Available online at www.sciencedirect.com
ScienceDirect
Brain & Language 109 (2009) 55–67
Brain & Language
www.elsevier.com/locate/b&l

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*

Accepted 14 January 2008
Available online 4 March 2008

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.

Keywords: Phonetic learning; Behavioral plasticity; Language; Individual differences; Non-native; Hindi dental–retroflex contrast

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 articulation, 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).

0093-934X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2008.01.005

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