Linguistic experience did not predict any of the pre-test, post-test, or training measures, suggesting that such individual differences may not be explained by differences in language background. This variability is consistent with the results of previous studies on non-native speech sound learning (Bradlow et al., 1997; Golestani & Zatorre, 2004;

Polka, 1991; Pruitt et al., 1990). We have previously shown a large range in performance in the learning of the dental-retroflex contrast in native English speakers in a functional imaging study in which participants received five 1-h training sessions, spread over the course of two weeks on non-consecutive days. The training procedure was the same as that used in this study, apart from the fact that it was much longer. We found that only half of the participants improved over the course of this much longer training (Golestani & Zatorre, 2004). Further, in two, related anatomical magnetic resonance imaging (aMRI) studies, it was found that people who are faster at learning to hear the dental–retroflex contrast have a greater left > right asymmetry in parietal lobe white matter volumes (Golestani et al., 2002), as well greater white matter volumes of left Heschl's gyrus, a part of the brain which includes primary auditory cortex (Golestani et al., 2007) than do slower learners. The left inferior parietal cortex has previously been shown to be involved in aspects of phonological processing (Démonet, Price, Wise, & Frackowiak, 1994; Démonet et al., 1992; Zatorre, Evans, Meyer, & Gjedde, 1992; Zatorre, Meyer, Gjedde, & Evans, 1996), and in particular in the storage of phonological information in verbal short-term memory (Jonides et al., 1998; Paulesu, Frith, & Frackowiak, 1993). Taken together, these findings suggest that there may be white matter connectivity differences between 'faster' and 'slower' phonetic learners, which could in turn either be due to greater myelination (Anderson, Southern, & Powers, 1999) or to a greater number of white matter fibers, between the auditory and parietal cortices or between these regions and other anatomically connected ones. Greater myelination would likely result in more rapid neural transmission between connected areas, which is likely critical for the processing of certain speech sounds such as stop consonants which contain very rapidly changing acoustic information. This connectivity hypothesis is currently being tested using diffusion tensor imaging (DTI) data, which allows to 'track' white matter fibers between brain regions, and to quantify connectivity differences between groups.

For the dental–retroflex contrast, we found a relationship between performance on the identification and discrimination tasks, suggesting that people who do well on one tend to do well on the other, and vice versa. The relationship was stronger after relative to before training, likely reflecting correlated training-related improvement on the two tasks across participants (i.e., less noisy performance after relative to before training). We also found relationships between performance during training and on test measures suggesting that individuals who were better at identifying and discriminating the sounds before and after training were more likely to perform more accurately during training, and/or to need fewer training blocks. Interestingly, we found evidence that the subset of 26/59 participants who showed successful learning of the Hindi stimuli on an identification task had qualitatively different discrimination curves than those who did not perform bet-

ter than chance on the identification task. Specifically, the former but not the latter group showed greater "boundary effects" in the middle of the stimulus continuum. In contrast, for the tonal stimuli, there was no evidence for boundary effects in the post-training discrimination data of the 32 participants showing improvement in post-training identification of this latter stimulus type. This is the case despite the fact that the amount of overall learning of the tonal contrast was at least as great as that seen on with Hindi contrast. In addition, individuals who improved on the Hindi stimuli were not the same as those who improved on the tonal stimuli, suggesting a dissociation between performance on the two stimulus types also during learning itself. The dissociation between the patterns of performance for the two stimulus types suggests that different mechanisms are involved in the learning of speech sounds compared to that of tonal, pitch differences. Specifically, the pattern of learning of speech sounds suggests the formation of categories, as opposed to some more general perceptual learning which would be expected to be more continuous, as is seen with the tonal stimuli. Our results suggest that those who successfully learned the Hindi contrast were the ones who managed to create some kind of long-term memory category representation for the native (dental) and non-native (retroflex) sounds. Note, however, that the tonal stimuli that we used were acoustically simpler than were the speech stimuli. Thus, the above dissociations that we find in performance with the two stimulus types may be due not only to the fact that one is speech and the other not, but they may also be due to stimulus complexity. In support of the idea that different learning mechanisms exist for speech and non-speech sounds, there exists evidence from other studies for speech-specific neural mechanisms when stimulus complexity is controlled (Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Scott, Blank, Rosen, & Wise, 2000; Scott, Rosen, Lang, & Wise, 2006), or when the *same* stimuli are first perceived as non-speech and then later as speech (Dehaene-Lambertz et al., 2005; Dufor, Serniclaes, Sprenger-Charolles, & Démonet, 2007). More generally, it is likely that the neural mechanisms underlying the processing of abstract, linguistically relevant properties versus of the underlying acoustic characteristics of stimuli interact in a complex and non-exclusive manner, and that they depend on linguistic experience as well as on neural top–down processing mechanisms which interact with afferent pathways which carry stimulus information (Zatorre & Gandour, 2007).

We like to advance some ideas that can be tested in future research to explain possible contributions to individual differences in non-native speech sound learning. It is possible that the observed individual differences on the different tasks are due to individual differences in performance on specific task components which differentially contribute to identification tests, discrimination tests, and training. For example, it is possible that individual differences in working memory capacity differentially influence performance on identification and discrimination tasks

because, as mentioned above, it is known that labeling tasks place greater demands on working memory than do discrimination ones. People may also differ in their ability to create long-term memory representations of the proto-typical sounds, and/or in their ability to later retrieve these representations when classifying or discriminating the sounds. There is evidence for individual differences in the ability to learn with feedback (Cohen, 2007), and in the neurochemical and physiological underpinnings of such learning (Pickering & Gray, 2001); such differences may predict performance during training more than during testing. Last, individual differences could be due to differences in task performance strategy, related for example to differential engagement of top–down (e.g., articulatory) versus bottom–up (e.g., acoustic, stimulus-driven) processing across individuals. Specifically, given that we did not instruct participants on how the non-native, retroflex sound is articulated, it is possible that some individuals may have realized or guessed how to articulate this sound and that this might have assisted perception, while others might not have. Alternatively, it is possible that 'faster' phonetic learners are precisely those who are able to stop hearing certain novel speech sounds in a speech-specific manner, which likely depends in part on how sounds are to be articulated (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985), and who instead are better able to focus on the sounds themselves, and thereby to hear the sounds in a more bottom–up, stimulus-driven mode (i.e., to attend more to the physical/acoustic characteristics of the auditory input). We have functional brain imaging data suggesting that the successful learning of a non-native phonetic contrast results in the recruitment of the same areas that are involved during the processing of native contrasts (Golestani & Zatorre, 2004). These results would support the former rather than the latter of the two above hypotheses, or at least they can help to exclude the latter, since they suggest that the learning of non-native speech sounds results in the recruitment of a speech-specific network. This is an example of how brain imaging data can be used to help to constrain models of perception, rather than the more traditional, opposite approach.

We also speculate that individual differences in performance on these different aspects of processing (e.g., working memory, ability to make use of feedback during learning, strategy differences, etc.) might predict different aspects of brain structure, and that they might arise from individual differences in the critical periods underlying these different processing components. For example, it is possible that the ability to parse rapidly changing sounds such as the stop consonants used in this study might be the highest predictor of left Heschl's gyrus white matter volumes, but it is possible that individual differences in performance on other cognitive dimensions which contribute to task performance might predict other aspects of brain structure. This idea can be tested in future studies on brain morphology–behavior relationships. In line with the idea

that individual differences in specific aspects of task processing might arise from individual differences in the critical periods underlying such processing, Werker and Tees (2005) propose that phonological acquisition involves the emergence of a series of nested capabilities which biology, experience, and functional use contribute to, each with its own sensitive period and each best explained at one of several different levels of specificity (Werker and Tees, 2005).

Taken together, our findings suggest that even if individuals are not phonemically exposed to certain speech sounds at an early age in the native language environment, some individuals can successfully learn these sounds with feedback during training. Further, our findings of greater white matter volumes in left Heschl's gyrus and of greater left > right asymmetry in parietal lobe white matter volumes in faster compared to slower phonetic learners suggest that aspects of brain white matter morphology, and likely of white matter connectivity, are the anatomical characteristics underlying the more efficient neural processing in the speech-dominant hemisphere in some individuals compared to others.

References

Achim, A., Bregman, A. S., & Ahad, P. (1995). *Manager of auditory perception and linguistic experiments*. Montreal: McGill University, Department of Psychology, Auditory Laboratory.

Anderson, B., Southern, B. D., & Powers, R. E. (1999). Anatomic asymmetries of the posterior superior temporal lobes: A postmortem study. *Neuropsychiatry Neuropsychology and Behavioral Neurology, 12*(4), 247–254.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*(3), 345–360.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*(4), 2299–2310.

Burnham, D. K. (1986). Developmental loss of speech-perception—exposure to and experience with a first language. *Applied Psycholinguistics, 7*(3), 207–239.

Carney, A., Widin, G., & Viemeister, N. (1977). Non-categorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America, 62*(4), 961–970.

Cohen, M. X. (2007). Individual differences and the neural representations of reward expectation and reward prediction error. *Social Cognitive and Affective Neuroscience, 2*(1), 20–30.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage, 24*(1), 21–33.

Démonet, J. F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J. L., Wise, R., et al. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain, 115*(Pt. 6), 1753–1768.

Démonet, J. F., Price, C., Wise, R., & Frackowiak, R. S. (1994). Differential activation of right and left posterior sylvian regions by semantic and phonological tasks: A positron-emission tomography study in normal human subjects. *Neuroscience Letters, 182*(1), 25–28.

Dufor, O., Serniclaes, W., Sprenger-Charolles, L., & Démonet, J. F. (2007). Top–down processes during auditory phoneme categorization in dyslexia: A PET study. *Neuroimage, 34*(4), 1692–1707.

Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America, 106*(5), 2973–2987.

Golestani, N., Molko, N., Dehaene, S., LeBihan, D., & Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex, 17*(3), 575–582.

Golestani, N., Paus, T., & Zatorre, R. J. (2002). Anatomical correlates of learning novel speech sounds. *Neuron, 35*(5), 997–1010.

Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: Reallocation of neural substrates. *Neuroimage, 21*(2), 494–506.

Henke, W. L. (1990). *Mitsyn language: A synergistic family of high-level language for time signal processing*. Belmont, MA: Language Processing Manual.

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /delta/-/theta/ contrast by francophones. *Perception & Psychophysics, 40*(4), 205–215.

Jonides, J., Schumacher, E. H., Smith, E. E., Koeppe, R. A., Awh, E., Reuter-Lorenz, P. A., et al. (1998). The role of parietal cortex in verbal working memory. *Journal of Neuroscience, 18*(13), 5026–5034.

Klatt, D. (1980). Software for a cascade-parallel formant synthesizer. *Journal of the Acoustical Society of America, 67*(3), 971–995.

Ladefoged, P., & Bhaskararao, P. (1983). Non-quantal aspects of consonant production: A study of retroflex consonants. *Journal of Phonetics, 11*, 291–302.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*(6), 431–461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1–36.

Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex, 15*(10), 1621–1631.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories.. *Journal of the Acoustical Society of America, 94*(3, Pt. 1), 1242–1255.

McNichol, D. (1972). *A primer of signal detection theory*. London: Allen and Unwin.

Morosan, D. E., & Jamieson, D. G. (1986). Efficient training of non-native speech discrimination with perceptual fading. *Paper presented at the Proceedings of the 12th international congress on acoustics*.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature, 362*(6418), 342–345.

Pickering, A. D., & Gray, J. A. (2001). Dopamine appetitive reinforcement and the neuropsychology of human learning: An individual differences approach. In A. Eliasz & A. Angleitner (Eds.), *Advances in research on temperament* (pp. 113–149). Lergerich, Germany: PABST Science Publishers.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 8*(2), 297–314.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America, 89*(6), 2961–2977.

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *Journal of the Acoustical Society of America, 119*(3), 1684–1696.

Pruitt, J. S., Strange, W., Polka, L., & Aguilar, M. C. (1990). Effects of category knowledge and syllable truncation during auditory training on Americans' discrimination of Hindi retroflex-dental contrasts. *Journal of the Acoustical Society of America, 87*(Suppl. 1), S72.

Rivera-Gaxiola, M., Csibra, G., Johnson, M. H., & Karmiloff-Smith, A. (2000a). Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Behavioral Brain Research, 111*(1–2), 13–23.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123*(Pt. 12), 2400–2406.

Scott, S. K., Rosen, S., Lang, H., & Wise, R. J. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech—a positron emission tomography study. *Journal of the Acoustical Society of America, 120*(2), 1075–1083.

Stevens, K. N., & Blumstein, S. E. (1975). Quantal aspects of consonant production and perception: A study of retroflex stop consonants. *Journal of Phonetics, 3*, 215–233.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics, 36*(2), 131–145.

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology, 38*(4), 579–590.

Terrace, H. S. (1963). Discrimination learning with and without "errors". *Journal of the Experimental Analysis of Behavior, 6*, 1–27.

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development, 52*(1), 349–355.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychobiology, 24*(5), 672–683.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*(1), 35–44.

Werker, J. F., & Tees, R. C. (1984a). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49–63.

Werker, J. F., & Tees, R. C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75*(6), 1866–1878.

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology, 46*(3), 233–251.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256*(5058), 846–849.

Zatorre, R. J., & Gandour, J. T. (2007). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*.

Zatorre, R. J., Meyer, E., Gjedde, A., & Evans, A. C. (1996). PET studies of phonetic processing of speech: Review, replication, and reanalysis. *Cerebral Cortex, 6*(1), 21–30.