Phonetic learning abilities:
Behavioral, neural functional, and neural anatomical correlates

Narly A. Golestani
Department of Psychology
McGill University

Montreal, Quebec
Cananda

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of PhD.

Submitted August 2001

0-612-78693-5

Dedication:

This work is dedicated to my family, who have always encouraged me to pursue my interests and passions.

Phonetic learning abilities:
Behavioral, neural functional, and neural anatomical correlates

Table of Contents:

Chapter 3: ***Learning new sounds of speech: Reallocation of neural substrates***

Chapter 4: ***Anatomical correlates of the ability to learn novel speech sounds***

Chapter 5: General Discussion

Abstract

The studies included in this thesis had as aim to elucidate how individual differences in phonetic learning abilities might be related to differences in more general, psychoacoustic learning abilities, and in how they might be related to differences in brain function and brain morphology.

We tested and trained English speaking volunteers to perceive the Hindi dental-retroflex phonetic contrast. We found evidence suggesting that the ability to accurately perceive "difficult" non-native contrasts is not permanently lost during development. We also tested and trained subjects to perceive the difference between non-linguistic rapidly changing and steady-state tonal sounds, and found evidence supporting the hypothesis that successful phonetic learning is in part a function of a more general psychoacoustic ability to process rapidly changing sounds.

The aim of the second study was to determine how the pattern of brain activity may change as a result of training with non-native speech sounds, and in whether it is possible to differentiate "learners" from "non-learners" on the basis of neural activation patterns. Results of this functional magnetic resonance imaging (fMRI) investigation suggested that successful learning of a non-native contrast results in the recruitment of the same areas that are involved in the processing of native contrasts; but the degree of success in learning is accompanied by more efficient neural processing in classical frontal speech regions, while making greater processing demands in left parieto-temporal speech regions.

In the final study, we correlated phonetic learning measures with brain morphology throughout the whole brain volume. We found evidence for overall larger parietal volumes in the left relative to the right hemisphere, and for more white relative to gray matter in the left hemisphere in the learners and not in the nonlearners. This finding is consistent with findings by other investigators suggesting that left-hemispheric dominance for speech may be in part accounted for by hemispheric differences in white matter connectivity, which may allow faster intra- and inter-hemispheric neural transmission. This latter feature may be critical for the processing of consonant speech sounds, which depends on the ability to process sounds that change on the time scale of 30-50 milliseconds.

Résumé

Cette thèse inclut trois études portant sur les corrélats comportementaux et neuronaux de la capacité d'apprendre des contrastes phonétiques venant d'une langue étrangère. Nous avons testé et entraîné des sujets anglophones à perçevoir le contraste phonétique dental-rétroflexe. Les résultats suggèrent que la capacité de percevoir un contraste phonétique "difficile" n'est pas complêtement perdue au cours du développement. Nous avons aussi entraîné les participants à percevoir la différence entre des sons non-linguistiques contenant des changements temporels rapides et des sons tonaux stationnaires. Les données suggèrent que l'apprentissage phonétique est en partie déterminé par une capacité psychoacoustique plus générale d'apprendre des sons contenants des changements temporels rapides.

Le but de la deuxième étude était de déterminer comment l'activité du cerveau peut changer en conséquence d'un entraînement phonétique, et de déterminer si l'activité est différente chez les sujets qui bénéficient d'un entraînement par rapport à ceux qui n'en bénéficient pas. Les résultats de cette étude d'imagerie par résonance magnétique fonctionnelle (IRMf) suggèrent que l'apprentissage de nouveaux contrastes phonétiques recrute les mêmes aires du cerveau que celles qui sont activées pendant la perception des contrastes natifs. Les résultats suggèrent aussi que chez les individus que réussissent à apprendre le contraste, il y une activité cérébrale plus efficace dans les aires frontales linguistiques, et que les aires postérieures linguistiques sont plus actives chez ces mêmes personnes.

Dans la troisième étude, nous avons corrélé des mesures comportementales d'apprentissage phonétique avec la morphologie du cerveau. Nous avons trouvé que le lobe pariétal gauche est généralement plus grand que le droit, et que chez les individus qui réussissent à apprendre les sons non-natifs, il y a plus de matière blanche par rapport à la quantité de matière grise dans l'hémisphère gauche. Ce résultat est intéressant car il suggère que la capacité d'apprendre des nouveaux sons phonétiques peut en partie être détérminée par la connectivité intra- ou inter-hémisphérique entre les aires linguistiques du cerveau. La connectivité peut donc influencer l'efficacité de la communication neuronale, ce qui est critique pour la capacité de percevoir certains phonèmes qui ont des changements acoustiques de très courte durée.

Contributions of Authors of to be published articles

McGill University requires that in the case that papers included in a thesis have been or will be published, as for all three studies in this thesis, an explicit statement be made regarding the contributions of each author.

Chapter 2: *Learning the Hindi dental-retroflex contrast: Phonetic or auditory factors?*

Narly Golestani, Robert J Zatorre, and Pierre Ahad

Narly Golestani:

- Outlined rationale / procedure of study in conjunction with R Zatorre
- Created stimuli with the help of Pierre Ahad
- Wrote stimulus presentation / data collection programs (with P Ahad's help)
- Recruited, scheduled, and tested subjects
- Analyzed data
- Interpreted result in conjunction with R Zatorre
- Wrote paper for journal submission

Robert J Zatorre:

- Participated in outlining rationale / procedure of study
- Participated in interpretation of results
- Proofed drafts of the article

Pierre Ahad

- Participated in the creation of the stimuli
- Provided guidance for writing stimulus presentation and data collection programs

Chapter 3: *Learning new sounds of speech: Reallocation of neural substrates*

Narly Golestani and Robert J Zatorre

Narly Golestani:

- Outlined rationale / procedure of study in conjunction with R Zatorre
- Created stimuli (with the help of P Ahad)
- Obtained approval from MRI working committee / MNI ethics committee
- Developed fMRI scanning protocol (with help of Pascal Belin and R Zatorre)
- Wrote stimulus presentation / data collection programs (with guidance of Pascal Belin)
- Recruited, scheduled, scanned, and tested subjects
- Analyzed data (with guidance of Valentina Petre)
- Interpreted result in conjunction with R Zatorre
- Presented results at conference (poster)
- Wrote paper for journal submission

Robert J Zatorre
- Participated in outlining rationale / procedure of study
- Provided funding for use of fMRI facilities
- Participated in interpretation of results
- Proofed drafts of the article

Chapter 4: *Anatomical correlates of the ability to learn novel speech sounds*

Narly Golestani, Tomas Paus, and Robert J Zatorre

Narly Golestani:
- Outlined rationale / procedure of study in conjunction with T Paus and R Zatorre
- Created stimuli (with the help of P Ahad)
- Wrote stimulus presentation / data collection programs
- Recruited, scheduled, scanned, and tested subjects
- Analyzed data
- Interpreted result in conjunction with T Paus and R Zatorre
- Wrote paper for journal submission

Tomas Paus:
- Participated in outlining rationale / procedure of study
- Provided guidance in analyses of results
- Participated in interpretation of results
- Proofed drafts of the article

Robert J Zatorre
- Participated in outlining rationale / procedure of study
- Participated in interpretation of results
- Proofed drafts of the article

Acknowledgements:

I would like to thank Robert Zatorre and Tomas Paus for providing me with invaluable guidance and inspiration for conducting these studies. Pierre Ahad has played an important in creating the acoustically complex stimuli used in all three of these studies, and in providing technical assistance in creating stimulus presentation programs; thank you Pierre. Thank you to Pascal Belin, who has also inspired me a great deal both with his enthusiasm and passion for science, and with his rigorous and lucid approach to research. Michael Petrides offered very helpful comments regarding morphological aspects of the results. Vali Petre, Keith Worsley, Alex Zijdenbos, Noor Kabani, JF Malouin, and Peter Neelin were very helpful regarding aspects of the brain imaging analyses. Thank you to Giovanna Locascio for handling bureaucratic aspects of my experience as a student at McGill. Thank you to my friends Martine, Abigail, Dana, Aghiles, Isaac, and Eduardo for offering me, through their presence in my life during the last few years, energy and inspiration for my work. Last but not least, a special thank you to my parents and to my brother Artia for always being there for me when I needed guidance and support.

Chapter 1

General Introduction

General Overview

There are large individual differences in how easily adults are able to learn speech sounds coming from foreign languages. The purpose of the three studies included in this thesis was to elucidate how these individual differences might be related to differences in more general, psychoacoustic learning abilities (Study 1), and how they might be predicted by differences in brain function (Study 2) and anatomy (Study 3). Previous work has already addressed aspects of these questions, such as for example how phonetic perception might be related to the perception of non-linguistic sounds, or such as what neural functional substrates underlie phonetic perception. To my knowledge, no one has to date addressed the question of how phonetic *learning* abilities may be related to more general auditory factors, or of how brain function, as measured by imaging techniques, may change as a result of phonetic training. In addition, no work has previously been done aimed at elucidating macroscopic morphological correlates of a continuous, speech-related ability in a population of normal, healthy adults. Study 3, which addresses this question, was exploratory relative to the first two since little previous work similar to this has been done.

In this general introduction, I will provide a brief overview of work on phonetic perception and learning in adults, on the categorical perception of speech and nonspeech sounds, and of factors influencing the ability to learn new speech sounds in adulthood. I will also briefly review neuropsychological literature on phonetic processing, and finally, will present results of previous work examining the relation between brain anatomy and behavior. I will finish with an overview of the aims and predictions of the three investigations included in this thesis.

I. PHONETIC LEARNING: BEHAVIOR

1. Phonetic perception and learning in adults

During development and starting as early as at six months of age, lack of experience with certain non-native speech sounds results in a developmental shift from a language-general to a language-specific pattern of phonetic perception (Kuhl et al, ’92, Best et al, ’88, Polka & Werker, ’94, Werker & Tees, ‘84a). Adults are more sensitive to acoustic differences among tokens of different native phonetic categories than to those of the same native category, even when the physical differences separating the stimuli have been equated. For example, most unilingual native Japanese speakers have considerable difficulty in hearing the difference between /r/ and /l/, since these sounds are not used to distinguish meaning in the Japanese language. The results of laboratory studies aimed at improving non-native phonetic perception in adults have shown that some contrasts are more easily learned than others. For example, it has been shown that training is effective in improving the ability to distinguish contrasts which differ along the voicing (VOT) dimension, suggesting that there is not a permanent sensory loss for certain speech sounds during development (Pisoni et al, ’82, McClaskey et al, ’83, Jamieson & Morosan, ’86, Carney et al, ’77). However, certain other contrasts which are distinguished in terms of place-of-articulation, such as the Hindi dental versus retroflex stop consonants, are much more difficult for adults to learn (Polka, ’92, Burnham, ’86, Tees & Werker, ’84).

2. Individual differences

There is a large range in performance across individuals in both the pre-training perception and in the amount of learning resulting from training, in particular for more “difficult” contrasts. For example, native Japanese speakers show a lot of variability in both pre-test and post-test identification and discrimination of the /r/ vs /l/ sounds (Strange and Dittman, '84, Jenkins et al, ’95). Large individual differences have also been found in the ability of native English speakers to distinguish the Hindi dental /da/ versus retroflex /da/ sounds (Polka, '91), and to improve in discriminating these phonemes after training (Pruitt et al, '90, Strange et al, '89). For example, Pruitt and colleagues (’90) found that a small number of subjects seem to benefit from training: their error rates dropped to less than 10% by day two of training. The majority of the subjects showed a gradual improvement over the course of the training. A third group showed no evidence of learning at all. Neither gender nor age seemed to be related to these differences.

Individual differences have also been shown in the ability to process complex, rapidly changing non-speech sounds. Näätänen and colleagues ('93), using an oddball paradigm, repeatedly presented subjects with a "standard" sound, composed of eight consecutive 50-ms segments of different frequencies. Occasionally, this sound was replaced with a "deviant" sound, which was similar to the standard one except that the frequency of the sixth segment was different. During the passive listening phase, subjects were reading a book and not attending to the stimuli. Two discrimination phases were interspersed between the oddball ones. Results showed that before any training, some individuals could discriminate the two complex sounds, that others developed this discrimination ability after several sessions of passive exposure, and that two subjects failed to develop this discrimination. It was also shown that the mismatch negativity (MMN) response, a physiological measure of discrimination which is independent of attention and voluntary response, emerged to variable degrees across adults. This increase in the MMN was paralleled by behavioral improvements in discrimination performance.

3. Psychophysics of categorical perception (CP) of speech sounds

We sort our sensory and cognitive experiences by categorizing them. Interest in the phenomenon of "categorical perception" (CP) was originally stimulated by work on the CP of colors and of phonemes. Speech sounds vary along a physical continuum; for example, the slope of the second formant transition (FT) varies when one compares the sounds /ba/, /da/, and /ga/. CP for speech sounds is defined as a quantitative discontinuity in discrimination at the boundaries of a physical continuum, as measured by a peak in discrimination acuity at the transition region for the identification of members of adjacent categories (Harnad, '87). Categorical perception helps to establish perceptual constancy among the acoustic variations of speech sounds and of words pronounced in different contexts by different speakers. For example, individuals can easily hear the difference between /ba/ and /da/, but perceptually assimilate two different instances of /ba/, even when the physical acoustic difference of the contrast is the same for both pairs. The motor theory of speech perception (Liberman et al, '63, Liberman & Mattingly, '85, '89) has been used to explain phonetic boundaries. Phonetic discontinuities are thought to arise from discontinuities required to pronounce sounds. A second explanation for CP is that of innate sensitivity, which attributes the discontinuities in perception to inborn enhancement and reduction of the sensory systems sensitivity in selected portions of certain psychophysical continua (Pastore, '76, Stevens, '81). Lastly, the label learning hypothesis, in its weaker form, suggests that labels are learned by exposure and association, and that these later come to influence identification and discrimination performance (Lane, '65). The stronger version of this theory claims that selective attention and learned expectations actually alter the perceptual similarity of stimuli (Fujisaki & Kawashima, '69, 71).

Some phonetic categories are innate. Research has demonstrated that very young, prelinguistic infants possess highly developed perceptual mechanisms for the perception of speech. Eimas and his colleagues ('71) showed that four month old infants discriminate speech sounds from different voicing categories that correspond to adult categories (ie; they demonstrated "categorical discrimination"). In contrast to innate CP, research on top-down processing and on human performance focuses on *learned* categories. It addresses the question of how learning based on a limited number of cases generalize to future cases.

Task, stimulus and subject parameters also influence CP (Repp, '84). For example, tasks or conditions that place fewer demands on working memory, such as smaller inter-stimulus intervals (ISI), or such as the use of truncated stimuli, lead to better within category discrimination (ie; to more acoustic rather than auditory processing). There are varying degrees of CP for different types of phonemes. Typically, individuals display the strongest degree of CP for stop consonants, intermediate levels for fricatives, affricatives, and liquids, and finally, no or very little CP for vowels (unless these latter are truncated) (Pisoni, '75, Fry et al, '62, Shankweiler & Studdert-Kennedy, '67). Syllabic position also influences CP: CP is more evident at syllable-initial than at the final position. This finding is consistent with the view that categorical perception is due in part to psychoacoustic masking effects of the vowel on the consonant (Tartter, '81).

4. Factors influencing the ability to learn non-native speech sounds

The tetrahedral model (Jenkins, '79) provides a framework for characterizing interacting factors that influence the perception of nonnative speech sounds, including stimulus materials, criterial variables, orienting variables, and differential subject characteristics. In the following section, I will focus on the first of these, and will address our motivations for selecting the other three factors in later, relevant sections of this thesis. Specifically, in this section, I will review evidence that both speech-specific as well as more general psychoacoustic factors influence the ability to hear and to learn non-native speech sounds. I will also describe two specific theoretical frameworks which have been developed to explain why some contrasts are more easily learned than others. Lastly, I will address the idea that *individual differences* in the ability to accurately perceive and learn certain new, non-native sounds may in part be related to differences in more general, rapid temporal processing abilities.

i. ***Relative influence of phonetic, phonemic, and acoustic processing.***

There is a body of work that has addressed the question of whether speech processing involves a special "speech mode" (ie; phonetic processing), whether it involves a more general "psychoacoustic processor" (ie; auditory processing), or whether it involves both. There are different lines of evidence for both of these single-factor models (Liberman, '82, Liberman et al, '67, Pastore, '77, Pisoni et al, '82, Jusczyk et al, '83). There is also evidence for the latter, dual-factor model, which suggests that both acoustic and phonetic factors are involved in speech perception, and that the relative influence of auditory versus phonetic factors during phonetic perception influence the ability to accurately perceive non-native speech sounds (Fujisaki & Kawashima, '70, Pisoni, '73). Further, it is thought that the degree to which auditory versus categorical information is used, or the ability to discriminate within versus between categories, respectively, is a complex function of stimulus characteristics, task demands, and subject factors (Zatorre, '83, Samuel & Tartter, '86, Pisoni, '73). For example, auditory perceptual traces are thought to decay more rapidly than phonetic ones. For this reason, under certain testing conditions such as when the ISI is long, the auditory information related to the first stimulus decays before the second stimulus is heard, and therefore subjects are forced to use more robust language-specific phonemic codes. In contrast, under other testing conditions, such as when the ISI is short or when truncated stimuli are used, finer discrimination abilities arising from acoustic processing can be shown. Note that the perception of natural speech involves relatively long intervals between verbal stimuli, imposing higher demands on working memory and making phonemic processing more likely.

Werker and colleagues (Werker and Tees, '84b, and Werker and Logan, '85) have extended the more traditional dual-factor framework to allow the distinction between three levels of speech processing. Within this model, the phonemic level refers to speech perception in terms of the listener's native phonology, the phonetic level refers to the perception of differences which are phonologically relevant in languages other than the listener's own language, and lastly, the auditory (psychoacoustic) level refers to perceived differences which are not phonologically relevant in any of the world's languages. This framework distinguishes between innate universal "phonetic" sensitivities versus learned linguistically relevant "phonemic" categories. Research on the development of CP has supported this three factor framework. Infants generally fail to differentiate sounds that do not exist in any of the worlds languages, yet can distinguish speech sounds that define phonetic classes across languages (Eimas et al, '71, Trehub, '76), even when variables such as speaker, speaking rate, or position of the phoneme in the word are manipulated to alter acoustic cues (Kuhl, '79, '80, '83, '85). In other words, infant speech perception is phonetically relevant, but is also language universal.

ii. *Best and colleagues' assimilation model.*

Best and colleagues ('88) have proposed a model which describes how phonemic, phonetic, and acoustic factors may underlie variation in the perception of non-native speech contrasts. They suggest that during speech perception, attention is normally focused at the phonemic level, such that listeners assimilate non-native phones to their native phonemic categories whenever possible. The pattern of assimilation is suggested to be a function of the phonetic similarity between native and non-native phones. Four types of assimilation patterns are suggested, and predictions are made about the perceptual difficulty of each. 1) The most difficult distinction is found in the "single category assimilation", in which both non-native phones are perceived as instances of the same native phoneme category. Performance on this distinction decreases at about 10-12 months of age. 2) The easiest distinction is for "two-category assimilation", in which the contrasting non-native phones are assimilated to two different native categories and the distinction is therefore perceived as a phonemic contrast. The 3rd and 4th types are of intermediate difficulty: 3) "Category-goodness assimilation", in which both phones are assimilated to the single native phonemic category, but the phones are differentiable because one phone is a better perceptual fit to the category than is the other phone (eg; perception of the g/G Farsi velar/uvular stop place contrast by English listeners). 4) "Non-assimilation", in which both phones are sufficiently dissimilar from any native category so as to be perceived as non-speech sounds, and the listener therefore differentiates them by attending to psychoacoustic differences between them (eg; English speaking adults are good at discriminating Zulu click contrasts). This type of assimilation is less common than the other three.

Best and colleagues suggest that the likelihood and direction of assimilation of non-native sounds can be predicted based on the degree of similarity in phonetic-articulatory features between the non-native item and native categories, and that phonetic similarity criteria should derive from phonetic-articulatory features established by phoneticians. During development, assimilation serves the purpose of structuring the phonological perceptual system, and of helping to establish perceptual constancy among different physical instances of the same phoneme.

iii. *Burnham's model: Robust versus fragile contrasts*

Burnham and colleagues (Burnham, '86, Burnham et al, '87) have developed a psychoacoustic hypothesis of speech perception, involving the distinction between "robust" and "fragile" phonetic contrasts. The ability to distinguish the sounds of a robust contrast is lost relatively late in development, between the ages of 4 and 8, at around the time of formal language training. The perception of these is relatively easy to train in adults. They tend to be easier to articulate than sounds of fragile contrasts, and have a clearer acoustic basis. They are more commonly used phonemically, and are also more likely to be allophonically represented across the languages of the world (ie; to be pronounced in languages in which the contrast is phonologically irrelevant). An example of a robust contrast is the prevoiced versus voiced bilabial stop, differing in VOT, which is phonemic in Spanish but not in English.

The perceptual sensitivity to sounds of "fragile" contrasts is lost earlier in life, at around the age of 6-12 months (Werker and Tees, '84a, Werker & Tees, '83, Werker et al, '81, Werker and Tees, '84b). These contrasts are harder to train in adults (Werker et al, '81, Werker & Tees, '83, Werker and Tees, '84a, Tees & Werker, '84), are more difficult to articulate, and have a weaker psychoacoustic basis. They are less likely to be part of the phonetic inventory of the world's languages, and are also less likely to be allophonically represented across languages. A good example of this type of contrast is the dental-retroflex place-of-articulation contrast which is used in languages of India such as Hindi or Urdu. Retroflex consonants require a relatively complex articulation, and their acoustic parameters are not well understood. They are rare across languages; only 11% of the world's languages include a retroflex consonant. Perceptually, English listeners perceptually assimilate the dental-retroflex sounds such that they perceive both sounds as instances of the dental consonant (Polka, '91, Werker & Lalonde, '88).

Burnham and colleagues (Burnham, '86, Burnham et al, '87) suggest that the perception of robust contrasts is thought to be lost due to a lack of phonetic *experience* with the sounds, whereas that of fragile contrasts is thought to be due to a lack of *exposure* to the sounds. As noted earlier, robust contrasts are more likely to be produced allophonically across languages. They suggest that simply being exposed to a phoneme, regardless of whether or not it serves to distinguish meaning from other sounds in a particular language, is sufficient to a) maintain perceptual sensitivity to that sound until a later age, and to b) facilitate the perceptual training of that sound during adulthood. In contrast, a lack of allophonic *exposure* to non-native "fragile" phonemes results in early perceptual loss for these sounds, as well as in difficulty in learning these sounds in adulthood. This hypothesis is consistent with the finding by Tees and Werker ('84) that adults having only been exposed to the Hindi dental-retroflex contrast at a very early age and not later were able to discriminate these sounds prior to studying the language, but that adults without early exposure to the retroflex sound did not improve in their ability to distinguish this sound from the dental one, even after one year of Hindi language courses.

iv. *Temporal versus spectral processing*

Certain speech sounds such as stop consonants contain frequency glides (formant transitions, FTs) or voice-onset-times which last as little as 40ms. The ability to distinguish these phonemes therefore depends critically on the ability to process very rapidly changing sounds. Vowels, on the other hand, are steady-state, and are distinguished based on differences in the frequency of the formants. The processing of such phonemes places relatively greater demands on spectral processing abilities. There is a large body of work which supports the hypothesis that what has traditionally been thought of as dominant hemispheric specialization for "speech" may in effect be a more general and multi-modal underlying specialization for the processing of rapidly changing information over time, which is critically necessary for the perception and production of speech (Schwartz & Tallal, '80, Tallal et al, '93, Belin et al, '98, Johnsrude et al, '97). It is possible that individual differences in the ability to *learn* certain difficult non-native speech sounds may be in part related to differences in the ability to parse rapidly changing temporal information, regardless of whether or not it is linguistic. Study 1 addresses this hypothesis.

Support for the hypothesis of a dissociation between temporal versus spectral processing abilities comes from research with clinical groups, as well as from functional imaging, electrophysiological, and dichotic listening studies with healthy individuals. Tallal and colleagues ('93) have reviewed evidence supporting the hypothesis that the speech processing impairments of patients with language disorders may result from problems with the processing of basic sensory information entering the nervous system in rapid succession, within the tens of milliseconds range. They have shown that language impaired children have deficits in processing rapidly presented acoustic information but not with the processing of steady-state stimuli. These children were shown to also have difficulty with the perception of vowel-vowel (VV) stimuli, where the duration of the two vowels was 40ms and 210ms, but not with VV stimuli with durations of 80ms and 170ms. This demonstrates that the temporal processing deficit is independent of phonetic classification. This deficit is pansensory, affecting several modalities, including the visual, tactile and cross-modal sensory integration of verbal and nonverbal stimuli. For example, the children have deficits in performing rapid sequential fine-grained manual or oral movements, within the 10s of ms.s time frame, and have impairments in their ability to control production of brief verbalizations. Tallal and colleagues hypothesize that the basic temporal deficits initially disrupt the normal development of the phonological system, and that this in turn affects the ability to read and speak. This is consistent with Merzenich and colleagues' ('93) hypothesis that a primary temporal processing deficit may result in a form of auditory deprivation that in turn alters neuronal mapping and connections across the auditory system with cascading effects on higher level auditory processes such as phonological perception and production abilities.

Results of electrophysiological (Liégeois-Chauvel et al, '99, Nicholls et al, '99) and of functional brain imaging (Belin et al, '98, Zatorre & Belin, '01, Johnsrude et al, '97) investigations have shown better temporal resolution for the left auditory cortex. In contrast, findings from electrophysiological (Liégeois-Chauvel et al, '01) and from lesion (Johnsrude et al, '00, Robin et al, '90) studies suggest that right auditory regions subserve aspects of pitch and spectral processing. There is also evidence from a dichotic listening investigation that longer inter-hemispheric transmission times (IHTT) from the right to the left hemisphere in the auditory modality are associated with greater left hemispheric specialization for linguistic perception, demonstrating the importance of rapid temporal processing in speech processing (Elias et al, 2000).

Dichotic listening involves the simultaneous presentation of different sounds to the two ears. The majority of right handed people more accurately perceive words (Kimura, '61, Kimura, '67) or nonsense syllables (Studdert-Kennedy & Shankweiler, '70, Shankweiler & Studdert-Kennedy, '67) presented to the right ear. This right ear advantage (REA) is thought to reflect a left-hemispheric specialization for speech processing, since crossed pathways from ear-to-cortex are more prominent during transmission than the uncrossed pathways. Conversely, findings in most right-handed people of a left-ear advantage (LEA) for familiar and unfamiliar melodic patterns and for environmental nonspeech sounds is thought to reflect a right hemispheric specialization for the processing of certain more slowly changing stimuli (Kimura, '64). Consistent with this is evidence that the perception of prosody in speech, in particular of affective prosody, is linked to right-hemispheric functioning (Blumstein & Cooper, '74, George et al, '96, Ross & Mesulam, '79, Ross et al, '97).

Dichotic listening studies of speech sounds have shown larger REAs for stop consonants than for steady-state vowels (eg; /ae/, /E/). Liquids (eg; /r/, /l/), semi-vowels, and fricatives produce REAs of intermediate strength compared to those of stop-consonants and of vowels (Shankweiler and Studdert-Kennedy, '67). The REA is significantly reduced for synthetic stimuli for which the rate of change of the formant transitions is extended (80 ms instead of 40 ms) while preserving the phonetic classification of the stimuli (Schwartz & Tallal, '80). Lastly, REAs have been shown in right-handed people for digits (Kimura, '61) and for trisyllabic nonsense words recorded and played backwards (Kimura & Folb, '68). Taken together, these results suggest that the magnitude of the REA is positively related to the rate of temporal change of stimuli, whether or not these are linguistic, suggesting that there is greater left hemisphere involvement during more general, psychoacoustic temporal relative to spectral processing.

II. PHONETIC LEARNING: BRAIN FUNCTION

Review of neuropsychological studies on phonetic processing

Linguists have proposed that language is decomposable into separate subsystems including the semantic, phonetic, and grammatical levels of processing (Chomsky, '65, Garrett, '80). Dissociations between these different aspects of language have been shown by linking variability in cognitive functions to concomitant and specific developmental alterations in cerebral function. For example, there is electrophysiological evidence that different subsystems within vision and within language display different degrees of experience-dependant modification, as well as different maturational time courses (Neville, '95, Weber-Fox & Neville, '92). Behavioral studies have also shown such differential effects of experience; acquisition of vocabulary appears to be least vulnerable to delays in language exposure, while other linguistic structures related to phonological and syntactical processing appear to be most affected (Johnson & Newport, '89). The three studies to be presented in this thesis address the lower level phonetic and psychoacoustic levels of processing. I will therefore, in the following section, review clinical, electrophysiological, and functional imaging work aimed at understanding the neural bases underlying the processing and learning of speech and nonspeech sounds.

1. *Functional imaging studies*

The neural correlates of phonetic perception have been studied using functional brain imaging techniques such as PET and fMRI. Generally, the results have shown the involvement of regions in and around what is classically known as "Wernicke's area", including left-sided activations in the superior temporal gyrus (STG), perisylvian area, and temporoparietal areas, the latter including the supramarginal and angular gyri. Activation of regions in and around the frontal speech area classically known as Broca's area has also been found during the performance of certain purely receptive language tasks (Démonet et al, '92 & '94, Zatorre et al, '92 & '96, Paulesu et al, '93, Fiez et al, '95, Burton et al, '00). The introduction to Study 2 provides a more detailed review of other previous functional studies on phonological perception.

A few imaging studies have examined the functional substrates underlying the perception of non-linguistic and linguistic stimuli with or without rapidly changing temporal changes. Belin and colleagues ('98) used PET functional brain imaging during passive listening to slow (200ms) and rapid (40ms) frequency changes in nonsense speech-like syllables. Their stimuli consisted of a central steady-state period, surrounded by initial and final formant transitions with randomized directions of frequency changes. These resulted in unpronounceable, speech-like syllables devoid of verbal content. They found that the consonant-like rapid frequency changes are better processed by the left than the right auditory cortex. They interpreted the results as suggesting that this low level, auditory processing asymmetry could explain high level language lateralization, and that the speech perception advantage of the left hemisphere may be related to superior temporal processing resulting in efficient coding of rapid acoustic changes such as the rapid FTs in consonants. Zatorre and Belin ('01) performed a functional imaging study aimed at examining the response of the human auditory cortex to spectral and temporal variation in pure tone sequences. They found evidence that responses in the core auditory cortex to the temporal features were weighted towards the left, while responses to the spectral features were weighted towards the right. These findings support the idea that the left hemisphere is specialized for rapid temporal processing, and that there is a complemetary hemispheric specialization of the right-hemisphere cortical areas for spectral processing.

Fiez and colleagues ('95) also found an association between left-lateralized activation and the processing of rapidly changing sounds. They manipulated two variables: 1) speech versus nonspeech, and 2) stimuli with versus without rapidly changing acoustic cues. They found that in the left hemisphere, the frontal opercular increases were larger when subjects performed an auditory detection task upon verbal and nonverbal stimuli that incorporated rapid temporal changes (words, syllables, and tone sequences) than steady state vowels.

2. *Electrophysiological studies*

The neurophysiological correlates of phonetic learning have been studied using ERP methodology. The mismatch negativity (MMN) evoked potential is an auditory cortical response to acoustic change that is introduced in a repetitive stimulus sequence (Näätänen et al, '78 & '93). It has been shown that behavioral training of two slightly different speech stimuli in adults results in a significant change in the duration and magnitude of this cortical potential (Kraus et al, '95), and that this physiological change precedes behavioral discrimination improvements (Tremblay et al, '98), suggesting that MMN is a measure of pre-attentive learning (see Kraus & Cheour, 2000). More generally, other techniques such as single cell recordings in animals (Recanzone et al, '93, Kraus & Disterhoft, '82) and magnetoencephalography (Pantev et al, '99) have shown plasticity of auditory cortex function resulting from training and experience. The mismatch generators are thought to involve the thalamo-cortical association areas of the auditory cortex (Giard et al, '90, Kraus et al, '94). In response to speech stimuli, some studies have shown the MMN response to be symmetrical (Aaltonen et al, '94, Tremblay et al, '97). Other studies, however, have shown slightly larger responses to vowels in the right hemisphere (Csépe, '95), and larger ones to consonants over the left hemisphere (Csépe, 95, Alho et al, '98). Interestingly, Tremblay and colleagues ('97) showed that MMNs elicited by nonnative speech syllables were initially symmetrical, but that they became enhanced over the left hemisphere following training. Taken together, these findings suggest that left-sided specialization for speech may be evident even in the representation of phonetic information.

To date, no research has been done using functional brain imaging techniques to investigate neural plasticity related to phonetic training. This will be addressed in Study 2. Based on the above described findings that phonetic training results in detectable neurophysiological changes, we predicted that we will find a change in the BOLD signal accompanying behavioral improvements in the perception of a non-native phonetic contrast.

3. *Dyslexia*

There is an abundance of evidence that shows that at least a subset of dyslexic individuals have problems performing tasks that require the processing of phonological information. Rack and colleagues ('92) reviewed a large number of studies examining phonological reading skills in dyslexic individuals. Based on their review, they have argued that *most* dyslexics have a specific deficit in phonological reading. I will review evidence for abnormal brain function and structure in language-related regions in this clinical group. These findings are relevant to some of the results of my studies, in particular to those of Study 3.

There is evidence that at least a subset of dyslexic individuals, in addition to more general language-based deficits, demonstrate specific *phonological* deficits (ie; problems in processing speech sounds). For example, dyslexic adults show deficits in the phonological aspects of reading, as detected by poor non-word reading and poor spelling (Pennington et al, '87). The dual route theory proposes that there are two possible neural pathways involved in reading. In the "phonological route", words are translated into their speech sound equivalents, and in the "orthographic route", it is thought that there are direct links between visual word forms and word meanings, without the involvement of a phonological recoding step. The former, "phonological route" is the only one available for reading nonwords, and this is the route that is uniquely deficient in many dyslexics (Høien & Lundberg, '89, Lundberg, '89, Calfee et al, '72). Results of functional imaging studies suggest that phonological coding depends on adequate functioning of inferior parietal regions including the left angular and supramarginal gyri (Paulesu et al, '93, Petersen et al, '88). There is evidence for functional connectivity between the angular gyrus and posterior language areas during the normal performance of language-related tasks (Démonet et al, '92, Rumsey et al, '99).

Abnormal functional activation of left temporoparietal regions has been shown in dyslexic relative to controls subjects during the performance of a phonological task (rhyme detection) (Rumsey et al, '92). In another study, reading skills were correlated with regional cerebral blood flow (rCBF) in temporal and inferior parietal regions of the brain. It was found that higher rCBF in the left angular region is associated with better reading skills in control subjects, but with worse reading skills in dyslexic subjects, suggesting functional abnormality of this region in dyslexia (Rumsey et al, '99). In a third study, while functional connectivity was shown between the left angular and temporal regions in normal individuals during word reading, such connectivity was not shown in a group of dyslexic adults (Horwitz et al, '98). The results of these three studies suggest abnormal left inferior parietal function, as well as abnormal functional connectivity between this area and temporal regions in dyslexia.

A number of studies have also demonstrated abnormal morphology in the left angular, parieto-occipital, and temporal regions in dyslexia. Duara and colleagues ('91) found a right-greater-than-left asymmetry in the cross-sectional area of a region corresponding to the angular gyrus in dyslexic individuals, whereas the area was symmetrical in normal control subjects. Other studies have shown that there is a lack of or a reversal of the normal asymmetry in the parieto-occipital region in dyslexic individuals (Hier et al, '78) and in children with language disorders (Jernigan et al, '91, Plante et al, '90), relative to normal control subjects

An unusually high incidence of temporal lobe volume symmetry (Rumsey et al, '86, Kushch et al, '93) has been reported in dyslexia. Finally, several groups have shown either a lack of or a reversed morphological asymmetry of the planum temporale in dyslexic relative to in normal individuals (Larsen et al, '90, Hynd et al, '90). Larsen and colleagues ('90) showed that 70% of dyslexic individuals showed morphological symmetry in the planum temporale, when only 30% of normal individuals showed symmetry in this region. They also showed that within the dyslexic group, all individuals who had pure phonological deficits in reading had symmetrical plana temporale, and that this symmetry appeared to result from a larger than "normal" right PT. Post-mortem findings have also shown an unexpected symmetry of the PT (Galaburda et al, '85, Galaburda, '88).

III. BRAIN ANATOMY AND BEHAVIOR

Review of literature on relation between anatomy and behavior

One of the earliest normative observations of brain morphology was the finding of a structural asymmetry between the left and right plana temporalia (PT) (Geschwind and Levitsky, '68, Wada et al, '75), auditory association areas which on the left side corresponds with the center of Wernicke's speech area (Geschwind, '70). This PT asymmetry is still conventionally thought to reflect a size asymmetry favoring the left over the right hemisphere. The majority of researchers define the posterior end of the PT as the point of upward deflection of the sylvian fissure, behind which lies an area defined as belonging to the planum parietale (PP). Several studies have shown that the sum of the PT and PP areas does not differ between the right and left hemispheres (Binder et al, '96, Loftus et al, '93, Steinmetz et al, '90). Moreover, Westbury and colleagues ('99) showed that when the definition of the posterior border of the PT is changed, there are no longer hemispheric differences in the area nor the volume of the PT, suggesting that the apparent size asymmetry may actually reflect an asymmetry in the shape of the PT and/or of adjacent structures. The shape of brain structures and of fiber tracts connecting various brain subregions may be as important as the size or length of those structures and tracts with respect to the pattern and speed of neural processing and transmission in and between functionally connected regions, and with respect to the integration of cerebral activity across these different regions.

It has been suggested that morphological asymmetry of the PT may play a role in language lateralization (Geschwind & Levitsky, '68, Galaburda et al, '78, Wada et al, '75, Steinmetz et al, '91), providing a structural basis for the functional specialization of auditory processing (Zatorre & Binder, 2000). In support of this, it has been shown that morphological asymmetries of the PT (Foundas et al, '94) and of blood vessels in the region of the sylvian fissure (Ratcliff et al, '80) are related to language lateralization for speech. As well, it has been shown that planum temporale morphological asymmetry is related to handedness such that left handers have a lesser degree of asymmetry than right handers (Steinmetz et al, '91, '95, Foundas et al, '95, Habib et al, '95). This latter finding was one of the first demonstrations of a behavioral correlate of brain structure. Planum temporale asymmetry has also been shown to be related to music-related abilities. Specifically, exaggerated leftward asymmetry has been associated with the capacity for perfect pitch (Schlaug et al, '95, Zatorre et al, '98).

The above examples of anatomy-behavior relationships, with the exception of the work by Zatorre and colleagues ('98), involved correlating *binary* behavioral measures with brain morphology. In Study 3 (below), we correlated a continuous measure of the ability to learn new speech sounds with brain morphology across a relatively large (N=59) sample of individuals.

IV. METHODOLOGICAL JUSTIFICATION

1. Advantages and disadvantages of using synthetic stimuli

The use of synthetic stimuli allows the experimenter to create a continuum of syllables that differ from each other in a known and controlled way (eg; by interpolating the physical parameter in equal steps between the endpoint stimuli). In other words, it allows manipulation of only the acoustic parameters that are relevant in distinguishing contrasting sounds, while keeping constant ones which are not necessary for differentiating them. The lesser degree of acoustic variability in synthetic relative to naturalistic stimuli may facilitate the training of difficult contrasts, since the unequal physical variability within and between nonnative phonetic categories present in naturalistic stimuli is eliminated in synthetic stimuli. For the above reasons, we used synthetic stimuli in the three studies to be presented.

A potential difficulty with the utilization of a synthetic dental-retroflex stimulus continuum is that the acoustic characteristics of retroflexion are not that well understood (Stevens & Blumstein, '75, Tees & Werker, '84). For example, it has been shown that there are difference degrees of retroflexion between the retroflex stops in Hindi, Telugu, and Tamil (Ladefoged and Bhaskararao, '83). This suggests that at least for this place-of-articulation contrast, there are not discrete places of articulation as might be suggested by theories supporting quantal perception of speech sounds. Stevens and Blumstein ('75) state that in constructing their synthetic stimuli, certain acoustic dimensions which may have served as essential cues for the identification of retroflex consonants may have been eliminated. We addressed this issue by validating our synthetic stimulus series, and showing that native speakers of Indian languages did successfully identify our dental and retroflex sounds. The use of naturalistic stimuli is preferred in studies which aim to study phenomena that will generalize to real speech. This was not a central concern in our studies.

2. Selection of the retroflex non-native sound

As described above, the retroflex consonant is unlikely to occur both phonetically and allophonically across languages. This allowed us to more easily screen subjects for both experience and exposure, respectively, to this phoneme. We also selected this sound, which has been shown to be difficult to perceive and to train in English-speaking adults, because we wanted to avoid both pre- and post-training ceiling effects. Instead, we wanted to ensure that we would obtain a large range performance across subjects. This was desired in order to allow us to correlate individual differences in performance with a) the ability to learn to perceive non-linguistic, rapidly changing, as well as steady-state, tonal stimuli (Study 1), b) functional brain activation (Study 2), and c) brain morphology (Study 3).

3. Selection of subjects

There is evidence that in bilinguals, the perception of phonetic categories reflects the influence of the phonetic categories existing in both languages (Caramazza et al, '73, Williams, '74, '77). In Study 2, we included only monolingual English speakers. Participants in Studies 1 and 3, however, were fifty-nine individuals, of whom approximately four-fifths were multilingual and came from a variety of language backgrounds. In selecting the non-native retroflex sounds, we were aware that, as a result of differences in our subjects' phonemic backgrounds, this phoneme may be assimilated to different degrees by some individuals relative to others. We did not expect, however, that such possible differences in assimilation would bias subjects in any systematic manner. Additionally, as described in the previous section, the retroflex sound is very unlikely to be pronounced allophonically in languages in which it is non-phonemic. We excluded subjects who spoke or who had been exposed to any languages that utilize any type of retroflex sound.

V. OVERVIEW OF THESIS PROJECTS

Study 1: *Learning the Hindi dental-retroflex contrast: Phonetic or auditory factors?*

The first study included in this thesis is a behavioral study with fifty-nine healthy volunteers. We first conducted a validation study with two synthetic versions of the dental-retroflex stimulus continuum in order to select the better one of these for the main, training study. We tested individuals who speak languages of India natively, and established that they could reliably identify the exemplar (continuum endpoint) sounds as belonging to the appropriate dental and retroflex phoneme categories. In the main part of Study 1, we addressed two principal questions; both have implications for theories of plasticity of phonetic perception in adults. Firstly, we asked whether or not the ability to accurately perceive "difficult" (ie; fragile) non-native speech sounds to which individuals have had no early or allophonic exposure is permanently lost in adulthood. Secondly, in order to elucidate the mechanism of phonetic learning in adulthood, we tested the hypothesis that individual differences in phonetic learning are a function of a more general ability to process rapidly changing sounds, regardless of whether or not they are linguistic in nature.

We trained individuals to learn to distinguish the following three stimulus types: the non-native dental-retroflex phonetic contrast, rapidly changing, non-linguistic sounds, and a tonal pitch contrast which involved steady-state frequency differences. We then examined associations and dissociations between two measures of performance for each stimulus type. These measures included pre- and post-training discrimination and identification performance. We predicted firstly that there would be a large range of individual differences in performance across subjects, and that given our large sample size, we might succeed in detecting a significant overall improvement in the ability to distinguish the dental-retroflex contrast as a result of training. Secondly, we predicted that we would find an association between performance using the Hindi speech sounds and that using the rapidly changing, non-linguistic sounds, and a dissociation between performance using these two stimulus types and performance using the steady-state, tonal, pitch contrast. In other words, we predicted that we would find a dissociation between the ability to process temporal versus spectral information, regardless of whether or not it is linguistic. This finding would provide evidence that phonetic perception depends, at least in part, on a generalized psychoacoustic processor, and not solely on a specialized linguistic processor.

Study 2: ***Learning new sounds of speech: reallocation of neural substrates.***

The aim of our second study was to determine how the pattern of brain activity may change as a result of training with speech sounds from a non-native language. Subjects were scanned using fMRI before and after a two-week period of phonetic training with a Hindi dental-retroflex contrast. During scanning, a native phonetic contrast was used as a control. A lower level noise control condition was also used in order to subtract out lower level acoustic processing of complex sounds and to make the results more comparable to those of previous studies on phonetic processing. We addressed the question of whether the identification of newly learned speech sounds recruits the same neural-substrates as does the identification of a known, native phonetic contrast, or whether new areas are recruited. We also wanted to see whether we could differentiate "learners" from "non-learners" on the basis of their pattern of activation while they classify the new speech sounds. We predicted firstly that the native identification task would reveal the bilateral involvement of superior temporal regions, stronger in the left than in the right hemisphere, of the left temporo-parietal region, and of the left frontal gyrus (IFG) in and adjacent to Broca's area. We predicted that after training, the pattern of activation for the non-native condition would be similar to that found in the native condition. This prediction was in part based on results of neuroimaging studies of language function in healthy bilinguals, showing that at the single word level, brain regions subserving the native language (L1) and the second language (L2) in fluent bilinguals appear to overlap (Klein et al, '94 & '95, Chee et al, '99, Illes et al, '99). Last, based on the assumption that more successful task performance recruits underlying neural substrates more actively, we predicted that correlations between a behavioral learning measure and the blood oxygenation level dependant (BOLD) signal during the post-training non-native task would reveal a positive relationship between learning and signal in left pre-frontal and left temporo-parietal speech areas.

Study 3: ***Anatomical Correlates of Phonetic Learning***

The aim of the third study was to determine whether individual differences in the ability to learn speech sounds coming from a non-native language may be in part accounted for by individual differences in brain morphology in language-related or other regions of the brain. We correlated three behavioral measures for the Hindi phonological contrast from Study 1, including pre-training discrimination performance, learning rate, and amount of improvement in identification performance after relative to before training, with brain morphology throughout the whole brain volume. In addition, we wanted to test the speech-specificity of any potential anatomical correlates of phonetic learning. Therefore, we only performed analyses that yielded significant results with the speech sounds with behavioral learning measures on rapidly changing sounds and steady-state tonal sounds from Study 1. We predicted that the ability to learn the phonological contrast would be correlated with differences in brain morphology in language-related cortical areas, including left parieto-temporal regions as well as left frontal opercular and inferior frontal regions.

Chapter 2

Study 1: *Learning the Hindi dental-retroflex contrast: Phonetic or auditory factors?*

During development and starting as early as at six months of age, lack of experience with certain non-native speech sounds results in a developmental shift from a language-general to a language-specific pattern of phonetic perception (Best et al, '88, Werker & Polka, '93, Polka & Werker, '94, Werker & Tees, '84a, Werker & Lalonde, '88, Werker et al, '81, Kuhl et al, '92). During adulthood, most individuals perceptually assimilate certain non-native speech sounds with similar ones from the native language (Fledge, '84, Pisoni et al, '82, Werker & Tees, '84b). In the present study, we were interested in determining firstly to what extent adults can learn to distinguish a "difficult" non-native phonetic contrast as a function of training, and secondly, whether individual differences in such an ability might be in part accounted for by more general auditory factors such as temporal processing abilities.

The results of laboratory studies aimed at improving non-native phonetic perception of contrasts have shown that some contrasts are more easily learned than others (Werker et al, '81, Best et al, '88, Burnham, '86). It has been argued that contrasts based on rapidly changing spectral cues, such as phonemes distinguished by the place of articulation, may be more difficult for listeners to acquire than ones based on rapidly changing temporal cues, such as phonemes distinguished by voice onset time (VOT) (Strange & Dittman, '84, Strange & Jenkins, '78). In support of this, it has been shown that training is effective in improving adults' ability to distinguish certain voicing contrasts, suggesting that there is not a permanent sensory loss for certain speech sounds during development (Pisoni et al, '82, McClaskey et al, '83, Jamieson & Morosan, '86, Carney et al, '77). Non-native voicing phonemes are likely to occur allophonically across languages (Lisker & Abramson, '67, Abramson & Lisker, '70, Repp & Liberman, '87). In other words, they are likely to be produced during speech, even though they do not carry phonemic information. It has been suggested that being exposed to certain non-native phonemes allophonically may be sufficient to maintain perceptual sensitivity to these sounds (Burnham, '86, Tees & Werker, '84).

In contrast to adults' relative facility in learning voicing distinctions, it has been shown that learning of certain contrasts which differ with respect to place of articulation is more slow and effortful. For example, it has been shown that even an extensive course of training (14 to 18 sessions) with the /r/ versus /l/ contrast with Japanese speakers results in only small improvements in the ability to distinguish these sounds (Logan et al, '91), and that the modification of the perception of this contrast in adulthood is slow and effortful (Strange & Dittman, '84). Similarly, it has been shown that the Hindi retroflex sound is difficult for English speaking adults to distinguish from the dental one, even after a considerable amount of training (Werker et al, '81, Werker and Tees, '84a, Tees & Werker, '84). Typically, there is a large range of individual differences across individuals in the amount of improvement as a function of training for both of these contrasts. Consistent with the allophonic exposure hypothesis, both of these non-native contrasts are not or are rarely pronounced allophonically in the native language. For example, the sounds /r/ and /l/ are phonetically distinctive in the English language, but are neither phonetically distinctive nor used allophonically in the Japanese language. Similarly, the retroflex stop consonant is phonetically relevant in languages of India such as Hindi or Urdu, but is not phonemic nor is it produced allophonically in English.

We address two principal questions in our main, training study; both have implications for theories of plasticity of phonetic perception in adults. First, we ask whether or not the ability to accurately perceive "difficult" non-native speech sounds to which individuals have had no early or allophonic exposure is permanently lost in adulthood. Secondly, in order to elucidate the mechanism of phonetic learning in adulthood, we will test the hypothesis that individual differences in phonetic learning are a function of a more general ability to process rapidly changing sounds, regardless of whether or not they are linguistic in nature.

It has been proposed that the ability to process certain consonant speech sounds depends on the ability to track rapidly changing acoustic information (Efron, '63, Tallal et al,'93, Benasich & Tallal, '96), and that what has traditionally been thought of as dominant hemispheric specialization for "speech" may in effect be a more general and multi-modal underlying specialization for the processing of rapidly changing information over time (Schwartz & Tallal, '80, Tallal et al, '93, Belin et al, '98, Johnsrude et al, '97, Zatorre & Belin, '01). Support for the hypothesis of a dissociation between temporal versus spectral processing abilities comes from research demonstrating that adults and children with language impairments have deficits in discriminating between speech sounds that incorporate rapidly changing acoustic information, but not between sounds that are composed of steady-state or slowly changing acoustic information such as is contained in vowels or glides. Interestingly, these individuals have deficits in perceiving certain types of non-linguistic information (Efron, '63). For example, they appear to be impaired at perceiving rapidly changing nonlinguistic, tonal stimuli, and the deficit in the ability to process rapidly changing information extends to production in other modalities (eg; performance of rapid, fine-grained manual or oral movements) (Tallal et al, '93). Results of electrophysiological (Liégeois-Chauvel et al, '99, Nicholls et al, '99) and of functional brain imaging (Belin et al, '98, Zatorre & Belin, '01) investigations have shown better temporal resolution for the left auditory cortex. In contrast, findings from electrophysiological (Liégeois-Chauvel et al, '01) and from lesion (Johnsrude et al, '00, Robin et al, '90) studies suggest that right auditory regions subserve aspects of pitch and spectral processing. There is also evidence from a dichotic listening investigation that longer inter-hemispheric transmission times (IHTT) from the right to the left hemisphere in the auditory modality are associated with greater left hemispheric specialization for linguistic perception, demonstrating the importance of rapid temporal processing in speech processing (Elias et al, 2000).

In the present study, we trained individuals to learn to distinguish three types of synthetic stimuli: a non-native dental-retroflex phonetic contrast, rapidly changing, non-linguistic sounds similar to those used by Belin and collegues ('98), and a tonal pitch contrast. We then examined associations and dissociations between measures of performance for each stimulus type. These measures included pre- and post-training discrimination and identification performance. The validity of the Hindi stimulus continuum was verified in Study 1 (see below). We predicted firstly that there would be a large range of individual differences in performance across subjects, and that given our large sample size, we might succeed in detecting a significant overall improvement in the ability to distinguish the dental-retroflex contrast as a result of training. Secondly, we predicted that we would find an association between performance using the Hindi speech sounds and that using the rapidly changing, non-linguistic sounds, and a dissociation between performance using these two stimulus types and performance using the steady-state, tonal, pitch contrast. In other words, we predicted that we would find a dissociation between the ability to process temporal versus spectral information, regardless of whether or not it is linguistic. This finding would provide evidence that phonetic perception depends, at least in part, on a generalized psychoacoustic processor, and not solely on a specialized linguistic processor.

Dental-retroflex phonetic contrast: Stimulus selection and previous work

The dental-retroflex place-of-articulation contrast is used in languages of India such as Hindi or Urdu. Retroflex consonants require a relatively complex articulation, and are rare across languages; only 11% of the world's languages include a retroflex consonant. They are also unlikely to be allophonically represented across languages (ie; to be pronounced in languages in which the contrast is phonetically irrelevant) (Burnham, '86). The non-native perception of the retroflex sound is lost early in development, below the age of one (Werker & Tees, '83, Werker & Tees, '84a, Tees & Werker, '84), and perceptually, English listeners assimilate the dental-retroflex sounds such that they perceive both sounds as instances of the dental consonant (Polka, '91, Rivera-Gaxiola et al, 2000a, Werker & Lalonde, '88). Research aimed at training the non-native perception of the dental and retroflex sounds in adults has shown that standard training approaches fail to ameliorate overall group performance (Werker et al, '81, Werker and Tees, '84a, Tees & Werker, '84), and that however, there are large individual differences in the ability to improve across subjects (Pruitt et al, '90, Strange et al, '84, Werker et al, '81). Overall group improvement has been reported under certain circumstances, including the use of shorter inter-stimulus intervals (ISI) (Werker and Tees, '84b, Werker and Logan, '85), truncated stimuli (Pruitt et al, '90), and extensive experience with the contrast (Tees and Werker, '84).

We selected the retroflex consonant because we wanted to ensure that none of our participants had had either allophonic or non-native exposure to this sound. Additionally, we wanted to ensure that no individual would be able to distinguish this sound from the dental consonant before training (ie; to avoid pre-training ceiling effects).

STUDY 1: VALIDATION STUDY

Few groups have synthesized the retroflex-dental contrast (Stevens & Blumstein, '75, Lisker, '85 (unpublished), Werker & Lalonde, '88). The purpose of Study 1 was to create two versions of the dental-retroflex continuum, and to determine which of the two series more accurately represents the speech sounds in question. We tested individuals who speak languages of India natively to establish whether or not, for each stimulus series, they could reliably identify the two exemplar (continuum endpoint) sounds as belonging to the appropriate phoneme categories. All of the above-mentioned training studies using the dental versus retroflex sounds have employed naturalistic stimuli. The use of synthetic stimuli will allow us to manipulate only the acoustic parameters that are relevant in distinguishing contrasting sounds, while keeping constant ones which are not necessary for differentiating them. This will allow us better experimental control over the stimuli than might the use of naturalistic stimuli. In addition, the lesser degree of acoustic variability in synthetic relative to naturalistic stimuli may facilitate the training of this contrast.

METHODS

Subjects

Subjects for the validation study included nine undergraduates (3 male and 6 female), ranging in age from 19 to 27 years. Participants spoke either Urdu (8 Ss), Kanuda (1 subject), or Gujarati (1 subject) natively, and some also spoke another language that has the dental-retroflex contrast (eg; Punjabi, Hindi). All subjects also spoke English fluently.

Stimulus Synthesis

The synthesis of Set 1 was based on parameters reported by Stevens and Blumstein ('75) for the voiceless, unaspirated dental and retroflex sounds. Parameters for stimulus Set 2 were based on ones reported by Werker and Lalonde ('88) for the voiced, unaspirated dental and retroflex sounds. In each of the two stimulus sets, there were seven stimuli varying in equal steps in terms of acoustic difference between adjacent items. Both sets of four-formant stimuli were constructed with the use of the Mitsyn (WLH) synthesizer. In both sets, stimulus 1 corresponded to the dental and stimulus 7 to the retroflex stop consonant prototypes, and all of the consonants were followed by the vowel /a/.

Both of our stimulus sets began with an initial noise burst. Previous studies have shown that the burst is necessary to create the retroflex-dental distinction. The noise burst had a bandwidth of 10 kHz and an exponential roll-off. In Set 1, the parameters that were manipulated to create the continuum are the frequency glides of the third formant (F3), as well as the center frequency of the burst. The initial noise burst lasted 5ms, and its center 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 40ms. The fundamental frequency (Fo) 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 Hz to 655 Hz during the transition, and then remained steady-state during the remainder of the sound. The second formant (F2) decreased from 1650 Hz 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 Figure 1 for schematized frequency-time representations of the dental and retroflex consonant-vowel (CV) syllables.

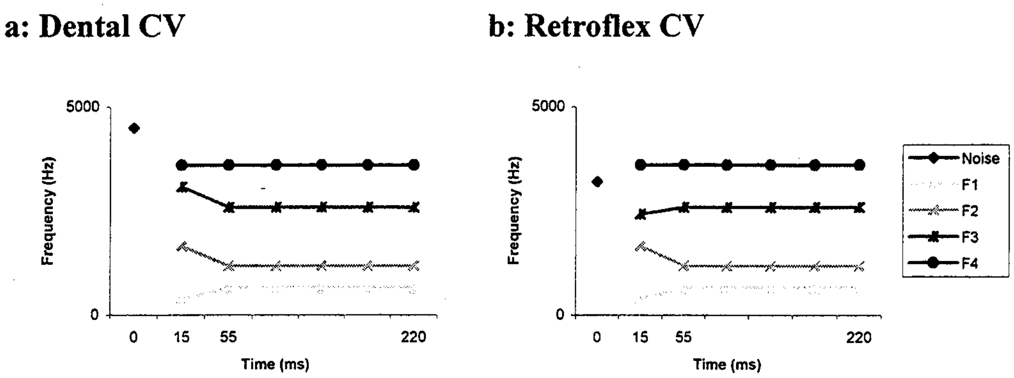


FIG. 1: Schematized frequency versus time diagrams of the (a) dental and (b) retroflex consonant-vowel (CV) syllables.

In stimulus set 2, the frequency glides of the 2nd and 3rd formants were manipulated. Each stimulus lasted 245 ms, and FTs lasted 40ms. The initial noise burst lasted 10ms, and its center frequency was 3150 Hz (BW=750Hz). Voicing began at the offset of the burst. The fundamental frequency (Fo) remained constant at 115 Hz throughout the syllables. F1 rose from 250 to 500 Hz during the first 40 ms, and then remained steady-state during the rest of the sound. F4 remained steady-state at 3500 Hz during the entire sound. The starting frequency of F2 varied in equal 50 Hz steps from 1500 Hz (stimulus 1) to 1800 Hz (stimulus 7), and its steady-state frequency was 1090 Hz. The starting frequency of F3 varied in equal 28 Hz steps from 2856 Hz (stimulus 1) to 3024 Hz (stimulus 7), and then remained steady-state at 2440 Hz.

Procedure

The following procedure was administered for each of the two stimulus continua. We began with familiarization, during which subjects heard 30 instances of each of the two endpoint stimuli, randomly presented. As each stimulus was presented, subjects saw its corresponding label on the computer screen, “A” for the dental sound, and “B” for the retroflex one. After familiarization, subjects were 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. They were asked to explain how they would articulate the A and B sounds, and if possible, to provide the written symbol in an Indian language for the sound that they heard.

In both the validation and training studies, sounds were presented binaurally through headphones at an intensity level of 70 dB SPL in a sound attenuated room.

RESULTS AND DISCUSSION

Results of the validation study revealed that for stimulus set 1, six out of the eight individuals from the Indian-language speaking group identified the endpoints as corresponding to the dental versus the retroflex voiced, unaspirated sounds (da/da). The results for stimulus set 2 were more variable. Only four out of nine subjects identified the endpoints as corresponding to the dental and retroflex sounds. The other subjects either perceived the vowel portions of the sounds to be different (two subjects), or perceived a difference in the consonant portion which they found difficult to explain (3 subjects). The results, taken together, supported the validity of stimulus Set 1 since native speakers reliably and accurately identified stimuli 1 and 7 as the retroflex and dental sounds, respectively. This stimulus continuum was therefore selected as the non-native contrast to be used for the training study.

STUDY 2: TRAINING STUDY

METHODS

Subjects

Subjects included 59 healthy right-handed adult volunteers (33 male and 26 female), screened for experience with any type of retroflex sound. All subjects spoke English, and a subset also spoke one or more additional languages. Linguistic and musical experience were evaluated with questionnaires. Hearing was tested and was found to be normal in all subjects prior to beginning the experiment.

Stimuli

We tested and trained subjects with three sets of stimuli: synthetic voiced, unaspirated dental-retroflex speech stimuli, nonlinguistic formant-based stimuli, and steady-state tonal stimuli. The non-linguistic stimuli were formant-based, and therefore had a voice-like quality, yet were non-linguistic in that they did not, to our knowledge, represent phonemes that exist in any existing language. When asked to qualitatively describe these sounds, subjects most often reported hearing either animal sounds, computer sounds, or human non-verbal sounds.

Stimulus Synthesis

Refer to Study 1 for acoustic parameters of the dental-retroflex stimulus continuum. The acoustics of the nonlinguistic stimuli were analogous to those of the speech syllables in that both of these 7-step continua, the frequency glides of the third frequency band was manipulated. They differed from the speech stimuli in that the frequencies of the "formants" were arbitrarily selected, and in that there was no noise burst preceding the sounds. These sounds were created in a manner similar to those used by Belin and collegues ('98). Our stimuli were different than theirs in two respects: first, the direction and slopes of our glides were different, and second, their stimuli had frequency glides both at the beginning and at the end of the sounds, whereas ours only had glides at the beginning.

The acoustics of the non-linguistic stimuli are the following. All of the stimuli lasted 220 ms. The frequency band analogous to Fo was steady-state at 210 Hz. The one analogous to F1 rose from 337.5 to 450 Hz during the first 40 ms, and then remained steady-state. The one analogous to F2 dropped from 2000 to 1500 Hz during the first 40 ms, and then remained steady-state. The one analogous to F4 remained steady-state at 3300 Hz throughout the entire sound. Finally, the starting frequency of the frequency band analogous to F3 decreased in equal 140 Hz steps from 3020 (stimulus 1) to 2180 (stimulus 7), and its steady-state frequency was 2400 Hz.

The steady-state pure tonal stimuli all lasted 220 ms, and had onset and offset times of five and twenty milliseconds, 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 2007 Hz (stimulus 1) to 2042 Hz (stimulus 7).

Procedure

Pilot testing served to equate the overall difficulty level of the three stimulus sets. Each subject was tested and trained with each of the three stimulus types, the order of which was counterbalanced across subjects. In some cases, all of the testing and training was done on a single day, and in others, it was done on two or on three separate days. For each stimulus type, subjects were first familiarized with the sounds by presenting subjects with 20 instances of each of the two endpoint stimuli, randomly presented. As each stimulus was presented, subjects saw its corresponding label on the computer screen, "A" for the dental sound, and "B" for the retroflex one. After familiarization, subjects were 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 discrimination and identification testing (described below) was then conducted. Following this, subjects were trained, using the adaptive training procedure described below. Discrimination and identification tests were re-administered after training.

Identification testing

Subjects 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 7 stimuli was presented 10 times in a random sequence, for a total of 70 trials. During pre-training testing, it was difficult for most subjects to distinguish even the endpoint sounds. They were therefore encouraged to think back to the descriptions of A and B that they had provided during familiarization.

Discrimination testing

Subjects were then 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 7 "same" pairs, five two-step "different" pairs (1 versus 3, 2 versus 4, 3 versus 5, 4 versus 6, and 5 versus 7), and one six-step different pair (1 versus 7).

Training

Training involved 20-trial blocks of identification with feedback. We implemented a "perceptual fading" training method (Jamieson & Morosan, '86, Morosan & Jamieson, '86, Terrace, '63), in which we progressively reduced the acoustic difference between the endpoint sounds as a function of successful performance. This approach allows subjects, 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 endpoint stimuli. Performance over the course of training itself provided us with a "learning" measure, in addition to the pre- and post-training identification and discrimination test measures.

Training began with the identification of the endpoint stimuli 1 and 7. 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 subject achieved criterion on this last contrast, or once they had completed a maximum of 200 trials (10 blocks). Discrimination and identification tests were re-administered at the end of training.

Psychophysical analyses

Laboratory training studies using synthetic phonemes typically involve creating a continuum of sounds, such that the endpoint sounds are reliably perceived exemplars of the two contrasting phonemes in question, and that there is an equal, controlled physical difference between each of the adjacent sounds in the series. Typically, before training, individuals can neither identify the endpoint sounds, nor can they discriminate pairs of sounds from the series. Identification is tested following training by presenting the sounds of the series one at a time, and successful training is reflected by the ability to reliably label one endpoint as one of the phonemes, and the other endpoint as the other phoneme. A steeper identification function indicates better performance than a less steep function. Successful discrimination is typically reflected by a discrimination function that indicates better performance across the phoneme boundary, and poor performance in discriminating two sounds belonging to the same phonetic category. For example, if the perceptual boundary is between stimuli 3 and 5 for a particular subject, they should be able to say that sounds 3 and 5 are "different", whereas they should find it difficult to discriminate sounds 1 and 3.

We developed a continuous learning rate variable ("L-scores") to quantify subjects' performance over the course of learning. This variable takes into account both the number of blocks at each level of difficulty and the accuracy of performance during 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 hardest blocks, respectively. Resulting values were then subtracted from 600, which was the maximum possible value, yielding an "L-score". This learning can range from zero (slowest possible learner) to 600 (fastest possible learner).

For each of the three stimulus types, we measured: a) pre-training and post-training discrimination performance (A'), b) pre-training and post-training identification slopes, and c) L-scores, reflecting rate of learning.

RESULTS

Discrimination performance

Each subject's performance was converted to an A' score, an index of sensitivity which corrects for individual differences in bias. A' is a non-parametric analogue 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, '72). Figure 2 illustrates the group average pre- and post-training A' scores.

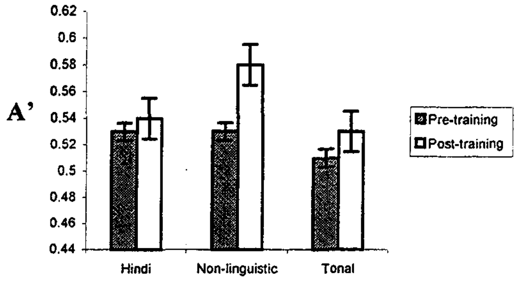


FIG.2. Pre-training and post-training discrimination performance

A 2X3X5 three-way repeated measures analysis of variance (ANOVA) was performed on the A' results, the within subjects factors being time (pre- vs post-training), stimulus type (Hindi, nonlinguistic, and tonal), and stimulus pair (1/3, 2/4, 3/ 5, 4/6, 5/7). Results showed significant main effects of time (F(1, 58)= 8.99, p<.005), and of pair (F(4, 232)=3.50, p<.01, Greenhouse-Geisser correction for all ANOVA results). The effect of stimulus type was not significant. The main effect of time reflects significantly better performance after relative to before training. In order to further analyse the effect of training, we performed tests of simple main effects on the effect of training for each stimulus type. We found that there was a significant improvement due to training in the ability to discriminate the rapidly changing non-linguistic stimuli (F(1,58)= 11.02, p<0.01), but that there was a non-significant improvement for the tonal (F(1,58)= 1.85, p>0.05) and for the Hindi stimuli (F(1,58)= 0.03, p>0.05).

Newman-Keuls post-hoc tests on the main effect of stimulus pair revealed that discrimination performance for the second (Q=3.77, Qc(4,232)=3.63, p<0.05), third (Q=3.60, Qc(3,232)=3.31, p<0.05), and fourth (Q=4.66, Qc(5,232)=3.86, p<0.05) stimulus pairs were significantly better than that for the last stimulus pair. Refer to Figure 3 for mean performance for each stimulus pair. Visual inspection of this Figure as well as the post-hoc results suggest that the performance is categorical-like, in that performance is better for the three pairs in the middle of the stimulus continuum relative to that for the pair at the end of the continuum.

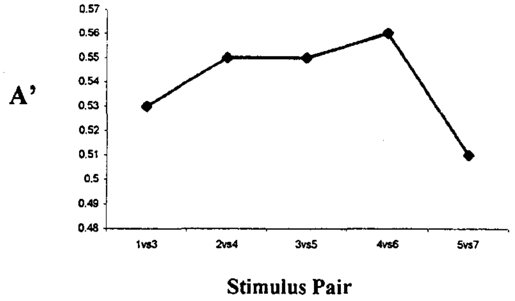


FIG.3. Main effect of stimulus pair

Correlational analyses revealed that within each stimulus type, pre- versus post-training performance was significantly correlated across subjects (Hindi: r=0.49, p<0.001, nonlinguistic: r=0.65, p<0.001, and tonal: r=0.56, p<0.001). Pre-training discrimination sensitivity was not related across stimulus types, whereas after training, only the A' values for the Hindi versus the non-linguistic stimuli were significantly correlated (r=0.35, p<0.05). The correlation between post-training performance on the Hindi and tonal stimuli was 0.25 (p>0.05). A test of part correlations revealed that there is a significant amount of unique variation associated between performance on the Hindi and non-linguistic stimuli while adjusting for performance on the tonal stimuli (t= 2.49, p<0.05). Refer to Figure 4 for a scattergram of post-training discrimination performance for the rapidly changing non-linguistic versus Hindi stimuli.

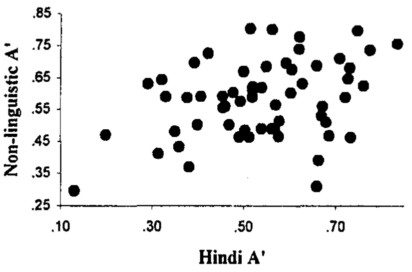


FIG.4. Scattergram showing post-training discrimination performance for rapidly changing non-linguistic versus Hindi stimuli.

Identification performance

Figure 5 presents the mean percentage of "A" responses for each of the seven stimuli during pre- and post-training identification of the three stimulus types. As described above, for the Hindi stimuli, responses "A" and "B" correspond to the dental and retroflex sounds, respectively.

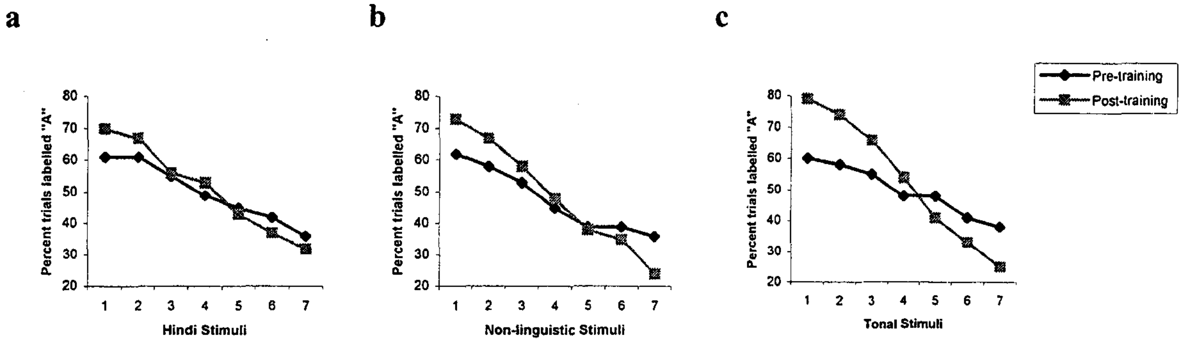


FIG.5. Pre-training and post-training identification functions for (a) Hindi stimuli, (b) non-linguistic stimuli, and (c) tonal stimuli.

The slope of individual subjects' identification curves and of identification curves averaged across subjects was calculated using a standard regression approach. Relatively steeper negative slopes reflect a relatively better ability to reliably identify the sounds in each stimulus set. Refer to Figure 6 for group average identification slopes.

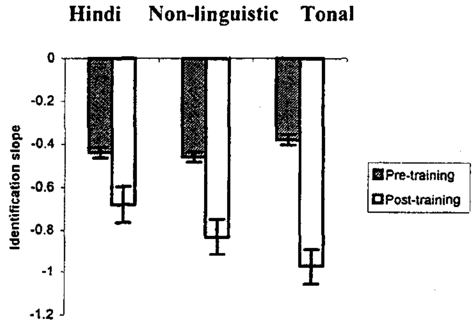


FIG.6. Pre-training and post-training identification slopes

A 2 X 3 two-way repeated measures ANOVA was performed on the identification slopes, with time (pre- vs post-training), and stimulus type (Hindi, nonlinguistic, and tonal) being the within subjects variables. Results revealed a main effect of time ($F(1, 58)=62.49$, GG-corrected, $p<0.001$), indicating that the post-training identification functions (slope, collapsed across stimulus type, was equal to -0.82) were steeper than the pre-training ones (mean slope= -0.43), reflecting improved performance after training. The effect of stimulus type was not significant. In addition, a significant interaction was found between time and stimulus type ($F(2, 116)(GG)=3.17$, $p<.05$). See Figure 6 for interaction means. We performed tests of simple in effects on the interaction. Results revealed that performance was significantly better after relative to before training for all three stimulus types (Hindi: $F(1,58)=8.47$, $p<0.01$, non-linguistic: $F(1,58)=16.56$, $p<0.01$, and tonal: $F(1,58)=26.94$, $p<0.01$). Tests of simple main effects on the effect of stimulus type at pre- and at post-training revealed that identification performance was similar across the three stimulus types before training ($F(2,116)=0.147$, $p>0.05$). After training, however, there was a significant difference in identification performance across the three stimulus types ($F(2,116)=3.23$, $p<0.05$). Newman-Keuls post-hoc tests revealed that identification performance was significantly better on the tonal relative to the Hindi stimuli ($Q=3.59$, $Qc(3,116)=3.36$, $p<0.05$).

Correlational analyses revealed that for each stimulus type, pre- versus post-training slopes were significantly correlated across subjects (Hindi: $r=0.50$, $p<0.001$, nonlinguistic: $r=0.59$, $p<0.001$, and tonal: $r=0.46$, $p<0.001$). During pre-training, there were no significant correlations in performance across the stimulus types, consistent with discrimination results reported above. At post-training, there were significant correlations between identification slopes on the Hindi and the nonlinguistic stimuli ($r=0.36$, $p<0.05$), as well as between the Hindi and the tonal stimuli ($r=0.30$, $p<0.05$). A test of part correlations revealed that there is a significant amount of unique variation associated between performance on the Hindi and non-linguistic stimuli while adjusting for variation associated with the tonal stimuli ($t=2.80$, $p<0.01$). Refer to Figure 7 for a scattergram of post-training identification performance for the rapidly changing non-linguistic versus Hindi stimuli.

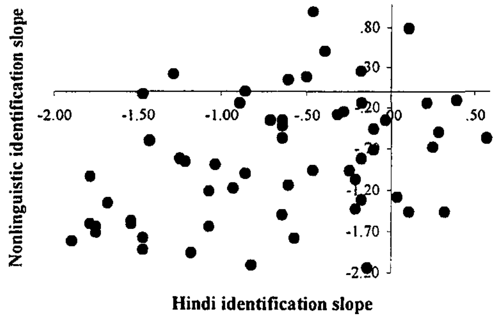


FIG.7. Scattergram showing post-training identification performance for rapidly changing non-linguistic versus Hindi stimuli.

Training results

Figure 8 illustrates the degree of variability in performance during training. The graphs show the frequency distributions for the L-scores for each of the stimulus types. It can be seen that there is a considerable range in the rate of learning for each of the three stimulus types. The following are the frequencies of subjects who achieved criterion performance on the third, most difficult training level over the course of training in 200 or fewer training trials: Hindi: 38/59, nonlinguistic: 33/59, and tonal: 29/59.

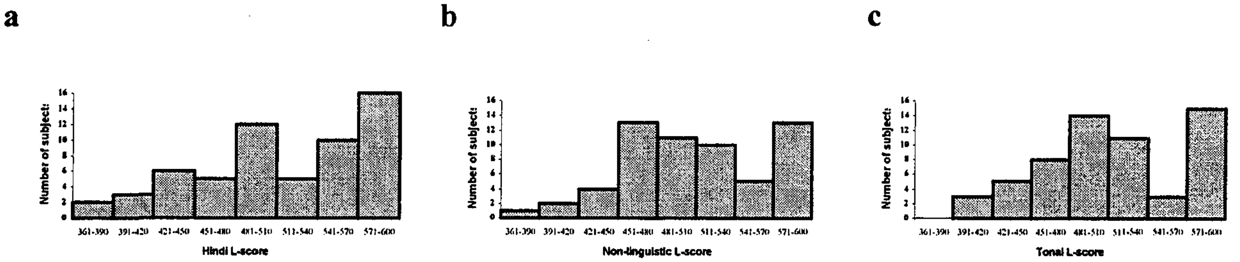


FIG.8. Frequency distributions of L-scores (N=59) for: (a) Hindi, (b) non-linguistic, and (c) tonal stimuli.

DISCUSSION

Two main findings emerge from our results. Firstly, as predicted, we showed that training results in overall group improvement in the ability to identify but not to discriminate the "difficult" non-native dental-retroflex contrast. There was a large amount of variability in performance across subjects. Secondly, we found support for the hypothesis that successful phonetic learning is in part a function of a more general ability to process rapidly changing sounds, regardless of whether or not they are linguistic.

The dental-retroflex contrast is difficult for non-native speakers to learn in adulthood, and it has been suggested that the difficulty in learning this contrast results from a lack of early exposure during development and from a lack of allophonic exposure to the retroflex sound in the native language environment. This hypothesis is consistent with the finding by Tees and Werker ('84) that adults having only been exposed to Hindi at a very early age and not later were able to discriminate this contrast prior to studying the language during adulthood, whereas that adults who had not had early exposure to the retroflex sound did not improve significantly in their ability to discriminate this contrast even after one year of second language experience during adulthood. Results of previous laboratory studies have shown that training using natural exemplars are effective only when the task or stimuli are modified so as to decrease working memory load during task performance (Werker et al, '81, Werker and Tees, '84a, Tees & Werker, '84, Pruitt et al, '90, Werker and Logan, '85). In contrast to the results of these training studies, we succeeded in showing an overall training-related group improvement in identification performance, though not in discrimination. Our sample size was larger than that included in previous training studies using the Hindi dental-retroflex contrast. We therefore had sufficient statistical power to detect even subtle improvements in performance. In addition, previous training studies with the dental-retroflex contrast have employed naturalistic stimuli. The use of synthetic stimuli in this study allowed us to manipulate only the acoustic parameters that are relevant in distinguishing the contrasting sounds, while keeping constant ones which are not necessary for differentiating them. The lesser degree of acoustic variability between our sounds relative to those used in previous studies likely allowed our subjects to attend to the acoustic parameters that are critical for distinguishing the dental and retroflex sounds, and thereby facilitated the training of this contrast. We did, as predicted, find considerable variability in both pre- and post-training performance, as well as in the learning rate during training, across subjects. Our results suggest that the ability to distinguish this difficult contrast is not permanently lost during development. This is consistent with findings in electrophysiological studies that native English speakers pre-attentively perceive the difference between the dental and retroflex sounds (Rivera-Gaxiola et al, 2000a, Rivera-Gaxiola et al, 2000b), despite not being able to differentiate the sounds in a behavioral detection task.

We had made predictions of a relationship between subjects' performance on the Hindi and on the non-linguistic rapidly changing stimuli, and of a dissociation in their performance on these two stimulus types versus the steady-state tonal ones. This prediction was supported by results of the post-training discrimination tests. Results of post-training identification tests were less clear; we found not only a relationship between performance on the Hindi and on the rapidly changing non-speech sounds, but also a relationship between performance on the Hindi and the tonal stimuli. Part correlations, however, revealed that there was significant variation in performance shared uniquely by the Hindi and non-linguistic stimuli when adjusting for variation associated with performance on the tonal stimuli. The finding of a relationship between the speech and rapidly changing non-speech sounds supports the idea that perceptual plasticity of the phonetic space is at least partly related to the more general psycho-acoustic ability to learn to track rapidly changing sounds over time. The additional relationship between performance on the Hindi and tonal stimuli may be a function of individual differences in general attentional and motivational factors, which are likely to influence performance on different stimulus types in a similar manner across subjects. There was no relationship between performance on any of the three stimulus types before training. Performance was poor before training, and we had conducted pilot tests in order to equate difficulty across stimulus types and to avoid pre-training ceiling effects. We suggest that pre-training floor effects made any possible cross-stimulus relationships in performance undetectable.

Taken together, our findings suggest that even if individuals are not exposed to certain non-native speech sounds either at an early age, or allophonically in the native language environment, some individuals can successfully learn these sounds with feedback during training. Further, they suggest that the more general auditory ability to track rapidly changing nonspeech sounds partly predicts the potential for phonetic perceptual remapping across individuals in adulthood. The physical parameters that distinguish certain phonetic contrasts such as voicing and place of articulation contrasts typically occur within the first thirty to forty milliseconds of the speech sounds. The ability to perceive the difference between these sounds therefore depends critically on cognitive and neural mechanisms which allow one to process such rapidly changing sounds. There is evidence from behavioral, functional imaging, and dichotic listening studies that the ability to process speech sounds is related to the ability to parse rapidly changing acoustic information. Our results extend these findings, and show that not only baseline phonetic perception but also *potential for change* in phonetic perception may be related to low-level psychoacoustic temporal processing.

Chapter 3

Study 2: *Learning new sounds of speech: Reallocation of neural substrates*

Infants aged six months or younger are able to discriminate speech sounds, including many that are not used to distinguish words in their native language. However, during development and starting as early as at six months of age, lack of experience with certain non-native speech sounds results in a developmental shift from a language-general to a language-specific pattern of phonetic perception (Best et al, '88, Polka & Werker, '94, Werker & Tees, '84a, Werker & Lalonde, '88, Jusczyk, '95, Kuhl et al, '92, Kuhl, 2000). Most adults can better distinguish two speech sounds belonging to different phonetic categories than ones belonging to the same category, even when the physical differences separating the stimuli have been equated (Liberman, '57, Liberman et al, '57, '67, Fledge, '84, Pisoni et al, '82, Werker & Tees, '84b). Despite native-language phonetic perception, adults are capable of learning new languages, and thereby of learning to distinguish non-native phonetic contrasts. Interestingly, even amongst adults with very similar language backgrounds, considerable individual differences exist in their ability to improve following phonetic training (Strange & Dittman, '84, Polka, '91, Pruitt et al, '90, Strange et al, '89, Werker et al, '81). This finding leads to important questions regarding the functional neural substrates underlying the perception of native versus newly learned, non-native speech sounds, and more specifically, regarding possible differences in functional anatomy between individuals who successfully learn new speech sounds and those who do not benefit from training.

The neural correlates of phonetic perception have been studied using functional brain imaging techniques such as PET and fMRI. These experiments have involved auditory presentation of stimuli including words, speech syllables, and meaningless speech sounds, and tasks used have included passive listening, phoneme monitoring, discrimination, or identification, and rhyming judgments. Generally, the results have shown the involvement of regions in and around what is classically known as "Wernicke's area", including left-sided activations in perisylvian temporoparietal areas including the supramarginal and angular gyri (Démonet et al, '94, Zatorre et al, '92, '96, Binder et al, '96, '97, Petersen et al, '88, Paulesu et al, '93). Consistent with functional imaging work, there is also evidence from lesion studies that deficits in phonological processing may arise from damage to perisylvian regions in and around Wernicke's area, including the left superior temporal gyrus and the supramarginal gyrus (Geschwind, '70, '71, Benson, '67, Benson et al, '73). Results of functional imaging work specifically examining phonetic perception have also typically shown activity in the superior temporal gyrus (STG) bilaterally (Binder et al, '94, Mazoyer et al, '93, Jäncke et al, '98, Mummery et al, '99). The precise function of secondary auditory regions in phonetic processing is not known, but it is thought to be specifically involved in aspects of speech processing, since it is not activated by simple tones or noise stimuli (Zatorre et al, '92, Démonet et al, '92, '94a, Binder et al, '96, '97, see Zatorre & Binder, '00). It has been suggested that it is involved in analysis of speech sounds leading to comprehension at the syllabic or whole-word level (Zatorre et al, '92, '96). Similarly, the exact function of the parietal region in phonetic processing is not known. Paulesu and colleagues ('93) suggested that the left supramarginal gyrus subserves the phonological short-term memory storage component of Baddeley's ('86) proposed "articulatory loop" model of verbal working memory.

The involvement of regions in and around the frontal speech area classically known as Broca's area in phonological processing has been the subject of controversy. Results of some studies involving receptive speech-related tasks have not shown activation in this region (Petersen et al, '89, Rumsey et al, '92). In contrast, a larger number of studies have shown its involvement in purely receptive language tasks that make certain specific demands (Zatorre et al, '92 & '96, Démonet et al, '92, '94b, Fiez et al, '95, Burton et al, '00). In addition, although speech perception has not extensively been investigated in aphasic patients with lesions in and around Broca's area, existing studies have shown deficits in phonetic discrimination (Blumstein et al, '77, Tallal & Newcombe, '78), and in temporal perception (Tallal & Newcombe, '78).

There are a number of hypotheses concerning the role of frontal speech regions in phonological perception. For example, Paulesu and colleagues ('93) have proposed that this area subserves subvocal rehearsal in verbal working memory. It has also been suggested that the left posterior temporal area is responsible for the initial stage of phonetic analysis, but that the frontal speech areas are additionally involved when portions of the speech signal are related to articulation during the active performance of tasks involving phonetic segmentation and/or working memory (Zatorre et al, '92, '96). Examples of such tasks include phonetic discrimination on pairs of syllables (Zatorre et al, '92, '96), phoneme monitoring (Zatorre et al,'96), or phonetic discrimination requiring overt segmentation of initial consonants from the rest of the word (Burton et al, 2000).

The neurophysiological correlates of phonetic learning have also been studied using ERP methodology. The mismatch negativity (MMN) evoked potential is an auditory cortical response to acou stic change that is introduced in a repetitive stimulus sequence (Näätänen et al, '78 & '93). It has been shown that behavioral training of two slightly different speech stimuli in adults results in a significant change in the duration and magnitude of this cortical potential (Kraus et al, '95), and that this physiological change precedes behavioral discrimination improvements (Tremblay et al, '98), suggesting that MMN is a measure of pre-attentive learning (see Kraus & Cheour, 2000). More generally, other techniques such as single cell recordings in animals (Recazone et al, '93, Kraus & Disterhoft, '82) and magnetoencephalography (Pantev et al, '99) have shown plasticity of auditory cortex function resulting from training and experience. The mismatch generators are thought to involve the thalamo-cortical association areas of the auditory cortex (Giard et al, '90, Kraus et al, '94). In response to speech stimuli, some studies have shown the MMN response to be symmetrical (Aaltonen et al, '94, Tremblay et al, '97). Other studies, however, have shown slightly larger responses to vowels in the right hemisphere (Csépe, '95), and larger ones to consonants over the left hemisphere (Csépe, 95, Alho et al, '98). Interestingly, Tremblay and colleagues ('97) showed that MMNs elicited by nonnative speech syllables were initially symmetrical, but that they became enhanced over the left hemisphere following training. Taken together, these findings suggest that left-sided specialization for speech may be evident even in the representation of phonetic information. Based on the finding that phonetic training results in detectable neurophysiological changes, we predict that we will find a change in the BOLD signal accompanying behavioral improvements in the perception of a non-native phonetic contrast.

The aim of the present study was to determine how the pattern of brain activity may change as a result of training with speech sounds from a non-native language. Subjects were scanned using fMRI before and after a two-week period of phonetic training with a Hindi dental-retroflex contrast and with a native phonetic contrast as a control. A noise control condition was also used in order to subtract out lower-level acoustic processing of complex sounds, and to make the results more comparable to those of previous studies on phonetic processing (Zatorre et al, '92, Binder et al, 2000). We wanted to address the following questions. First, does the identification of newly learned speech sounds recruit the same neural substrates as does the identification of a known, native phonetic contrast, or are new areas recruited? The second question relates to whether we can differentiate "learners" from "non-learners" on the basis of their pattern of activation while they classify the new speech sounds. We predict firstly that the native identification task will reveal the bilateral involvement of superior temporal regions, stronger in the left than in the right hemisphere, of the left temporo-parietal region, and of the left frontal gyrus (IFG) in and adjacent to Broca's area. Second, based on the above reported lateralization of the MMN response to nonnative speech sounds following training, we predict that before training, the neural response to nonnative speech sounds will be bilateral, but that it will be more left lateralized after training. We also predict that after training, the pattern of activation outside of the auditory regions (ie; in the left temporoparietal and inferior frontal regions) will be similar to that found in the native condition. This prediction is also based on results of neuroimaging studies of language function in healthy bilinguals, showing that at the single word level, brain regions subserving the native language (L1) and the second language (L2) in fluent bilinguals appear to overlap (Klein et al, '94 & '95, Chee et al, '99, Illes et al, '99). Last, based on the assumption that more successful task performance recruits underlying neural substrates more actively, we predict that correlations between a behavioral learning measure and the blood oxygenation level dependant (BOLD) signal during the post-training non-native task will reveal a positive relationship between learning and signal in left pre-frontal and left temporo-parietal speech areas.

METHODS

Subjects Ten right-handed monolingual English-speaking participants (4 men), ranging in age from 20 to 29 participated in the study. None had been exposed to or had experience with languages in which the retroflex non-native speech sound is phonologically represented.

Stimuli

Stimulus Selection: We selected the dental-retroflex place-of-articulation contrast which is used in languages of India such as Hindi or Urdu. Retroflex consonants require a relatively complex articulation, they are rare across languages; only 11% of the world's languages include a retroflex consonant, and they are unlikely to be allophonically represented across languages (ie; to be pronounced in languages in which the contrast is phonetically irrelevant) (Burnham, '86). The non-native perception of the retroflex sound is lost early in development, below the age of one (Werker & Tees, '83, Werker & Tees, '84a, Tees & Werker, '84, see Burnham, '86), and perceptually, English listeners assimilate the dental-retroflex sounds such that they perceive both sounds as instances of the dental consonant (Polka, '91, Werker & Lalonde, '88).

Research aimed at training the non-native perception of the dental and retroflex sounds in adults has shown that standard training approaches fail to improve performance (Werker et al, '81, Werker & Tees, '83, Werker and Tees, '84a, Tees & Werker, '84). Improvement has been reported under certain circumstances, including the use of shorter inter-stimulus intervals (ISI) (Werker and Tees, '84b, Werker and Logan, '85), truncated stimuli (Pruitt et al, '90), and extensive experience with the contrast (Tees and Werker, '84).

We selected this non-native retroflex consonant which is rare and unlikely to occur allophonically across languages because we wanted to ensure that none of our participants had had either allophonic or non-native exposure to this sound. Additionally, we wanted to ensure that no individual would be able to distinguish this sound from the dental consonant before training (ie; to avoid pre-training ceiling effects), and that we would find a considerable range in improvement across subjects so that we could then examine brain activity as a function of success in learning. In a previous study (Golestani et al, unpublished), we synthesized the retroflex consonant, and showed firstly that phonetic training resulted in overall improvement in the ability to identify this sound versus the native dental one in a large group of subjects, and secondly, that there were considerable individual differences in the amount of learning across subjects.

Stimulus Synthesis: Synthesis of the non-native phonetic contrast was based on parameters reported by Stevens and Blumstein ('75) for the synthetic retroflex-dental place-of-articulation continuum. Refer to Figure 1 for schematized frequency-time representations of the dental and retroflex sounds. There were seven stimuli varying in equal steps in terms of acoustic difference between adjacent items. The set of four-formant stimuli was constructed with the use of the Mitsyn (Henke, '90) synthesizer. Stimulus 1 corresponds to the dental and stimulus 7 to the retroflex voiced, unaspirated stop consonant prototype, and all of the consonants were followed by the vowel /a/.

Stimuli began with an initial noise burst. The parameters that were manipulated to create the continuum are the frequency glides of the third formant (F3), as well as the center frequency of the burst. The initial noise burst lasted 5ms, and its center 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 40ms. The fundamental frequency (Fo) started at 130 Hz and decreased linearly 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 Hz to 655 Hz during the transition, and then remained steady-state during the remainder of the sound. The second formant (F2) decreased from 1650 Hz 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.

The native control stimulus contrast was constituted of the dental /da/ sound described above (stimulus #1), and a synthetic voiceless stop consonant /t/, followed by the vowel /a/. (/ta/). Stimulus characteristics for the latter are the following. The stimulus lasted 220 ms, and the FTs lasted 30ms. The initial noise burst lasted 10 ms, and its center frequency was 4000 Hz. Voicing began 50 ms after the onset of the burst. The Fo started at 130 Hz, decreased linearly for 40 ms to 125 Hz, remained steady state for 40ms, and then decreased linearly to 90 Hz over the last 90 ms of the sound. Formants 1, 2, and 3 remained steady state after the FT. F1 rose from 530 Hz to 655 Hz during the FT, F2 decreased from 1480 Hz to 1185 Hz, and F3 decreased from 2857 Hz to 2585 Hz. F4 remained steady-state at 3600 Hz during the entire stimulus.

The noise stimuli consisted of amplitude modulated white noise matched in sound pressure level (SPL), duration, and amplitude with the CV stimuli. They were generated by using the same wave envelope as the speech stimuli. The interstimulus interval for the noise condition was also matched with that of the phonetic condition.

Stimuli were presented binaurally with electrostatic, MR-compatible headphones (Koss) at an 87 dB sound pressure level (SPL) using Media Control Function (Digivox).

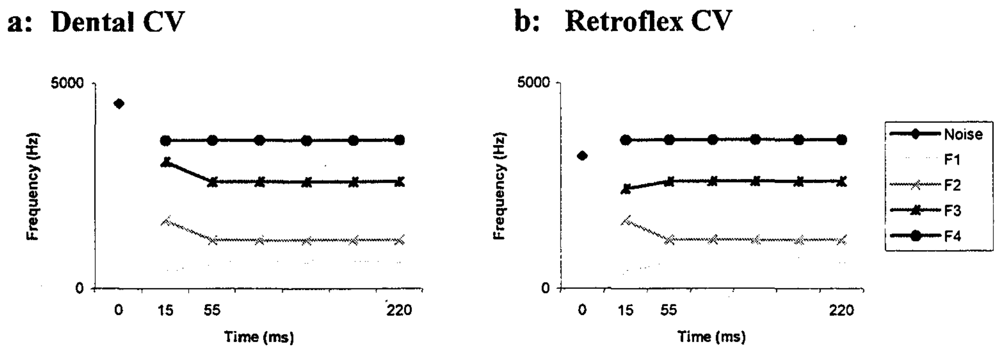


FIG. 1: Schematized frequency versus time diagrams of the (a) dental and (b) retroflex consonant-vowel (CV) syllables.

Scanning Protocol Imaging was performed on a 1.5-Tesla Siemens Vision scanner. After obtaining high resolution T1 anatomical scans, two series of 128 gradient-echo images of blood-oxygenation-level-dependent (BOLD) signal were acquired (TE = 50ms, head coil, matrix size: 64 X 64; voxel size: 5 X 5 X 5 mm; 14 slices acquired in the orientation of the Sylvian fissure). A long, 8 second interacquisition interval (TR) was used to ensure low signal contamination by noise artifacts of image acquisition (Belin et al, '99). Two stimuli were presented during each 8-second TR. The following are the temporal parameters of the stimulus presentation times relative to the acquisition. The acquisition lasted 1.4 seconds, and a 250 ms consonant-vowel (CV) stimulus was presented 0.1 seconds after the end of the acquisition period. There was then a 3 second silent period during which subjects made their classification response. The second CV was then presented, again followed by 3 seconds of silence, after which the next image was acquired.

Based on previous research (Belin et al, '99), we expected the optimal BOLD signal response in the primary auditory cortex (AI) to occur 2 to 3 seconds after stimulus presentation, and for the response in the secondary auditory cortex to occur approximately one second after the response in AI. The hemodynamic response to the first sound would therefore likely be contaminated by the scanning noise of the preceding acquisition; however, we did not expect that the following acquisition would detect the BOLD response to this first sound. The first sound was simply included in order to increase the number of trials that subjects performed.

Behavioral Testing and Training procedures

Identification task during pre- and post-training scans: During each of the two echo-planar imaging (EPI) series, 32 images were acquired during each of the following four conditions: native identification, non-native identification, noise baseline, and silent baseline. The conditions were distributed such that each of the three conditions was presented during sets of four consecutive acquisitions, followed sometimes by one and sometimes by two silent baseline acquisitions. Conditions were pseudo-randomly ordered. Prior to the initial scanning session, subjects were briefly familiarized with the stimuli in order to ensure that they could accurately identify the two native sounds (/da/ and /ta/) and that they could not distinguish the non-native retroflex /da/ sound from the native dental /da/ sound. During the native identification task, subjects were required to identify the /da/ versus /ta/ sounds and to indicate their response by pressing one of two mouse button keys. During the non-native condition, subjects heard either the dental or retroflex prototypes (ie; stimuli 1 or 7, respectively). Prior to training, no subjects could distinguish the dental from the retroflex CVs. For this reason, although they were engaging in the classification task in that they were prepared to press the key corresponding to /ta/ if and when it was presented, during the pretraining non-native condition, subjects in effect pressed the mouse key corresponding to the dental /da/ sounds after each stimulus. After training, they were required to identify the dental versus retroflex sounds (by again pressing mouse button keys). During the noise baseline condition, subjects were to press alternating mouse button keys to the noise bursts.

Phonetic Training: Five one-hour sessions of identification training with feedback spread over the course of two weeks every 3 days on average were administered between the two scanning sessions. During the first session, subjects heard endpoint stimuli (sounds 1 or 7), and indicated their response by pressing a mouse button. Feedback was presented on the computer screen as to the accuracy of their response. If criterion performance, defined as at least 16/20 correct responses on a block, occurred during three consecutive training blocks, then the task was made more difficult by requiring subjects to identify sounds 2 versus 6 (these sounds are a smaller step-size apart, and are acoustically more similar than are the prototypes). The subjects who achieved criterion on this 5-step pair were then trained on an even more difficult 3-step pair, using stimuli 3 versus 5. This "perceptual fading" training method has been shown to result in behavioral improvement in previous studies (Jamieson & Morosan, '86, Morosan & Jamieson,

’86). It allows individuals who successfully learn the contrast early during the course of training to then learn additional information regarding more subtle perceptual aspects of the acoustic parameters which critically distinguish the prototypical stimulus pair which they will later be tested with.

Analysis BOLD signal images were smoothed (6-mm FWHM), corrected for motion artifact and transformed into standard stereotaxic space (Talairach & Tournoux, ’88) using in-house software (Collins et al, ’94). Statistical t-maps were obtained in each individual using a voxel-based analysis based on a linear model with correlated errors, with direct specification of the design matrix (Worsley et al, ’96). Output across runs within sessions was combined, and then that of subjects within a population in a hierarchical random effects analysis. Runs were combined with another linear model for the run effects, weighted inversely by the square of their standard errors. The t-map images where then registered with a standardized anatomical image, averaged across the 9 subjects. Criteria of $t=4.5$ in the whole brain volume and of $t=4.3$ in the regions of interest (ROIs) were calculated (Worsley *et al.,* 1996) based on 125 degrees of freedom, a voxel size of one cubic millimeter, smoothness of 6-mm, 1000cc, and a significance level of $p<0.05$. The volume of interest for the whole brain was 1000cc. The ROIs included bilateral superior temporal regions, the left temporoparietal region, and the left inferior frontal region. These ROIs were defined anatomically as cubical volumes of interest based on the upper and lower x, y, and z-value coordinates for these structures derived from a standardized stereotaxic atlas (Talairach & Tournoux, ’88). Superior temporal coordinate limits were: x= 30 to 70 and –30 to 70, y= 25 to –40, and z= 20 to –35. Left parietal limits were: x= 0 to –42, y= -40 to –80, and z= 20 to 75. Finally, left frontal limits were: x= -30 to –60, y= 10 to 60, and z= -2- to 30. The total volume of interest for the regions of interest was 556cc.

RESULTS

Behavioural results

One out of the ten subjects did not respond to a large proportion of the pre- and post-training identification trials during scanning, therefore we excluded this subject’s results from all of the analyses. Each subject’s performance was converted to an A’ score, which is a nonparametric unbiased index of sensitivity that ranges from 0 - 1.0, with 0.5 corresponding to chance performance (McNichol, '72). Mean A' scores and standard errors for the native and the non-native contrasts at pre- and at post-training are provided in Figure 2 below. A (2X2) two-way repeated measures ANOVA, with time (pre- and post-training) and stimulus (native and non-native) being the two independent variables, was performed on the A' scores. Results revealed a significant main effect of stimulus (F (1,8)=61.38, $p<0.001$), indicating that overall, performance was better for the native control condition than for the non-native one, and a significant time by stimulus interaction (F (1,8)=13.34, $p<0.01$). Post-hoc Tukey tests on the significant interaction revealed that the post-training non-native mean A' measure was significantly higher than the pre-training one (Q (4,8)=8.02, $p<0.05$), indicating that subjects' performance improved as a result of training while the native performance remained near ceiling. There was a large range of individual differences with respect to the amount of improvement resulting from training; post-training A' values for the identification of the non-native contrast ranged from 0.30 to 0.87.

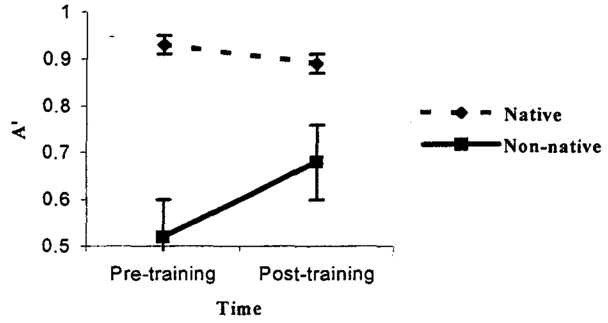


FIG. 2: Identification performance during pre- and post-training scans.

Imaging results

Group Subtractions

Table 1 provides the stereotaxic coordinates for the pre-training and post-training native minus noise voxel-based subtractions, respectively (Fig. 3). During the pretraining scan, the native minus noise subtraction yielded significant positive activations bilaterally in the insula/frontal opercular (FO) regions, in the superior temporal gyri, inferior frontal gyri, as well as in the left superior parietal gyrus. Superior temporal activations, based on inspection of probabilistic maps for Heschl's gyrus (Penhune et al, '96) and for the planum temporale (Westbury et al, '99), appeared to be at the junction of these two structures.

The same subtraction after training (Table 1, Fig.3) yielded a similar pattern of results with the exception of additional activations in the left caudate nucleus, and in the right inferior frontal gyrus, at a location not symmetrical to the left-sided inferior frontal activation.

Table 2 and Figure 4 present the pre- and post-training results for the non-native minus noise subtraction. During pretraining, the only significant result was a positive activation in the left insula/frontal operculum. After training, the subtraction yielded positive activations bilaterally in insula/frontal opercular areas, as well as in the left IFG, STG, and caudate nucleus. The right STG activation was marginally significant. Additionally, there was a right IFG activation which was once again considerably more anterior than that in the left hemisphere.

In order to directly compare activity associated with identifying the non-native contrast before versus after training, we subtracted the pre-training non-native minus noise difference image from the post-training one. In this compound comparison, we found significant activation in the left and right inferior frontal gyri, in the left superior parietal gyrus, in the right globus pallidus, and in the left caudate nucleus. Additionally, there was marginally significant activation in the left supramarginal region (refer to Table 3). We also performed such a control compound comparison; we subtracted the post-training native minus noise difference image from the pre-training one. This comparison did not reveal any significant differences.

Correlation of brain activity with learning

In order to examine the relationship between brain activity and the degree of learning across individuals, we performed a correlational analysis between post-training A' minus pretraining A' scores and the BOLD signal in the whole brain volume during the post-training non-native minus noise subtraction (refer to Table 4). We found positive correlations between this behavioral measure of learning and activity in regions within the left and right angular gyri. Refer to Figure 5 for an image of this result and for a scatter diagram showing the relationship between the voxel value at the peak voxel-value location and the degree of learning across subjects. Additionally, there were marginally significant correlations between the learning measure and activity in bilateral insula/frontal opercular areas, at similar locations to the signal found in the group subtraction results. There was also a significant inverse relationship between the learning measure and activity in the posterior part of the left middle temporal gyrus (see Fig.6)

DISCUSSION

Behavioral results

The behavioral results followed the expected pattern. They indicate that the training procedure was effective in producing an overall improvement in subjects' identification of the dental-retroflex contrast during the post-training relative to the pre-training fMRI test sessions, although not all subjects learned to the same extent. This finding is consistent with results of a previous behavioral study (Golestani et al, unpublished), in which we showed that training with the dental-retroflex Hindi contrast improved the ability of a larger group (n=59) of non-native adult speakers to identify the dental and retroflex sounds.

Imaging results

Classifying native speech sounds As predicted, during both pre- and post-training scans, the native minus noise comparison yielded a similar pattern of activation in areas including the left insula/frontal operculum, the left inferior frontal gyrus, and the left superior parietal region, and bilaterally in the superior temporal gyri. Unexpected findings were right insula/frontal opercular and inferior frontal activations during both scans (refer to Table 1). During the post-training scan (refer to Table 1), additional activations were observed in the right inferior frontal gyrus and in the right thalamus. A compound analysis involving the subtraction of the pre-training native minus noise difference image from the post-training one revealed no significant results, showing that, as expected, these differences were minimal.

The bilateral superior temporal activation is consistent with previous findings of bilateral STG involvement in processing both speech and certain types of nonspeech sounds (Binder et al, '94, '96, Jäncke et al, '98, Zatorre et al, '92 & '96, see Zatorre & Binder, 2000). Zatorre and colleagues ('92) found that regions of the STG and superior temporal sulcus (STS) anterior to the primary auditory regions bilaterally, and posterolateral to the primary auditory cortex on the left side, are activated by speech stimuli but not by acoustically matched noise. They suggested that the primary auditory cortex contributes to the early acoustic processing of all auditory stimuli, whereas the associative areas are involved in higher order processing of sounds. Other groups have also shown that areas in secondary auditory regions including regions of the temporoparietal cortex, of the superior temporal cortex, and of the superior temporal sulcus (STS) are not activated by simple auditory stimuli such as clicks, tones, or noise bursts (Zatorre et al, '92, Démonet et al, '92, '94a, Binder et al, '96, '97). Our finding of STG activation in secondary auditory regions outside of Heschl's gyrus is consistent with these results, and suggest that these areas are activated by the phonetic condition more than by the noise condition.

We found two distinct frontal activations; one in BA 44, in what we will refer to as the inferior frontal gyrus (IFG), and the other in what we will refer to as the insula/frontal operculum (FO). Previous functional imaging studies examining phonetic perception have more typically shown activation in regions that converge around the location of our "IFG", BA 44 activation (Zatorre et al, '92, '96, Burton et al, '00, Démonet et al, '92, '94b). Other studies, however, have shown activation at locations similar to our second insula/FO activation (Fiez et al, '95, Chee et al, '01). These two regions likely represent functionally distinct subregions within Broca's speech area. This interpretation is consistent with evidence that activations in this frontal speech region is not confined to the classic Broca's area (Binder et al, '97), and that there are functionally distinct subregions within the left inferior prefrontal cortex (Buckner et al, '95, see Fiez, '97 and Zatorre et al, '96).

According to the classical models of speech, the left inferior frontal "Broca's area" is thought to be involved in the programming and preparation of speech output. There is accumulating evidence, however, that regions in and around the IFG are activated during purely receptive speech tasks such as phonetic discrimination on pairs of syllables (Zatorre et al, '92), sequential phoneme monitoring of nonwords (Démonet et al, '92), and phonetic discrimination requiring overt segmentation of initial consonants from the rest of the word (Burton et al, '00).

Explanations have been offered for the role of the left IFG in such non-productive language tasks. A prominent explanation has been derived from Baddeley's "articulatory loop" theory (Baddeley, '86) of verbal working memory. Paulesu and colleagues ('93) examined the neural correlates of two components of Baddeleys model, and suggested that the left supramarginal gyrus is the location of the phonological store, and that Broca's area is involved in subvocal rehearsal. Démonet and colleagues ('94b) interpreted results from one of their studies in a manner consistent with this; they found that when phoneme monitoring tasks involved perceptual ambiguity, there was activation in the IFG. They attributed this to a greater reliance on verbal rehearsal strategies in working memory during the ambiguous relative to the unambiguous phoneme monitoring conditions. Another prominent explanation for the role of frontal areas derives from the motor theory of speech perception (Liberman and Mattingly, '85), which proposes that phonetic perception arises from a system which detects articulatory gestures, and which predicts activation of speech production substrates during certain purely perceptual speech tasks. Consistent with this theory, Zatorre and colleagues ('92 and '96) propose that the left posterior temporal regions may be involved in the initial stage of phonetic analysis during passive listening to speech, while the regions including Broca's area are additionally involved when phonetic segments must be extracted and manipulated in relating the phonetic information to articulation. This hypothesis was supported by an fMRI study by Burton and colleagues (2000), in which results suggested that the IFG is recruited during phonological processing tasks only when these latter require either phonetic segmentation or working memory processes.

As already mentioned, several imaging studies have also shown activity at locations closer to our insula/FO activations (Fiez et al, '95, Chee et al, '01). For example, Fiez and collegues ('95) showed that there was bilateral FO activity when subjects were required to detect stimuli including words, syllables, and tone sequences that incorporated rapid temporal changes. This second frontal region is also thought to be involved in aspects of articulatory recoding during phonological processing. Fiez and Petersen ('98) reviewed neuroimaging studies on word reading, and found convergence of results in FO areas at locations very similar to ours. Of interest is the finding that several groups found that this area was activated when subjects read low frequency exception words and when they read pronounceable non-words, but not when they read low-frequency consistent words (Herbster et al, '97, Fiez et al, unpublished data (see Fiez & Petersen, '89), Rumsey et al, '97). Fiez and Petersen suggested that reading of the former two classes of words may activate phonological processing (the 'assembled" route, where the word is "sounded out" by translation of letters to sounds), whereas that the latter type of word may involve the "direct route" for reading, in which the entire word form is translated into phonological representations. Once again, it is likely that during the performance of our identification task, the insula/FO and IFG regions subserve distinct functions related to phonetic processing. Future studies may be designed to elucidate the differential roles of these regions.

Right frontal activations

Although we did not expect right frontal activations, this result is consistent with those of several previous studies. Fiez and colleagues ('95) found bilateral frontal opercular activation at locations very close to those in our group subtractions while subjects detected words, syllables, and tone sequences that incorporated rapid temporal changes. Chee and colleagues ('01) showed, in a group of bilingual individuals, that there was greater activation bilaterally in frontal opercular regions during a semantic task in the least proficient of two languages. Right prefrontal regions have been implicated in functions related to aspects of pitch and tonal working memory processing (Zatorre et al, '92, Zatorre & Samson, '91). It is possible that during the performance of our tasks, subjects were processing the sounds partly in an acoustic mode, thereby recruiting right frontal regions. This explanation may in part explain findings by Mazoyer and colleagues ('93), who performed a PET study in which subjects listened to speech in the native language and in an unknown language, to semantically anomalous sentences, and to sentences with pseudo-words. They found that only when listening to the sentences with pseudo-words was there an additional activation in the right inferior frontal gyrus.

Learning new phonetic contrasts

During the pre-training scan, the non-native versus noise comparison yielded significant activation only in the left insula/frontal operculum (Table 2 and Fig. 4). As predicted, after training, the pattern of activation in this comparison was more similar to that in the native versus noise comparison (Table 2 and Fig.4). Significant peaks were observed bilaterally in the insula/frontal opercula and in the left inferior frontal and superior temporal gyri. The right STG signal was marginally significant. Activity was also detected in the left caudate nucleus, and in the right IFG, at a location not symmetrical to the left-sided inferior frontal location. Our findings have implications for neuro-linguistic theories of second language learning. As mentioned in the introduction, the ability to distinguish certain non-native speech sounds from native ones is lost during development (Best et al, '88, Werker et al, '81, Werker & Tees, '83, Werker & Tees, '84a). There is evidence that the dental-retroflex distinction is lost very early in development, below the age of one (Werker & Tees, '83, Tees & Werker, '84). Our results suggest that learning a new speech sound in adulthood results in the recruitment of the same neural substrates as those involved during the classification of native speech sounds. These findings are based on the learning of a single non-native consonant-vowel (CV) syllable. We predict that we would have found similar results had several non-native CVs been trained; these results may or may not generalize to the lexical and semantic aspects of learning a new language. In addition, short-term learning effects such as the ones that we studied may not generalize to expert performance, nor to performance with naturalistic exemplars of the non-native retroflex sound.

Change in brain function associated with learning new speech sounds

In order to confirm the effects seen in the subtraction analyses presented above, we performed a compound analysis in which we subtracted the pre-training non-native minus noise difference image from the post-training one. We found significant activity bilaterally in the inferior frontal regions, as well as in the left superior parietal region, the right globus pallidus, and the left caudate nucleus. There was also marginally significant activity in the left supramarginal gyrus. The inferior frontal activations were at similar locations to those found in the post-training native-minus-noise group subtraction reported above (Table 1), and suggest that classifying newly learned speech sounds recruits the inferior frontal regions bilaterally, as does the classification of native speech sounds. This analysis does not, however, support the group average non-native-minus-noise finding described above suggesting that the second, more inferior insula/frontal opercular (FO) region is also recruited when newly learned speech sounds are classified. As already reported, we found large individual differences in the behavioral performance of our subjects; only approximately half of our subjects improved in their ability to classify the sounds after training. It is possible that activity in certain brain regions such as the insula/FO increased for some subjects and decreased for others as a function of success in learning, and that such learning-related differences have cancelled each other out such that they are not detectable in this compound comparison. This interpretation was confirmed in the correlational analysis to be described below.

Correlations of brain activity with learning

We found positive correlations between a behavioral learning measure and activity in the left and right angular gyri (Fig.5). There were also concurrent marginally significant negative correlations with activity in insula/frontal opercular regions bilaterally, as well a significant negative correlation with activity in the posterior left middle temporal gyrus. Refer to Figure 6 for these results, as well as for scatter diagrams showing the relationship between the voxel value at the peak voxel-value location and the degree of learning across subjects. The negative correlation between the BOLD signal in the left middle temporal gyrus and the amount of learning suggests that in more proficient learners, there is less activity in this brain region relative to in non-learners. A number of studies have shown involvement of the left middle temporal region during the performance of language-related tasks involving lexico-semantic processing (Démonet et al, '92, '94a, Binder et al, '96, '97). Binder and colleagues (2000), using fMRI, compared patterns of auditory cortex activation in response to different types of speech stimuli. Based on their results, they suggested that bilateral secondary auditory superior temporal regions play a role in processing simple temporally encoded auditory information, that bilateral superior temporal sulci play a role in the acoustic processing of words, pseudowords, and reversed speech, and finally, that left-lateralized ventral temporal regions including the middle temporal gyrus play a role in processing lexical-semantic or syntactic information associated with words. These results implicate the left middle temporal gyrus as a left-lateralized component of a bilateral language network. Our finding suggests that activity in this speech-related region is modulated by the degree of phonetic learning.

Angular gyrus lesions are well known to produce language deficits, particularly for written material (Rumsey et al, '99, Horwitz et al, '98, Galaburda et al, '85, Geschwind, '65), and many PET studies have demonstrated activation associated with phonological processing in the left temporoparietal regions outside the superior temporal gyrus, including in the angular gyrus, supramarginal gyrus, and middle and inferior temporal gyri (Démonet et al, '94a, Zatorre et al, '92, '96, Binder et al, '96, '97, Petersen et al, '88, Paulesu et al, '93). The strong positive correlation between learning and activation in the left angular gyrus supports the idea that "good learners" recruit these more posterior temporoparietal regions relatively more than do "poor learners", and that activity in this region is modulated by learning. As will be further discussed, activity in the insula/FO is also modulated by learning, and it is possible that functional connectivity between these two regions is modified by learning. It has been suggested, based on functional imaging data, that normal functional connectivity between the left inferior parietal regions and Broca's area is disrupted in dyslexic individuals (Paulesu et al, '96). Functional connectivity may or may not indicate anatomical connectivity between these two regions, since activity in these regions may be modulated by input from a third part of the brain.

We had expected a positive correlation between learning and insula/frontal opercular activation because we had predicted that more successful task performance would place greater processing demands on speech regions. In contrast, we found an inverse relationship with bilateral activity in the same insula/frontal opercular regions as were detected in the group subtraction results. This finding suggests that although all subjects recruit these areas for performing the task, learners engage frontal speech regions to a lesser extent than do poor learners. Individual differences in activity in this region likely resulted in a training-related group average signal that was too weak to be detected in the compound comparison analysis reported above. These results are based on correlations, and do not inform us about the direction of causality, if any, between differences across subjects in brain activation and in performance. We will, however, speculate about the mechanism of this function-behavior relationship, and suggest a number of possible explanations for this finding. The first is that in learners, processing in these frontal regions may be more efficient and/or may require less processing time than in non-learners. Several mechanisms may underlie this difference in functional anatomy. It may be that "poor learners" do not succeed in developing a stable representation of the new non-native sound in their long term memory store, and that this results in a relatively greater involvement of "articulatory" strategies while subjects attempt to classify the sounds. Additionally, poor learners may engage subvocal rehearsal strategies in phonological verbal working memory relatively more than good learners. In other words, maybe they rely more on the strategy of comparing each sound to the previous one as they attempt to classify the sounds, and are less able than the good learners to match each sound to long-term memory "templates" for the native versus non-native sounds. The latter explanation is consistent with reports during training by some subjects that they did compare each sound to the previous one, and that their performance tended to deteriorate every time that they received negative feedback on an item.

The hypothesis of more efficient processing in frontal areas in learners is consistent with the results of several studies which have shown that prefrontal regions are relatively more active in tasks requiring "top-down" processing. In a study by Frith and collegues ('91), activity in the left IFG, at a location similar to our IFG area (location: -44, 8, 28), was found to be reciprocal with and was thought to modulate activity in the STG bilaterally. This area, in turn, was relatively more active in a task requiring more "bottom-up" processing. In a second study (Raichle et al, '94), a reciprocal pattern of activation was found between left prefrontal and sylvian/insular cortices before versus after the learning of a verbal generation task. Once again, it was proposed that the left prefrontal activation arises from top-down, effortful or naïve task performance, whereas that sylvian-insular activation arises from more bottom-up, practiced or automated task performance. Lastly, Chee and colleagues ('01) showed that in a group of bilinguals, there was greater activity bilaterally in the frontal operculum during a semantic task in the less proficient of two known languages. The authors suggesed that processing words in the less familiar language requires greater cognitive effort, that it is subserved by less well-tuned neural representations, and that it therefore requires greater neuronal activity than processing words in the more familiar language. We thus suggest a second interpretation for our results which is not exclusive with the first. It may be that "poor learners" continue to engage relatively more "top-down", effortful cognitive processes, resulting in stronger activation in regions including frontal speech areas. In contrast, maybe "good learners" engage relatively more "bottom-up" processes, more automatically attending to the acoustic information which critically distinguishes the two sounds, and thereby engage posterior speech regions to a greater extent than prefrontal ones.

The group subtraction analysis as well as the correlational analyses yielded significant peaks bilaterally in similar insula/frontal opercular areas. This suggests that activity in this area is both a) recruited in the classification task, and b) modulated by learning. In contrast, activations in the left inferior frontal gyrus (BA 44) and superior temporal regions were only detected in the group subtraction analysis, suggesting that these areas are recruited by performance of the identification task but that they are not modulated by learning. These findings also support the idea presented above that within the frontal speech area, the IFG and the insula/FO areas subserve different aspects of phonetic perception. Phonetic processing, which is only one aspect of speech processing, may involve different neural substrates depending on the task demands. This idea is supported by evidence for the involvement of functionally distinct subregions within the left inferior prefrontal region during different speech production tasks (Buckner et al, '95). More generally, it has been shown that activations attributed to semantic processing have been most often located anteriorly within the ventral inferior prefrontal cortex (BA 47/10), whereas activations attributed to phonological processing have been more frequently located posteriorly within the triangular and opercular portions of the inferior frontal gyrus (BA 44 and 45) (Fiez et al, '97). Consistent with this, we detected activity during a phonetic identification task in these latter frontal subregions, including parts of Brodmann's areas 44 and 45.

Finally, the finding of inverse modulation of insula/FO activity with learning highlights the importance of examining both increases as well as decreases in activity when attempting to elucidate the neural mechanisms of aspects of human cognitive functioning and of plasticity of such functions, in particular when examining processes as complex as human speech. Avenues for future work include combining functional imaging with electrophysiological measures to examine the timing of activity of speech-related brain regions and sub-regions, as well as to better understand the chronology of these events in order to gain knowledge regarding the temporal aspects of functional connectivity related to speech.

Summary and Conclusions

We were able to demonstrate overall behavioural improvement, despite large individual differences in performance, in the ability of monolingual English speakers to identify the Hindi dental-retroflex contrast after five hours of phonetic training. Functional imaging results during the classification of a native control phonetic contrast revealed the involvement of speech-related brain regions including the superior temporal and insula/frontal opercular regions bilaterally. Activity associated with the post-training classification of the newly learned phonetic contrast revealed a pattern of activation very similar to that found during the classification of the native phonetic contrast. This finding has important implications for neuro-linguistic theories of second language learning, and suggests that only five hours of phonetic training in adulthood with a non-native phonetic contrast, the perception of which is likely to have been lost as early as at six months of age, is sufficient to recruit the same substrates underlying the perception of speech sounds to which one has been exposed since birth.

In examining differences in brain function across subjects related to differences in the degree of success in learning the new speech sound, we found that in more successful learners, there was relatively more activity in the left angular gyrus, a language-related brain area, and relatively less activity in insula/frontal opercular regions as well as in the left middle temporal gyrus relative to non-learners. These results suggest that "good learners" recruit posterior temporoparietal speech regions relatively more than do "poor learners", and conversely, that frontal speech regions are relatively less active in "good learners" relative to in "poor learners".

This latter finding suggests that processing in these regions is more efficient and/or may require less processing time when successful phonetic learning has taken place. This in turn suggests that in successful learners, there is either less need for or more automatized processing related to subvocal rehearsal strategies in verbal working memory and/or related to mapping the perceived speech sounds onto an articulatory template. More generally, the concurrent modulation of activity in posterior and anterior speech regions as a function of learning may suggest that the functional connectivity between posterior and anterior speech regions is modulated by learning new speech sounds. Future studies can be designed to address this possibility.

Table 1: Native versus noise subtraction: Pre-training: Post-training:

Structure	*Talaraich Coordinates:*			*t-value*	***BA***	*Talaraich Coordinates:*			*t-value*
	X	*Y*	*Z*			*X*	*Y*	*Z*	
Left insula/ frontal operculum	-34	16	4	6.4*	**In/44**	-34	20	4	5.05*
Left inferior frontal gyrus	-48	14	24	4.50*	**44**	-42	26	20	4.51*
Right insula/frontal operculum	32	22	10	5.45*	**In/44**	30	20	-2	5.58*
Right middle frontal gyrus					**46**	38	40	14	4.87*
Left superior temporal gyrus	-66	-38	4	4.63*	**22**	-56	-22	-2	5.18*
Right superior temporal gyrus	46	-32	4	5.94*	**22**	54	-24	0	5.47*

Table 2: Non-native versus noise subtraction: Pre-training: Post-training:

Structure	*Talaraich Coordinates:*			*t-value*	***BA***	*Talaraich Coordinates:*			*t-value*
	X	*Y*	*Z*			*X*	*Y*	*Z*	
Left insula/ frontal operculum	-34	16	4	5.44*	**In/44**	-34	20	4	5.01*
Left inferior frontal gyrus					**44**	-44	16	0	4.54*
Right insula/frontal operculum					**In/44**	30	22	4	5.36*
Right middle frontal gyrus					**46**	40	44	8	5.67*
Left superior temporal gyrus					**22**	-58	-22	2	4.6*
Right superior temporal gyrus					**22**	52	-24	-1	3.68
Left caudate nucleus						-12	-2	16	5.14*

Table 3: Compound comparison: Post-training non-native versus noise comparison minus pre-training non-native versus noise comparison

Structure	*Talaraich Coordinates:*			*t-value*	***BA***
	X	*Y*	*Z*		
Left inferior frontal gyrus	-44	26	20	4.32*	**45**
Left precentral gyrus	-58	4	18	3.96	**6**
Right middle frontal gyrus	44	38	8	4.04*	**46**
Left superior parietal gyrus	-26	-74	38	4.19	**7**
Left supramarginal gyrus	-44	-36	36	3.71	**40**
Left caudate nucleus	-12	4	14	4.69*	
Right globus palladus	24	-4	-6	4.84*	

Table 4: Correlation between learning measure (post-A' minus pre-A') and post-training non-native minus noise subtraction

Structure	*Talaraich Coordinates:*			*t-value*	***BA***
	X	*Y*	*Z*		
Positive correlations:					
Left angular gyrus	-54	-66	26	5.47*	**39**
Right angular gyrus	44	-70	34	5.11*	**39**
Negative correlations:					
Left insula/ frontal operculum	-40	10	4	-3.8	**In/44**
Right insula/frontal operculum	30	20	8	-3.76	**In/44**
Left poster. middle temporal gyrus	-52	-50	8	-5.13*	**21**

* indicates $p<0.05$

Figure captions

Figure 3: Pre- and post-training native minus noise subtractions.

Figure 4: Pre- and post-training non-native minus noise subtractions.

Figure 5: Positive correlations between learning measure (L-score) and BOLD signal during post-training non-native minus noise comparison.

Figure 6: Negative correlations between learning measure (L-score) and BOLD signal during post-training non-native minus noise comparison.

GROUP SUBTRACTIONS:
NATIVE versus NOISE

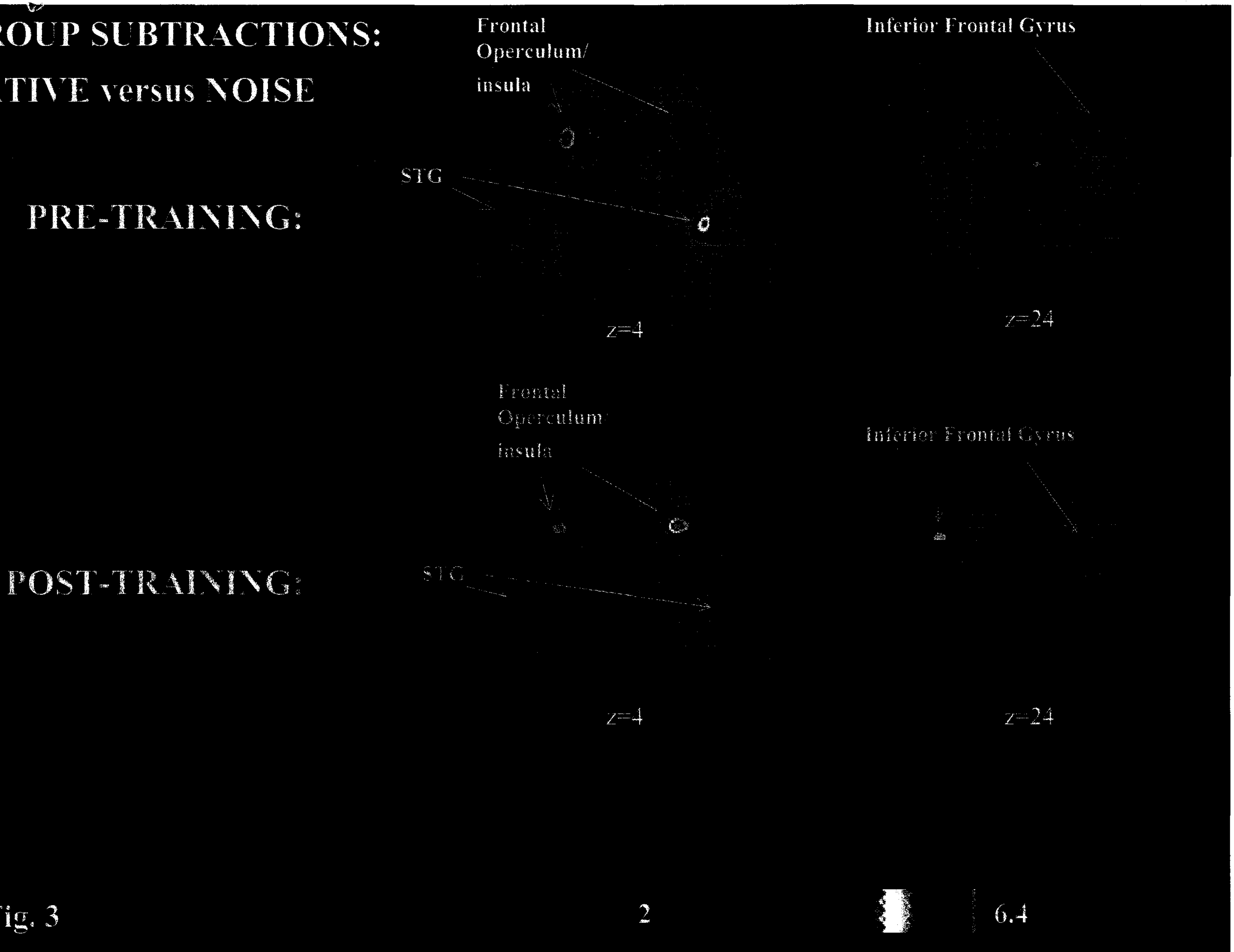


Fig. 3

GROUP SUBTRACTIONS:
NON-NATIVE versus NOISE

PRE-TRAINING:

POST-TRAINING:

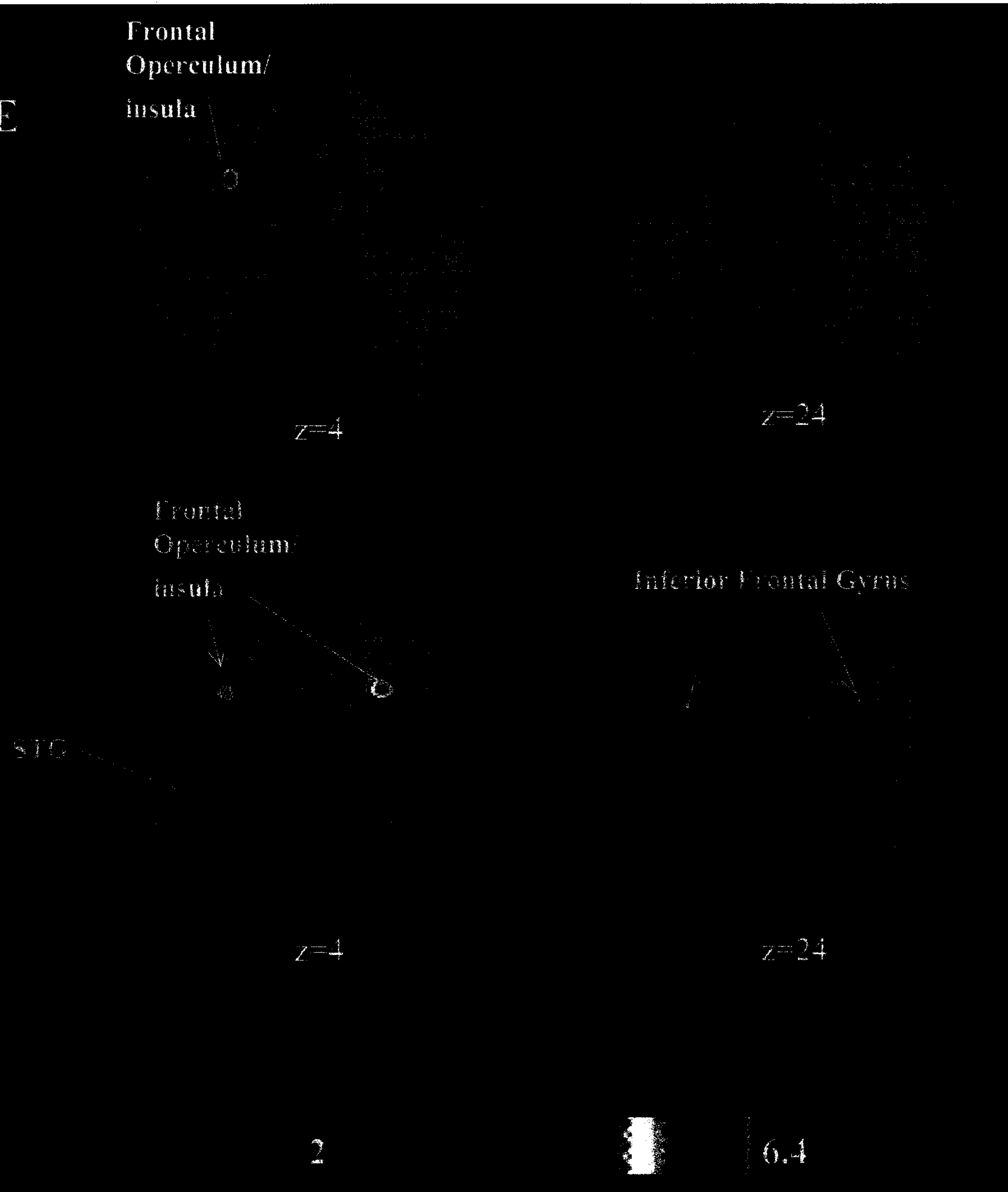


Fig. 4

Correlation between learning measure and BOLD signal during Post-training Non-native minus Noise comparison

Positive Correlation:

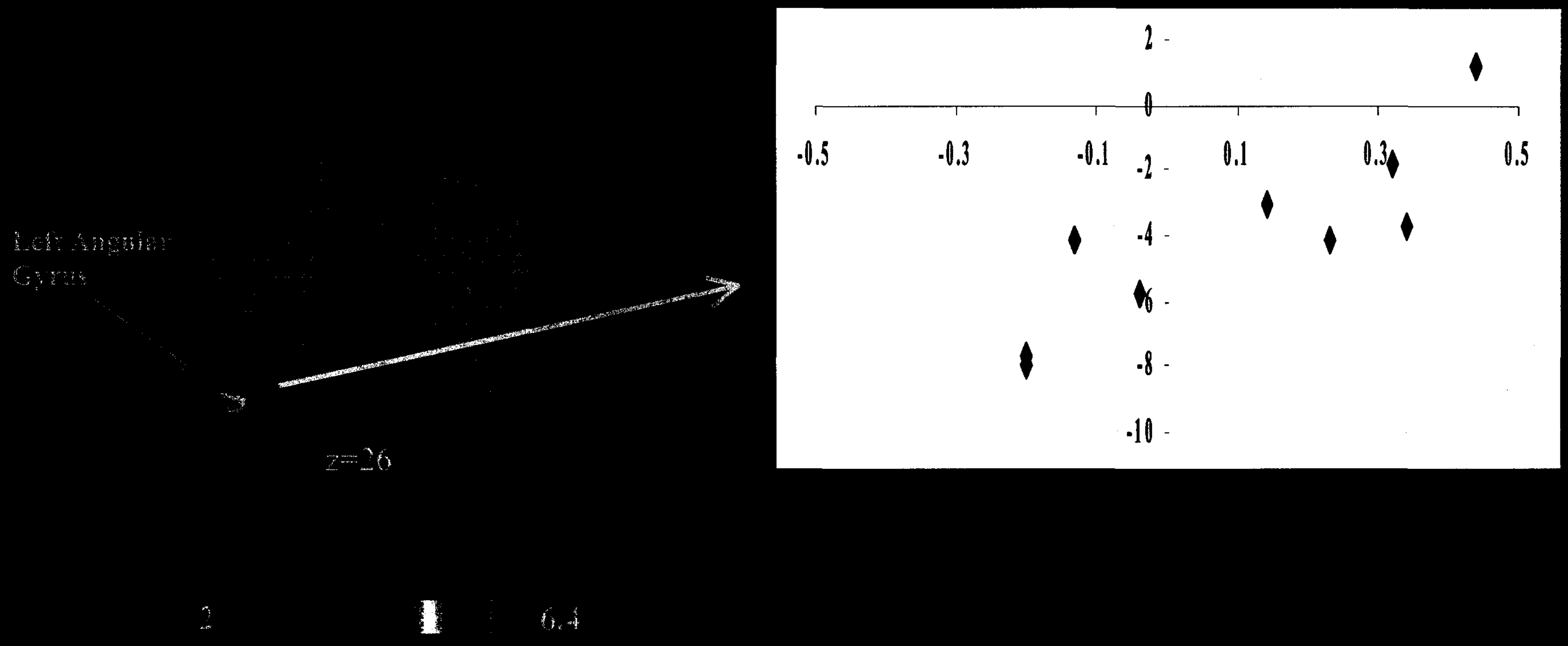
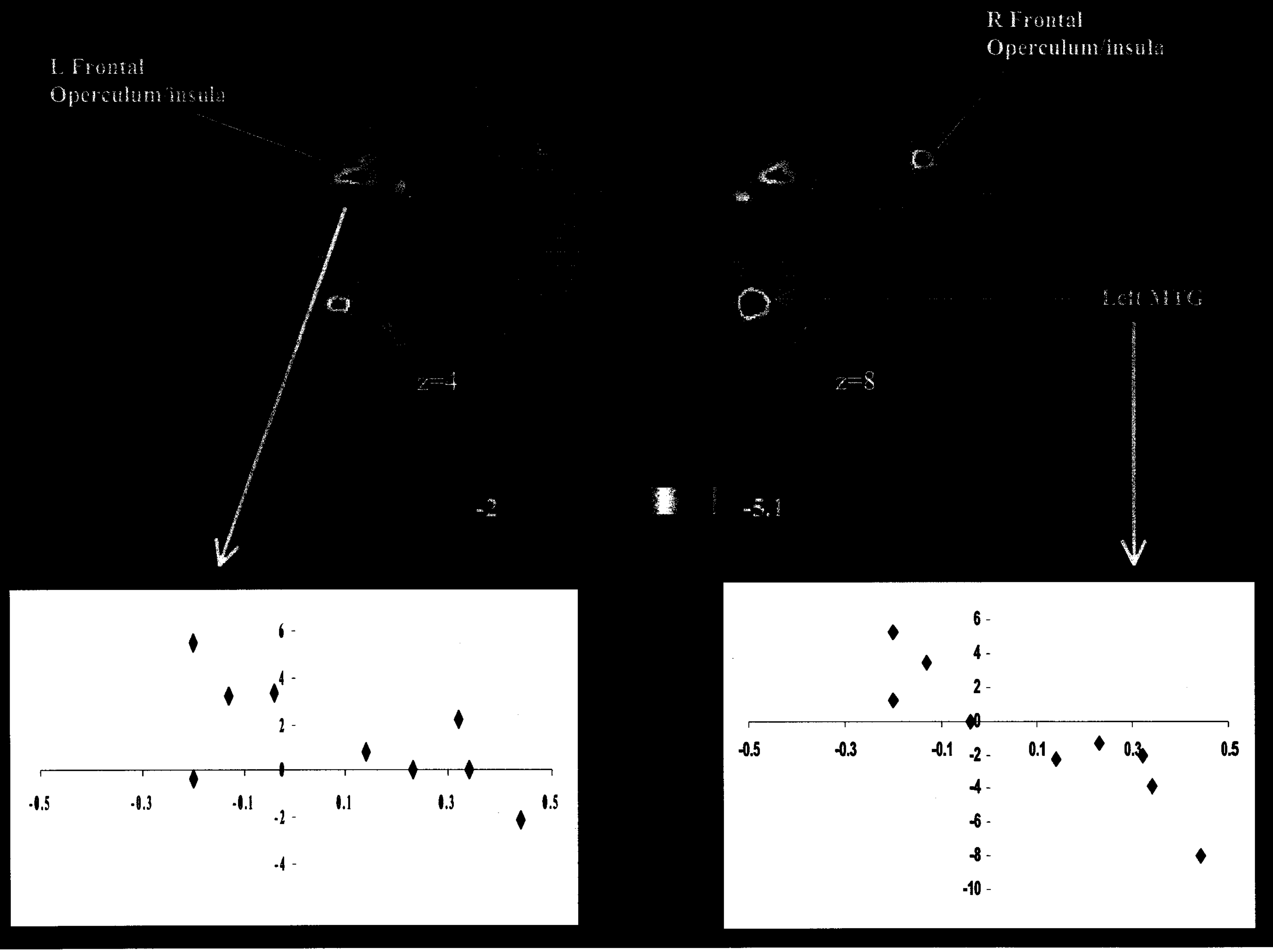


Fig. 5

Negative Correlations:

Fig. 6



Chapter 4

Study 3: *Anatomical correlates of the ability to learn novel speech sounds*

With the development of high resolution three-dimensional anatomical MRI (aMRI), there has been a growing interest in studying normative inter-individual variations in the size and shape of various brain structures (eg; Paus et al, '96, Giedd et al, '99, Penhune et al, '96, Westbury et al, '99), and in relating such variations to differences in aspects of behavior across individuals (Schlaug et al, '95, Steinmetz et al, '91, '95, Foundas et al, '95, Habib et al, '95). The aim of the present investigation was to determine whether individual differences in the ability to learn speech sounds coming from a non-native language may be in part accounted for by individual differences in brain morphology.

During development and starting as early as at six months of age, lack of experience with certain non-native speech sounds results in a developmental shift from a language-general to a language-specific pattern of phonetic perception (Best et al, '88, Polka & Werker, '94, Werker & Tees, '84a, Werker & Lalonde, '88, Kuhl et al, '92). During adulthood, most individuals cannot distinguish certain non-native speech sounds from similar ones belonging to the native language. The results of laboratory studies aimed at improving non-native phonetic perception have shown that some contrasts are more easily learned than others. Interestingly, even amongst adults with very similar language backgrounds, considerable individual differences exist in the ability to learn to perceive "difficult" non-native speech sounds following phonetic training (Polka, '91, Pruitt et al, '90, Jenkins et al, '95). We addressed the question of whether such individual differences might be predicted by differences in brain morphology. To achieve this, we used voxel-based morphometry instead of manually segmenting particular brain structures. This method allows an exploratory search of the whole brain for relationships between behavioral measures and brain morphology. It does not require subjective identification of tissue boundaries, and hence, the use of arbitrary or conventional definitions of particular brain structures. Futhermore, it allows examination of white-matter morphology, which is otherwise difficult to do due to the lack of clear tissue boundaries between adjacent white matter subregions (except for the corpus callosum). Voxel-based morphometry has been used by other investigators to demonstrate normative brain asymmetries (Watkins et al, '01), maturation of white matter tracts (Paus et al, '99), structural correlates of arithmetic calculation abilities (Isaacs et al, '01), and differences in brain morphology in normal versus clinical groups (Vargha-Khadem et al, '98, Wright et al, '95).

In the present study, we trained individuals to distinguish the non-native dental-retroflex phonetic contrast. Previous functional imaging work on phonetic perception has shown the involvement of several temporoparietal regions of the left hemisphere (Démonet et al, '94, Zatorre et al, '92, '96, Binder et al, '96, '97, Petersen et al, '88, Paulesu et al, '93), the superior temporal gyri bilaterally (Binder et al, '94, Mazoyer et al, '93, Jäncke et al, '98, Mummery et al, '99), and left inferior frontal regions in and around Broca's area (Zatorre et al, '92 & '96, Paulesu et al, '93, Fiez et al, '95, Burton et al, '00). We characterized subjects' performance using the following three measures: pre-training discrimination performance, learning rate, and amount of improvement in identification performance after relative to before training. We correlated two of these measures (identification performance and learning rate) with brain morphology throughout the whole brain volume. We predicted that the ability to learn the phonological contrast would be correlated with differences in brain morphology in language-related cortical areas, including left parieto-temporal regions as well as left frontal opercular and inferior frontal regions. In addition, we wanted to test the speech-specificity of any potential anatomical correlates of phonetic learning. Therefore, we also trained subjects to perceive two other sets of non-linguistic stimuli, consisting of rapidly changing sounds and steady-state tonal sounds. For these two stimulus types, we only performed analyses that yielded significant results with the speech sounds, as the aim was to test the specificity of the effects.

METHODS

Subjects

Our subjects included 59 healthy right-handed adult volunteers, 33 male and 26 female. Ages ranged from 18 to 39 years (mean age 22), screened for experience with any retroflex sounds. All subjects spoke English, and a subset also spoke one or more additional languages that do not employ the retroflex sound. For each subject, hearing was assessed with an audiometer before testing began, and was found to be normal.

Stimuli

Selection of the Non-native Linguistic Stimulus: We selected the dental-retroflex place-of-articulation contrast that is used in languages of India such as Hindi or Urdu. Retroflex consonants require a relatively complex articulation and are rare across languages; only 11% of the world's languages include a retroflex consonant, and they are unlikely to be pronounced in languages in which the contrast is phonetically irrelevant (Burnham, '86). The non-native perception of the retroflex sound is lost below the age of one year (Werker & Tees, '83, Werker & Tees, '84a, Tees & Werker, '84). Perceptually, English listeners assimilate the dental-retroflex sounds such that they perceive both sounds as instances of the dental consonant (Polka, '91, Werker & Lalonde, '88).

Research aimed at training the non-native perception of the dental and retroflex sounds in adults has shown that standard training approaches fail to improve performance (Werker & Tees, '83, Werker and Tees, '84a, Tees & Werker, '84). Improvement has been reported under certain circumstances, including shorter inter-stimulus intervals (Werker and Tees, '84b, Werker and Logan, '85), truncated stimuli (Pruitt et al, '90), and having had extensive experience with the contrast (Tees and Werker, '84).

We selected this non-native retroflex consonant, which is rare and unlikely to occur allophonically across languages, because we wanted to ensure that none of our participants have had either allophonic or non-native exposure to this sound. Additionally, we wanted to ensure that no individual would be able to distinguish this sound from the dental consonant before training, and that we would find a considerable range in performance across subjects so that we could then correlate such individual differences with brain morphology. In a previous study (Golestani et al, unpublished), we synthesized the retroflex consonant, and showed that phonetic training resulted in overall improvement in the ability to identify this sound versus the native dental one in a large group of subjects. A considerable range in performance was observed across subjects.

Synthesis of Stimuli:

Non-native Hindi contrast: Synthesis of the non-native phonetic contrast was based on parameters reported by Stevens and Blumstein ('75) for the retroflex-dental place-of-articulation continuum. Refer to Figure 1 for schematized frequency-time representations of the dental and retroflex consonant-vowel (CV) syllables. There were seven stimuli varying in equal steps in terms of acoustic difference between adjacent items. The set of four-formant stimuli was constructed with the use of the MITSYN synthesizer (Henke, '90). Stimulus 1 corresponds to the dental voiced, unaspirated stop consonant prototype, and stimulus 7 to the retroflex prototype. All of the consonants were followed by the vowel /a/.

Stimuli began with an initial noise burst. The parameters that were manipulated to create the continuum are the frequency glides of the third formant (F3), as well as the center frequency of the burst. The initial noise burst lasted 5ms, and its center 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 40ms. The fundamental frequency (F0) started at 130 Hz and decreased linearly over the first 40 ms to 125 Hz, 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 Hz to 655 Hz during the transition, and then remained steady-state during the remainder of the sound. The second formant (F2) decreased from 1650 Hz 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.

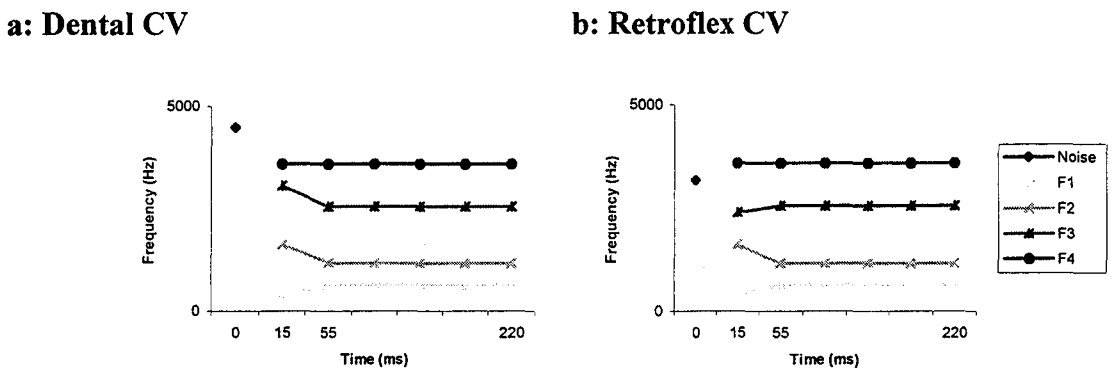


FIG. 1: Schematized frequency versus time diagrams of the (a) dental and (b) retroflex consonant-vowel (CV) syllables.

Rapidly changing non-linguistic stimuli:

Non-linguistic stimuli were formant-based, and therefore had a voice-like quality, yet did not, to our knowledge, represent phonemes that exist in any existing language. The acoustics of the nonlinguistic stimuli were analogous to those of the speech syllables in that for both of these 7-step continua, the frequency glides of the third frequency band was manipulated. The non-linguistic stimuli were different from the speech ones in that the frequencies of the "formants" were arbitrarily selected, and in that there was no noise burst preceding the sounds. These sounds were created in a manner similar to those used by Belin and colleagues ('98). Our stimuli were different than theirs in two respects: first, the direction and slopes of our glides were different, and second, their stimuli had frequency glides both at the beginning and at the end of the sounds, whereas ours only had glides at the beginning.

The acoustics of the non-linguistic stimuli are the following. All of the stimuli lasted 220 Hz. The fundamental frequency (Fo) was steady-state at 210 Hz. The first formant (F1) rose from 337.5 to 450 Hz during the first 40 ms, and then remained steady-state. The second formant (F2) dropped from 2000 to 1500 Hz during the first 40 ms, and then remained steady-state. The forth formant (F4) remained steady-state at 3300 Hz throughout the entire sound. Finally, the starting frequency of the third formant (F3) decreased in equal 140 Hz steps from 3020 (stimulus 1) to 2180 (stimulus 7), and its steady-state frequency was 2400 Hz.

Steady-state tonal stimuli:

The steady-state stimuli consisted of pure tones with durations of 220 ms, and had onset and offset times of five and twenty milliseconds, 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 2007 Hz (stimulus 1) to 2042 Hz (stimulus 7).

Procedure

All subjects were scanned several months prior to behavioral testing and training. Linguistic and musical experience were evaluated with questionnaires before testing began. Identical testing and training procedures were followed for each of the three stimulus types, the order of which was counterbalanced across subjects. For each stimulus type, subjects were first familiarized with the sounds; they were randomly presented with 20 instances of each of the two endpoint stimuli. As each stimulus was presented, subjects saw its corresponding label on the computer screen, “A” for the dental sound, and “B” for the retroflex one. After familiarization, subjects were 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 discrimination and identification testing (described below) was then conducted. Following this, subjects were trained, using the adaptive training procedure described below. Discrimination and identification tests were re-administered after training.

Identification and discrimination testing:

During the identification task, subjects were encouraged to use the descriptions of A and B that they had provided during familiarization. They 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 the “dental” sound, and the other to the “retroflex” one. Each of the 7 stimuli was presented 10 times in random sequence, for a total of 70 trials. They were then tested using a 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 7 "same" pairs, five two-step "different" pairs, and one six-step different pair. Discrimination and identification tests were again administered at the end of training.

Training:

During the training phase of the experiment, we progressively reduced the acoustic difference between the dental and retroflex sounds as a function of successful performance by subjects. This "perceptual fading" training method adapts to subjects' performance (Jamieson & Morosan, '86, Morosan & Jamieson, '86, Terrace, '63). It allows them, if they do well using the endpoint stimuli, to try to distinguish dental and retroflex sounds which are a smaller step-size apart. This training paradigm is likely to help individuals attend to the relevant acoustic differences between the dental and retroflex sounds. Training involved 20-trial identification training blocks. Subjects were presented with either sound 1 or sound 7, and were asked to press one of two mouse buttons corresponding to the presented sound. They received feedback on the computer as to whether their response was correct or incorrect. If and when an individual achieved criterion performance, defined as 16/20 correct responses on a given block, we administered stimuli 2 versus 6 during the next training block. Again, if and when criterion was reached with this slightly more difficult pair, we used stimuli 3 versus 5 for the next training block. Training was discontinued either once a subject achieved criterion on this last contrast, or once they had completed a maximum of 200 trials (10 blocks).

Psychophysical analyses

Laboratory training studies using synthetic phonemes typically involve creating a continuum of sounds, such that the endpoint sounds are reliably perceived exemplars of the two contrasting phonemes in question, and that there is an equal, controlled physical difference between each of the adjacent sounds in the series. Before training, the endpoint sounds are difficult to identify, and pairs of sounds from the series are difficult to discriminate. Identification is tested following training by presenting the sounds of the series one at a time, and successful training is reflected by the ability to reliably label one endpoint as one of the phonemes, and the other endpoint as the other phoneme. A steeper identification function therefore indicates that individuals are better at reliably classifying the sounds at one end of the continuum as belonging to the "dental" category, and the sounds at the other end as belonging to the "retroflex" category. Successful discrimination is reflected by higher scores across the phoneme boundary as well as by lower scores during the discrimination of sounds belonging to the same phonetic category. For example, if the perceptual boundary is between stimuli 3 and 5 for a particular subject, they should be able to say that sounds 3 and 5 are "different", whereas they should find it difficult to discriminate sounds 1 and 3.

Each subject's discrimination performance was converted to an A' score, an index of sensitivity which corrects for individual differences in bias. A' is a non-parametric analogue 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, '72). The slopes of individual identification curves were calculated using a standard regression approach. Once again, a steeper slope reflects better identification performance. We developed a continuous learning rate variable ("L-scores") to quantify subjects' performance over the course of learning. This variable takes into account both the number of blocks at each level of difficulty and the accuracy of performance during 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 hardest blocks, respectively. Resulting values were then subtracted from 600, which was the maximum possible value, yielding an "L-score". This learning can range from zero (slowest possible learner) to 600 (fastest possible learner).

For each of the three stimulus types, we measured: a) pre-training discrimination performance (A'), reflecting the ability to distinguish the non-native sound from the native one before any training has taken place, b) post-minus-pre-training slopes, reflecting the amount of improvement in the ability to classify the two sounds resulting from training, and c) L-scores, reflecting rate of learning. We correlated the latter two of these measures for the Hindi stimuli with brain morphology using a voxel-wise analysis.

This analysis was followed by regional morphometric analyses in brain areas detected by the voxel-based analysis.

Scanning Protocol & Morphometric Analyses

We used a voxel-based statistical analysis (Wright et al, '95, Vargha-Khadem et al, '98, Paus et al, '99, Ashburner & Friston, 2000, Watkins et al, 2001, Isaacs et al, '01), in order to determine whether there is a systematic relationship between behavioral measures and regional amounts of gray matter or white matter. The following three MR sequences were used: T1-weighted (3D fast field echo scan with 140-160 slices, 1-mm isotropic resolution, TR = 18ms, TE = 10ms, flip angle = 30°), T2-weighted (2D multi-slice fast spin echo scan with 140-160 2-mm slices with a 1mm overlap, TR = 3300ms, TE = 35ms), and proton density (as for T2 scan, but with TE = 120ms). Each of these image volumes underwent a non-uniformity correction (Sled et al, '98) to remove variations in signal intensity related to radio-frequency inhomogeneity. Each T1-weighted image volume was linearly transformed with an automatic image-registration method (Collins et al, '94) based on multi-scale 3D cross-correlation with an average (n=305) MR image aligned with the Talairach stereotaxic space (Talairach and Tournoux 1988). The transformed images were then classified into gray matter, white matter, and cerebrospinal fluid using INSECT (Zijdenbos et al, '98), an automatic tissue-classification paradigm. The binary tissue-classified gray matter (GM) and white matter (WM) maps were then smoothed using a Gaussian smoothing kernel of 10-mm full-width at half-maximum, resulting in three-dimensional (3D) maps of GM and WM "density". Smoothing serves the purpose of converting binary into continuous data, which is necessary for correlating the signal with a continuous behavioral measure. It also serves to weigh the signal at each voxel according to the signal in neighboring voxels, thereby reflecting the amount of gray or of white matter within the smoothing kernel (ie, GM or WM regional "density"). The GM and WM maps were then correlated with each of three measures of performance on a voxel-by-voxel basis using a liner regression model, yielding three gray matter and three white matter maps. The three behavioral measures were: pre-training A' scores, post- minus pre-training identification slopes, and L-scores. Thresholds for the t-statistic of >5.3 for gray matter and of >5.4 for white matter were calculated (Worsley *et al.,* 1996) based on 57 degrees of freedom, a voxel size of one cubic millimeter, smoothness of 10-mm, a significance of $p<0.05$, and volumes of interest of 540 cubic centimeters for gray matter and of 810 cubic centimeters for white matter.

In order to facilitate the interpretation of the voxel-based analyses, we also performed regional morphometric analyses using DISPLAY (MacDonald et al, '94). This manual segmentation software allows simultaneous real-time viewing and voxel-labeling on sagittal, coronal, and horizontal planes. We also correlated measures of performance with existing midsagittal corpus callosum (CC) area measurements (Bermudez & Zatorre, '01), as well as with the volumes of the gray and white matter in the left and right parietal lobes. The CC areas were acquired by manual segmentation using DISPLAY. The parietal volumes were measured using an approach to automatic segmentation which relies on the ANIMAL and INSECT programs. This technique involves combining automatic tissue classification (see above) with non-linear registration to a template brain for which brain regions had been manually defined (Collins et al, '95, '97, '99).

RESULTS

Behavioral Results

Discrimination:

A 2x3x5 three-way repeated measures analysis of variance (ANOVA) was performed on the A' discrimination scores, the within subjects factors being time (pre- vs post-training), stimulus type (Hindi, nonlinguistic, and tonal), and stimulus pair (1/3, 2/4, 3/ 5, 4/6, 5/7). Results showed significant main effects of time ($F(1, 58)= 8.99, p<0.005$), and of pair ($F(4, 232)=3.50, p<0.01$, Greenhouse-Geisser correction for all ANOVA results). The main effect of time reflects significantly better performance after relative to before training. In order to further analyze the effect of training, we performed tests of simple main effects on the effect of training for each stimulus type. We found that there was a significant improvement due to training in the ability to discriminate the rapidly changing non-linguistic stimuli ($F(1,58)= 11.02, p<0.01$), but that there was a non-significant improvement for the tonal ($F(1,58)= 1.85, p>0.05$) and for the Hindi stimuli

(F(1,58)= 0.03, p>0.05). There was a large range in performance both before and after training across subjects.

Correlational analyses revealed that for each stimulus type, pre- versus post-training performance was significantly correlated across subjects (Hindi: r=0.49, p<0.001, nonlinguistic: r=0.65, p<0.001, and tonal: r=0.56, p<0.001). Pre-training discrimination sensitivity was not related across stimulus types, whereas after training, only the A' values for the Hindi versus the non-linguistic stimuli were significantly correlated (r=0.35, p<0.05). The correlation between post-training performance on the Hindi and tonal stimuli was 0.25 (p>0.05). A test of part correlations revealed that there is a significant amount of unique variation associated between performance on the Hindi and non-linguistic stimuli while adjusting for performance on the tonal stimuli (t= 2.49, p<0.05).

Identification:

Figure 2 presents the mean percentage of "A" responses for each of the seven stimuli during pre- and post-training identification of the three stimulus types. As described above, for the Hindi stimuli, responses "A" and "B" correspond to the dental and retroflex sounds, respectively.

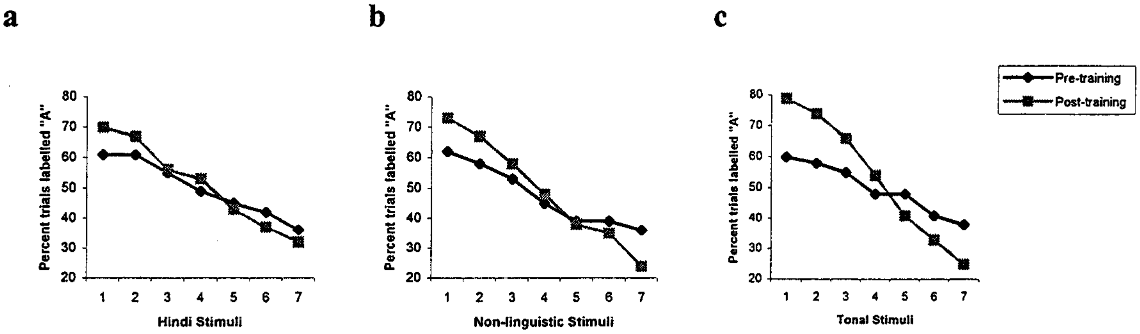


FIG.2. Pre-training and post-training identification functions for (a) Hindi stimuli, (b) non-linguistic stimuli, and (c) tonal stimuli.

The slope of individual subjects' identification curves and of identification curves averaged across subjects was calculated using a standard regression approach. Relatively steeper negative slopes reflect a relatively better ability to reliably identify the sounds in each stimulus set. A 2 X 3 two-way repeated measures ANOVA was performed on the identification slopes, with time (pre- vs post-training), and stimulus type (Hindi, nonlinguistic, and tonal) being the within subjects variables. Results revealed a main effect of time ($F(1, 58)=62.49$, GG-corrected, $p<0.001$), indicating that the post-training identification functions (slope, collapsed across stimulus type, was equal to -0.82) were steeper than the pre-training ones (mean slope= -0.43), reflecting improved performance after training. The effect of stimulus type was not significant. In addition, a significant interaction was found between time and stimulus type ($F(2, 116)(GG)=3.17$, $p<0.05$). We performed tests of simple main effects on the effect of training for each stimulus type in order to ensure that there was improvement in performance for all three stimulus types. Results revealed that performance was significantly better after relative to before training for all three stimulus types (Hindi: $F(1,58)=8.47$, $p<0.01$, non-linguistic: $F(1,58)=16.56$, $p<0.01$, and tonal: $F(1,58)=26.94$, $p<0.01$). Tests of simple main effects on the effect of stimulus type at pre- and at post-training revealed that identification performance was similar across the three stimulus types before training ($F(2,116)=0.147$, $p>0.05$). After training, however, there was a significant difference in identification performance across the three stimulus types ($F(2,116)=3.23$, $p<0.05$). Newman-Keuls post-hoc tests revealed that identification performance was significantly better on the tonal relative to the Hindi stimuli ($Q=3.59$, $Qc(3,116)=3.36$, $p<0.05$).

Correlational analyses revealed that for each stimulus type, pre- versus post-training slopes were significantly correlated across subjects (Hindi: $r=0.50$, $p<0.001$, nonlinguistic: $r=0.59$, $p<0.001$, and tonal: $r=0.46$, $p<0.001$). During pre-training, there were no significant correlations in performance across the stimulus types, consistent with discrimination results reported above. At post-training, there were significant correlations between identification slopes on the Hindi and the nonlinguistic stimuli ($r=0.36$, $p<0.05$), as well as between the Hindi and the tonal stimuli ($r=0.30$, $p<0.05$). A test of part correlations revealed that there is a significant amount of unique variation associated between performance on the Hindi and non-linguistic stimuli while adjusting for variation associated with the tonal stimuli ($t=2.80$, $p<0.01$).

Training:

As described above, L-scores were calculated for each subject, providing a learning rate measure. Scores can range from zero to six-hundred, the higher scores indicating faster learning. Figure 3 illustrates the large range in performance across subjects for the Hindi stimuli.

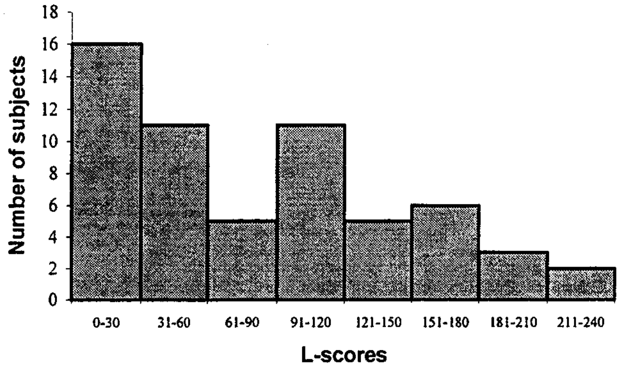


FIG. 3: Frequency distribution of L-scores (learning rate) across 59 subjects.

Imaging Results

Given that there was behavioral improvement for identification but not for discrimination performance, we correlated post-minus pre-training identification slopes as well as L-scores, but not discrimination performance, with brain morphology. Refer to Table 2 for list of morphological analyses and main results. There was a negative correlation between L-scores and the gray matter density in a region just anterior to the parieto-occipital sulcus bilaterally (Fig.4a-c, Table 1a), suggesting that there is less gray matter and/or more white matter in these regions in faster relative to slower learners. As expected, the correlation with the white matter maps yielded similar results with t-values in the opposite direction (Fig.4c-d, Table 1b). The t-value in the left hemisphere was marginally significant, however, given that the shape of the signal is symmetrical and is at similar, homologous locations (two millimeters apart) in the two hemispheres, it is likely that this result represents a bilateral morphological correlation. The results did not change when a measure of linguistic experience was covaried out. Even when lowering the threshold to >4.5, we did not find an association between measures of phonetic learning and gray or white matter density in primary speech regions of the brain.

Correlations with post- minus pre-training identification performance for the Hindi stimuli did not yield any significant results. Similar analyses using L-scores for learning non-linguistic rapidly changing and tonal stimuli did not yield any significant findings.

The finding of a correlation between the Hindi L-scores and gray matter density anterior to the parieto-occipital sulcus may be due to differences across subjects either in the location or in the gray matter volume of this sulcus. In order to distinguish between these two possibilities, we performed regional morphometric analyses of the parieto-occipital sulcus of twenty individuals selected in the following manner. We chose ten of the fastest learners who also had the lowest gray matter density values, and ten of the slowest learners who also had the highest white matter density values at the following parieto-occipital location in the left hemisphere: $x=-14$, $y=-66$, $z=39$, $t= 5.38$, $r= 0.57$ (Fig. 5). These two groups constituted what we will, from now on, refer to as the "learner" and the "non-learner" groups.

We manually labeled the parieto-occipital sulcus in T1-weighted MR images of these twenty individuals (Fig. 6). We averaged these labels separately for the learners and the non-learners, thus creating probability maps for this sulcus (Fig.7). It can be seen from these images that the sulcus appears to be located more posteriorly in the fast learners as compared with the slow learners. In order to test this apparent location difference, we performed a mixed, two-factor analysis of variance on the Y-axis location of the center of gravity of the sulcus. The between-subjects variable was group (learners vs non-learners), and the within-subjects variable was hemisphere (right vs left). We found a main effect of group ($F(1,18)=19.88$, $p<0.001$), confirming that the sulcus is located more posteriorly in the learners relative to the non-learners, as well as main effect of hemisphere ($F(1,18)=30.87$, $p<0.001$), indicating that the sulcus is located more posteriorly in the left relative to the right hemisphere. There was also a significant group by hemisphere interaction ($F(1,18)=5.10$, $p<0.05$) (Fig.8a). Newman-Keuls post-hoc tests confirmed the two main effects, and showed that the sulcus is located more posteriorly in the left relative to the right hemisphere for both learners ($Q=7.81$, $Qc(2,18)=2.97$, $p<0.05$) and non-learners ($Q=4.69$, $Qc(3,18)=3.61$, $p<0.05$), and that in both the left ($Q=9.03$, $Qc(3,18)=3.61$, $p<0.05$) and the right ($Q=4.69$, $Qc(3,18)=3.61$, $p<0.05$) hemispheres, the sulcus is located more posteriorly in learners relative to in nonlearners. Results indicate that the significant group by hemisphere interaction (see Fig.8a) arises from a larger difference in the location of the sulcus for the learners relative to the nonlearners. In other words, there is a greater asymmetry (left more posterior than right) in the position of the sulcus in learners. We also performed a mixed, two-factor analysis of variance on the volume of the parieto-occipital sulcus. These volumes were obtained by summing across all labeled voxels for each subject. We found no difference in the volume of the sulcus between groups ($F(1,18)=0$, $p>0.05$) nor between hemispheres ($F(1,18)=0.43$, $p>0.05$).

The difference in the location of the parieto-occipital sulcus across subjects may also arise from differences in the volume of the parietal lobes, which lie anterior and adjacent to the sulcus. It is possible that larger parietal lobes in fast learners result in a posterior "positional shifting" of the sulcus relative to slow learners. In order to test this possibility, we performed a mixed factor analysis of variance on parietal lobe volumes. The between-subjects variable was group (learners versus non-learners), and the within-subject variables were hemisphere (right versus left) and tissue-type (gray versus white matter). We found a main effect of hemisphere ($F(1, 18)=63.39$, $p<0.001$), reflecting larger volumes in the left relative to the right hemisphere. There was a significant tissue by group interaction ($F(1,18)=5.44$, $p<0.05$), as well as a significant tissue by hemisphere interaction ($F(1,18)=8.73$, $p<0.01$). There was no group by hemisphere interaction ($F(1,18)=0.082$, $p>0.05$). Lastly, there was a significant three-way interaction between group, hemisphere, and tissue. Tests of simple interactions were performed for the two groups separately. Results revealed that for the learners, there was a two-way hemisphere by tissue interaction ($F(1,9)=18.00$, $p<0.005$). Newman-Keuls post-hoc tests showed that in both the left ($Q=16.96$, $Q_c(3,9)=3.95$, $p<0.05$) and right ($Q=8.48$, $Q_c(3,9)=3.95$, $p<0.05$) hemispheres, there is more white relative to gray matter. The hemisphere by tissue-type interaction for the learners arises from a greater difference in the amount of white relative to gray matter (WM>GM) in the left relative to the right hemisphere (Fig.8b). Tests of simple interactions revealed that for non-learners, the hemisphere by tissue interaction was not significant.

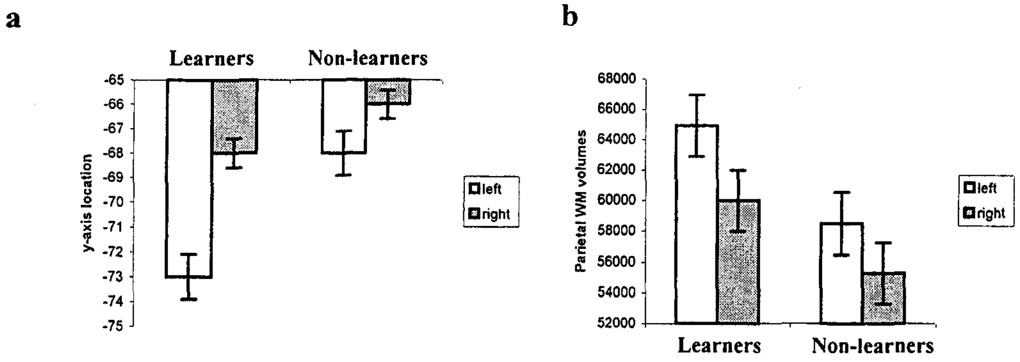


FIG. 8: Group by hemisphere interactions for a) y-axis location of parieto-occipital sulcus, and b) white matter parietal lobe volumes.

Correlational analyses between the gray and white matter volumes of the left and right parietal regions with the rate of learning across the twenty subjects confirmed the above pattern of results. They revealed that in the left hemisphere, there is less gray matter ($r= -0.49$, $p<0.05$) and more white matter ($r= 0.50$, $p<0.05$) in the learners relative to in the non-learners (refer to Figure 3). There was a similar, marginally significant pattern of results for gray ($r= -0.43$, $p=0.06$) and for white ($r=0.41$, $p=0.07$) matter volumes in the right parietal region. The same analysis across all fifty-nine subjects revealed that there is significantly less gray matter in the left ($r= -0.31$, $p<0.05$) and in the right hemisphere ($r= -0.25$, $p<0.05$) in the learners relative to in the non-learners.

Correlations between the y-axis location of the parieto-occipital sulcus and the volumes of the left and right parietal lobes across the twenty learner and nonlearner subjects did not reveal the presence of a significant relationship in the left ($r= -0.21$, $p>0.05$) nor in the right hemisphere ($r= 0.10$, $p>0.05$).

In order to determine whether a similar pattern of results exists for the non-linguistic and for the tonal stimuli, we correlated gray and white matter volumes of the left and right parietal lobe with the rate of learning for these other two stimulus types across all fifty-nine subjects. We found that there is more white matter in both the left ($r=0.25$, $p<0.05$) and in the right ($r=0.27$, $p<0.05$) hemisphere in individuals who learn the non-linguistic stimuli more rapidly relative to those who learn these sounds more slowly. A test of part correlations revealed that in the right hemisphere, there is a significant amount of unique variation associated between performance on the non-linguistic stimuli and white matter volume, even when adjusting for variation in performance associated with the Hindi stimuli (t=2.00, p<0.05). In the left hemisphere, white matter volumes can be predicted based on performance on the non-linguistic stimuli even when adjusting for shared variation in performance on the Hindi stimuli with a p-value of 0.07 (t=1.87, p=0.07). There was no association between parietal lobe volumes and performance on the tonal stimulus type.

Anterior to the parieto-occipital sulcus are white matter tracts connecting several cortical regions both inter- and intra-hemispherically (Talairach & Tournoux, '93). The fibers connecting the temporal and parietal regions of the two hemispheres cross the caudal part of the corpus callosum (Waddington, '84, de Lacoste et al, '85). It is possible that in fast learners, the more posterior position of the parieto-occipital sulcus is related to a greater number of interhemispheric fibers adjacent to the sulcus. Visual inspection of the statistical parametric maps from the voxel-based analyses (Fig. 4) reveal that significant bilateral differences in morphology in the parieto-occipital region appear to lead inferiorly and medially to the corpus callosum, and extend into the opposite hemisphere. In order to test the possibility that the parieto-occipital finding is related to a difference across the groups in interhemispheric connectivity, we compared midsagittal area measurements of various subdivisions of the corpus callosum (CC) for the learner versus non-learner groups. Corpus callosum subdivisions included: total CC, anterior, middle, and posterior third of the CC, and the splenium of the CC (Bermudez & Zatorre, '01). None of the comparisons were significant; however, there was a non-significant trend (t(18)=1.43, p=0.09, one-tailed) indicating that learners tended to have a larger midsagittal area in the middle third of the CC relative to nonlearners.

DISCUSSION

We found a significant relationship between the ability to learn novel speech sounds and the position of the parieto-occipital sulcus. Individuals who are better able to learn to distinguish the non-native from the native sound have a more posterior position of this sulcus relative to non-learners. We did not find an association between measures of phonetic learning and gray or white matter density in primary speech regions of the brain, and this result did not change even when we lowered the statistical threshold in order to perform a directed search in regions of interest. The parieto-occipital finding was specific to the rate of learning; there was no association between the position of this sulcus and pre-training performance, nor between its position and post-minus pre-training performance.

At least two non-exclusive possibilities exist to explain the parieto-occipital results. Anterior to the parieto-occipital sulcus lie white matter fiber tracts connecting temporal and parietal regions of the brain interhemispherically. These fibers cross the caudal part of the corpus callosum (Waddington, '84, de Lacoste et al, '85). It is possible that in fast learners, the more posterior position of the parieto-occipital sulcus is related to a thicker fiber tracts adjacent to this sulcus. This may suggest, in turn, that in fast learners, there is greater inter-hemispheric connectivity in temporal and temporo-parietal auditory-related brain regions.

We tested the hypothesis of greater inter-hemispheric connectivity in learners by examining possible relationships between area measurements of various subdivisions of the corpus callosum and learning measures. We did not find a relationship between phonetic learning ability and the area of the splenium or of the posterior third of the corpus callosum. It is possible that regional differences in the number of fibers in one region (eg; caudal end) of the corpus callosum result in a positional shift of adjacent fibers. For this reason, area measurements of specific subdivisions of the CC may not be well suited to detect possible regional differences. It is also possible that existing differences between our groups in the number of interhemispheric fiber tracts "funnel" together and are more densely packed at the mid-sagittal level, such that they do not result in a detectable overall difference in the area of the CC. In sum, despite our non-significant CC area results, the possibility that there is greater inter-hemispheric connectivity in learners relative to non-learners may still hold. The t-statistic at the midsagittal level of the corpus callosum was not significant using a stringent correction for multiple comparisons; a t-value of 2.73 at the location x= 0, y= -41, and z= 14 suggests that there tends to be more white matter at this location in learners relative to nonlearners. Despite nonsignificance, these results are qualitatively nonetheless interesting and support the interhemispheric connectivity hypothesis.

The second interpretation, not exclusive with the first, for the more posterior position of the parieto-occipital sulcus in learners relative to non-learners is that it may indirectly result from differences in the size or shape of structures adjacent to this sulcus. The parietal lobe, including Brodmann's areas (BA) 39 and 40, roughly corresponding to the angular and supramarginal gyri, respectively, is posteriorly demarcated by the parieto-occipital sulcus. It is possible that in learners, the parietal lobes are larger than in non-learners, resulting in a relatively more posterior location of the parieto-occipital sulcus in the former relative to the latter group. Analyses of variance on parietal volumes did not reveal a main effect of group. We did, however, find differences between groups in the tissue-type (ie; gray versus white matter) of the parietal lobes. Specifically, correlational analyses revealed that in the left hemisphere, there is less gray matter and more white matter in the learners relative to in the non-learners. There was a similar, marginally significant pattern of results in the right parietal region. These results support the idea that the difference in location of the sulcus between our groups is driven by differences in the size of parietal regions. They also support the idea presented above that learners have more interhemispheric white matter fiber tracts than do non-learners because, given that there appears to be more parietal white matter bilaterally, it can be expected that there are also more interhemispheric fibers connecting homotopic parietal regions.

Analyses of variance revealed that for both groups of subjects, there are overall larger parietal volumes in the left relative to the right hemisphere. Furthermore, we found evidence that the asymmetry is present for both learners and nonlearners, but that in learners, there is also a tissue-type by side interaction suggesting that there is more white than gray matter in the left hemisphere but not in the right. Separate analyses of variance on the location of the parieto-occipital sulcus revealed that for both groups, this sulcus is located more posteriorly in the left relative to the right hemisphere, and in addition, that there is a greater asymmetry in the location of the sulcus in the learners relative to in the nonlearners. These findings suggest that larger overall parietal volumes in the left hemisphere correspond to overall more posteriorly located sulcus in this hemisphere relative to the right one. Correlational analyses, however, did not support this interpretation. Furthermore, the above results suggest a differential effect of both sulcal location and of white matter volume in the learners relative to the nonlearners. Specifically, it appears that a greater asymmetry in the location of the sulcus in the learners corresponds to a greater asymmetry in the amount of parietal white matter (ie; more white relative to grey matter in the left hemisphere) in this group relative to in the nonlearners. This overall morphological asymmetry, as well as the differential degree of asymmetry when comparing individuals who are better able to learn new speech sounds to those individuals who are less able, may be related to the known functional asymmetry for speech-related functions in the temporo-parietal regions in and around Wernicke's area. Functional activation studies have more typically shown left-lateralized involvement of parieto-temporal regions in phonetic processing tasks (Paulesu et al, '93, Petersen et al, '89, Démonet at al, '92, '94, Zatorre et al, '92, '96). In the study by Paulesu and colleagues ('93), it was suggested that the left supramarginal gyrus is the substrate underlying the phonological store component of verbal working memory. They also, however, found weaker, non-significant activity in a symmetrical right supramarginal region, suggesting that this region in the non-dominant hemisphere is also involved in some aspect of phonological processing. Our morphological results are also consistent with this latter finding in that we demonstrated that in learners, there are greater white matter volumes *bilaterally* relative to in non-learners.

We found a similar pattern of results for the ability to learn rapidly changing non-linguistic stimuli, but not for the ability to learn steady-state tonal stimuli. Left and right parietal white matter volumes were larger in individuals who learned the non-linguistic stimuli rapidly relative to those who learned them slowly, even when adjusting for performance on the Hindi stimuli. These results suggest that the morphological correlates of phonetic learning in parietal regions are not specific to speech sound learning, but rather, that they are related to a more general, psychoacoustic ability to learn rapidly changing sounds. They are also consistent with behavioral results, which showed an association between phonetic learning abilities and the ability to learn rapidly changing sounds, but no association with the ability to learn steady state sounds. It has been proposed that the ability to process consonant speech sounds depends in part on the ability to track rapidly changing acoustic information (Tallal et al,'93, Schwartz & Tallal, '80, Benasich & Tallal, '96, Belin et al, '98, Johnsrude et al, '97, Zatorre & Belin, '01). Results of electrophysiological (Liégeois-Chauvel et al, '99, Nicholls et al, '99) and of functional brain imaging (Belin et al, '98, Zatorre & Belin, '01, Johnsrude et al, '97) investigations have shown better temporal resolution for the left auditory cortex. In contrast, findings from electrophysiological (Liégeois-Chauvel et al, '01) and from lesion (Johnsrude et al, '00, Robin et al, '90) studies suggest that right auditory regions subserve aspects of pitch and spectral processing. Taken together, these results suggest that different functional substrates underlie temporal versus spectral processing. Our results extend these findings by demonstrating an association in brain *morphology* between phonetic learning ability and the ability to learn rapidly changing nonspeech sounds, together with a dissociation between the morphological correlates of these two stimulus types and those related to the ability to learn steady-state pitch differences.

The finding of more white matter in the left and right parietal region in learners relative to in nonlearners, as well as the finding of a greater asymmetry (left > right) in the amount of white matter relative to gray matter in learners, illustrates the importance of examining white matter morphology in relation to language-related behaviour. In a study examining the morphometry of Heschl's gyrus, Penhune and colleagues ('96) found that the left-greater-than-right asymmetry of the total volume of this gyrus can be explained by greater white but not gray matter volume in the left hemisphere. These results suggest that the well known left-hemisphere dominance for speech may be related to a larger volume of cortical connecting WM fibers, rather than to differences in the size of speech-related GM structures. Several other investigations have also revealed interesting morphological findings in the white matter regions of language-related neural substrates. Paus and colleagues ('99), using structural magnetic resonance images, examined age-related increases in white matter density of fiber tracts in the human brain. They found age-related increases in white matter density along the arcuate fasciculus in the left hemisphere, and suggested that these increases represent a structural component of the corticocortical pathway mediating sensory-motor interactions between the anterior and posterior speech regions. Similarly, Moore and colleagues ('97), based on postmortem data from children ages 5 to 11 years, observed gradual maturation of axons originating in the superficial layers of the auditory cortex.

Differences in WM density or volume can be due to greater myelination and/or to a greater number of WM fibers connection brain regions inter- or intra-hemispherically. The former allows faster conducting of neural signals, resulting in greater efficiency of neural processing. Anderson and colleagues ('99) showed anatomical asymmetries of the posterior superior temporal lobes in postmortem data. They measured myelination sheath and axon diameters with electron microscopy. They found that the posterior superior temporal lobe white matter volume was greater on the left than on the right side. Furthermore, they found that this asymmetry appeared to be due to thicker axonal myelination in the left relative to the right hemisphere, and not to an isolated proliferation of glia, nor to the density of cortical to cortical projections of neurons in the overlying cortex. These findings suggest that an asymmetry of myelination may explain morphological and functional asymmetries in language-related temporal areas of the brain.

Processing of speech sounds depends on the ability to process sounds that change on the time scale of 30-50 milliseconds. Rapid intra- or interhemispheric transfer of information between components of a language network is critical for the processing of such rapidly changing sounds. Future studies can be designed to test the hypothesis of greater interhemispheric connectivity in individuals who are good relative to these who are poor at learning new speech sounds. For example, diffusion tensor mapping techniques can be used to examine the relationship between aspects of behavior and the thickness and/and shape of white matter fiber tracts connecting brain regions inter- and intra-hemispherically. Additionally, electrophysiological recordings can be combined with transcranial magnetic stimulation to measure differences in inter-hemispheric transfer time related to the ability to learn new speech sounds across individuals.

We have demonstrated that differences in white matter morphology in the left and right parietal regions of the brain can in part predict the ability of normal, healthy adults to learn non-native speech sounds. To our knowledge, this is the first illustration of a morphological correlate of a continuous, speech-related behavioral ability which lies within the normal range. This finding raises interesting questions regarding whether such morphological differences exist at birth, or whether there is some degree of experience-dependent structural plasticity during development. It is likely that a combination of both constitutional and experience-related factors are involved. These findings open exciting avenues for future research on morphological correlates of other aspect of human cognition and behavior.

Table 1a: GM maps correlated with Hindi L-Score (rate of learning)

Talaraich Coordinates: X	Y	Z	*t-value*	*Structure*
16	-52	23	-5.73 *	anterior to right parieto-occipital sulcus
18	-61	39	-5.05	"
-14	-66	39	-5.38 *	anterior to left parieto-occipital sulcus
-15	-49	20	-4.74	"

Table 1b: WM maps correlated with Hindi L-Score (rate of learning)

Talaraich Coordinates: X	Y	Z	*t-value*	*Structure*
17	-50	23	5.84 *	anterior to right parieto-occipital sulcus
-16	-64	38	5.09	anterior to left parieto-occipital sulcus

Table 2: Morphometric Analyses

Stimulus type: ***Analysis:***	***Hindi***	***Rapidly changing non-linguistic***	***Tonal***
1. Voxel based morphometry	* parieto-occipital result	ns	ns
2. Anova on position of parieto-occipital sulcus (group, side)	*** more posterior in learners relative to nonlearners *** more posterior in left relative to right hemisphere	(not performed)	(not performed)
3. Anova on volume of parieto-occipital sulcus (group, side)	ns	(not performed)	(not performed)
4. Correlation between corpus callosum (CC) areas and L-scores	ns trend in middle 3rd of CC (larger area in learners)	(not performed)	(not performed)
5. Anova on parietal volumes (group, tissue (GM vs WM), and hemisphere)	*** left > right * learners: R WM > R GM & L WM > L GM		
6. Correlations between parietal GM/WM volumes and L-scores	* left: WM>GM in learners relative to non-learners - right: similar trend as in left hemisphere ($p=0.06$ & 0.07)	* left and right: WM > GM in learners relative to in non-learners, even when adjusting for performance with Hindi stimuli.	ns

ns indicates no significant findings
* indicates $p<0.05$
** indicates $p<0.01$
*** indicates $p<0.001$

Figure Captions

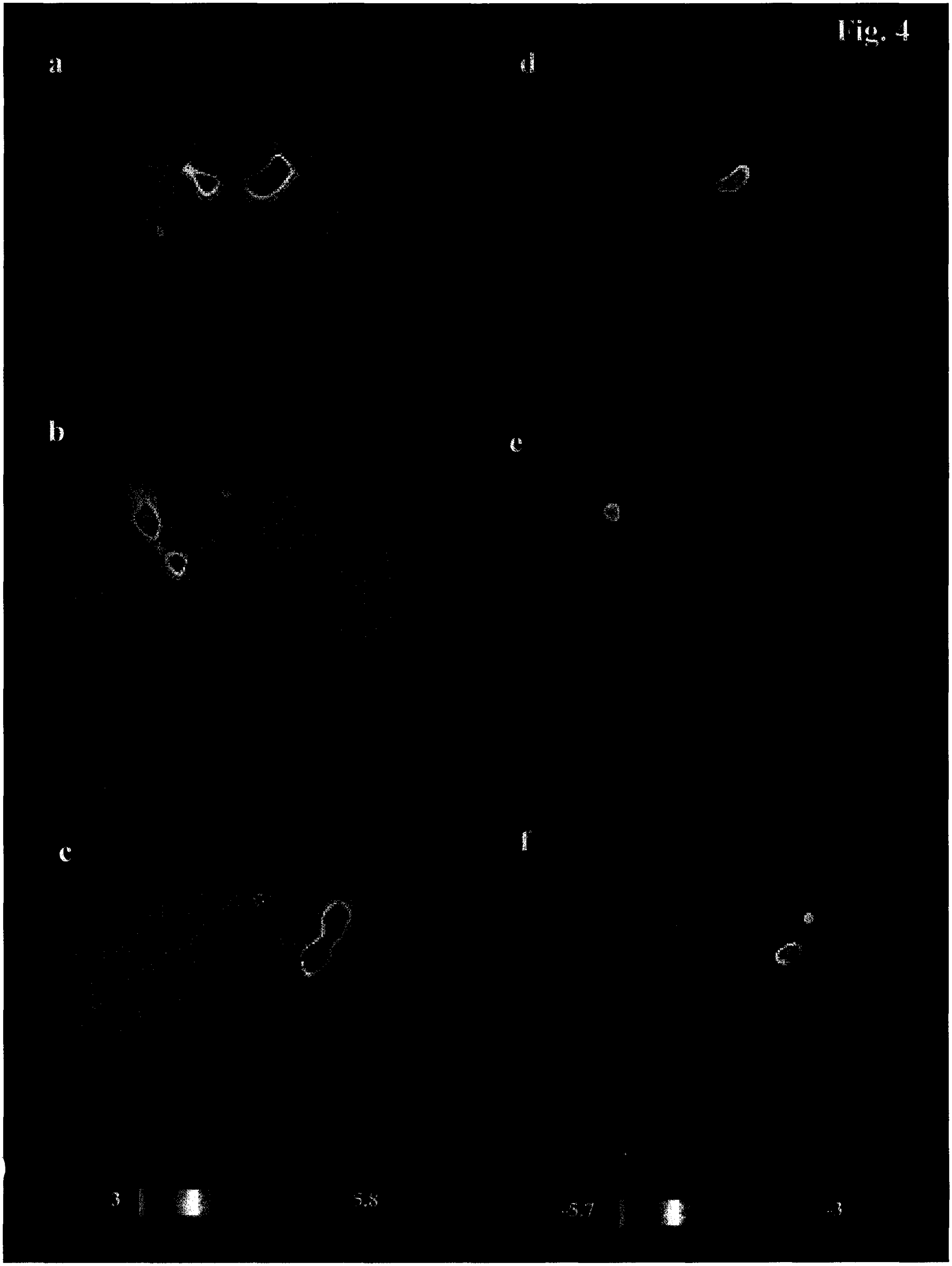
Figure 4: Correlation between behavioral Hindi learning rate scores (L-scores) and: **a-c**: white matter tissue classification maps, and **d-e**: gray matter tissue classified maps (N=59). Locations: **a** & **d**: y= -52, **b** & **e**: x= -14, and **c** & **f**: x= 16. Note difference in scale for figures 4a-4c versus those for figures 4d-4e.

Figure 5: White matter density at peak t-value locations: a) x= -14, y= 66, & z= 39 (left hemisphere), and b) x= 16, y= -52, & z= 23 (right hemisphere), versus L-scores.

Figure 6: Illustration of manual segmentation of parieto-occipital sulcus on individual magnetic resonance images.

Figure 7: Probability maps for parieto-occipital sulcus for the learners and for the nonlearners in the a) left (x= -14) and b) right hemisphere (x= 14).

Fig. 4



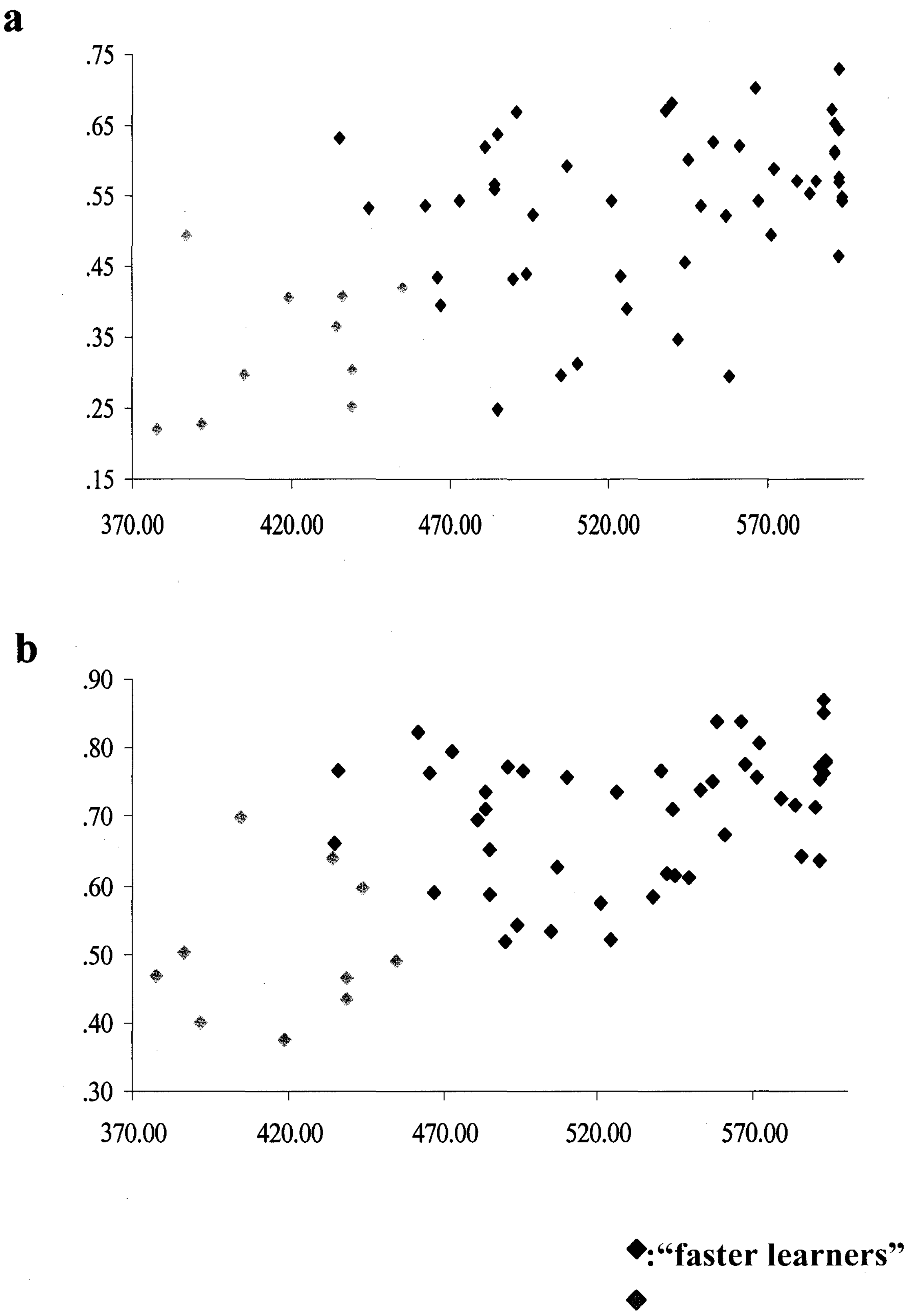

a
.75
.65
.55
.45
.35
.25
.15
370.00
420.00
470.00
520.00
570.00
b
.90
.80
.70
.60
.50
.40
.30
370.00
420.00
470.00
520.00
570.00
◆:"faster learners"
◆

Fig. 5

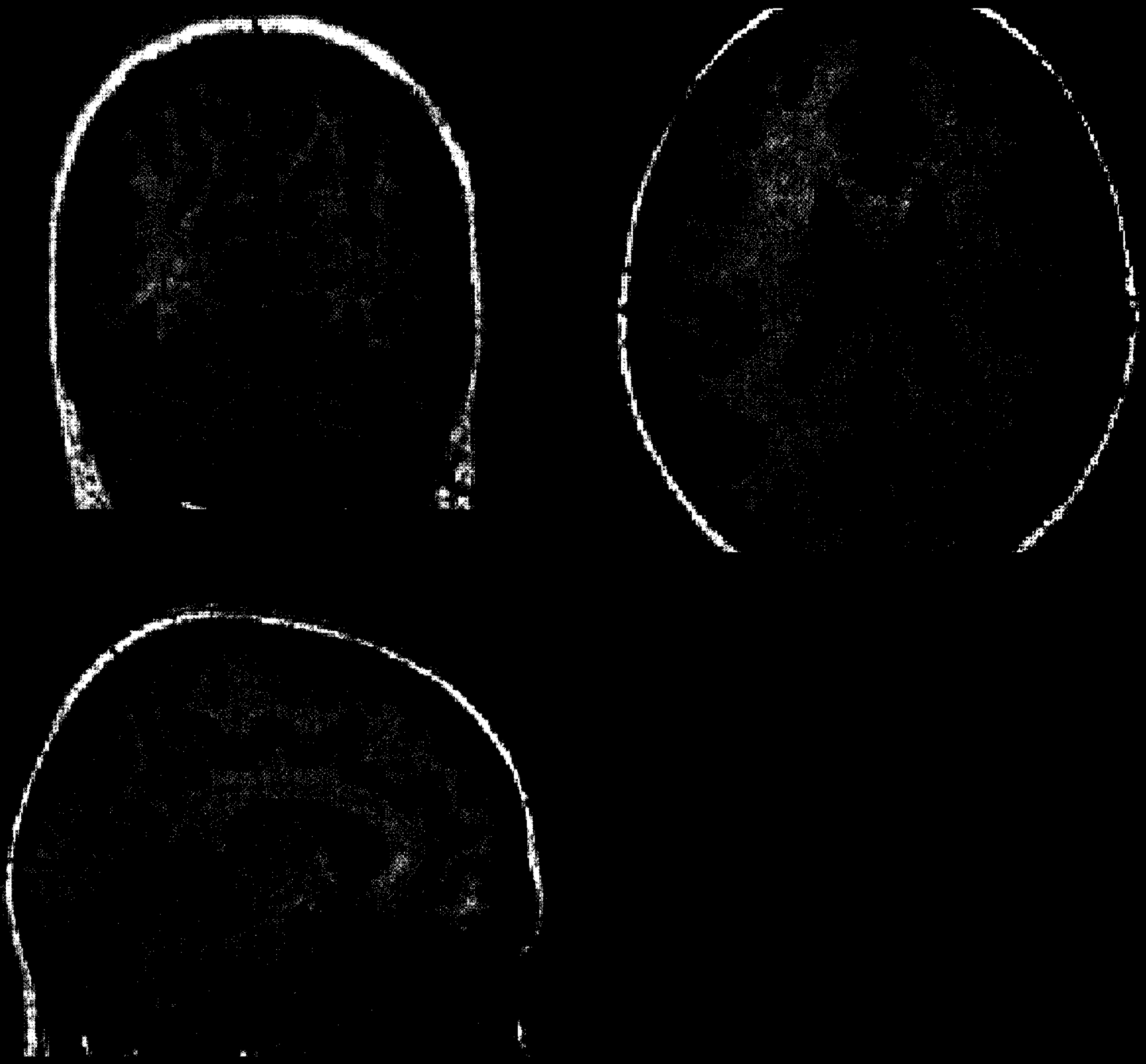


Fig. 6

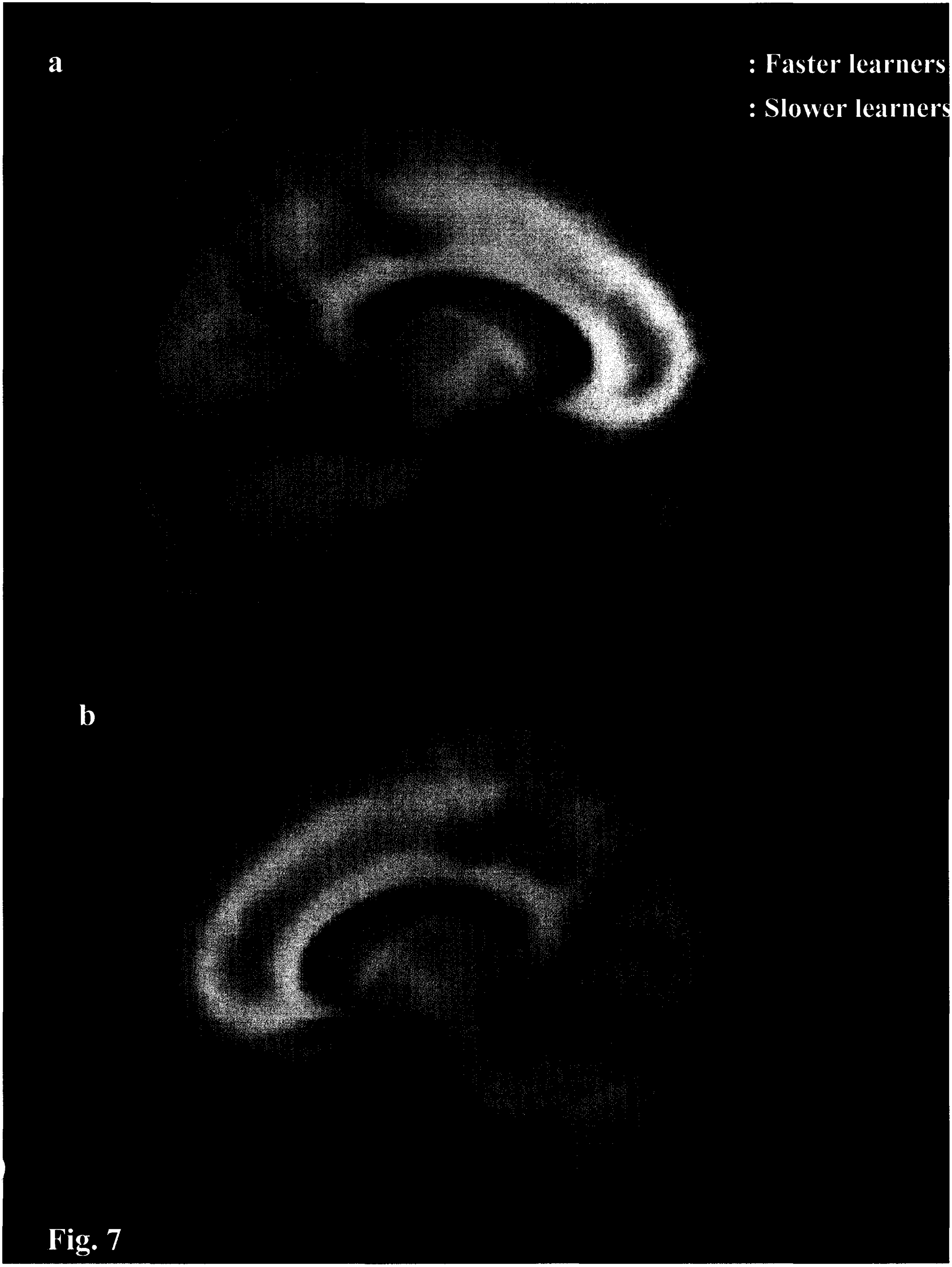


Fig. 7

Chapter 5

GENERAL DISCUSSION

I. Summary of the current findings and integration with previous literature

A. Behavioral study on the learning of speech and non-speech sounds

Results from Study 1 showed that phonetic training improves the ability to identify but not to discriminate the non-native dental-retroflex Hindi contrast. There was, as expected, a large range in both pre- and post-training performance on both measures. Results also showed that the extent to which an individual can learn to discriminate and to identify non-native speech sounds can be in part uniquely predicted by their ability to learn to discriminate *non-linguistic, rapidly changing* stimuli.

The finding of an improvement in the ability to identify the difficult, "fragile" non-native Hindi contrast has important implications for theories of plasticity of phonetic perception in adults. It suggests that the ability to accurately perceive such "difficult" non-native phonemes is not permanently lost in adulthood, even in individuals who have had no early phonemic or allophonic exposure to the retroflex sound. This result is different than those of other studies aimed at ameliorating the perception of the dental-retroflex contrast in non-native speakers; these have shown that standard training approaches do not improve performance, but that more extensive training (eg; five years of language courses), or manipulations which decrease working memory load during task performance, do. These previous training studies were all, however, performed using naturalistic stimuli, which contain more within- and between-category acoustic variability than do synthetic stimuli. In creating synthetic stimuli, only parameters critical to distinguishing the contrast were manipulated. This likely made it easier for subjects to attend to the relevant portions of the sounds, and thereby facilitated learning. Additionally, our large sample size likely allowed sufficient statistical power to detect an overall improvement, even if only a subset of our subjects actually benefited from training.

We were surprised, given that we used a short ISI (500ms), which poses fewer demands on working memory abilities than would the use of a longer ISI, that performance did not improve on the discrimination measure as a function of training. Previous work (Worker & Tees, '84b) has shown that discrimination of the Hindi voiceless, inspirited dental /at/ and retroflex /at/ sounds was significantly better when using an ISI of 500ms than when using an ISI of 1500ms, even without training. Our non-significant effect of training on the A' discrimination measure could be because our stimuli were perceptually more similar (i.e.; more difficult) than the ones used by Worker and Tees. We did, however, find improvement on identification measures. We had expected that the discrimination task would be easier than the identification one because during the former, the first sound can be compared directly with the second sound, whereas during identification, each individually presented sound is compared to encoded templates of the dental and retroflex sounds. It is possible that there was improvement in identification but not in discrimination because the former were more similar to the training paradigm (identification with feedback) than were the latter. Training-related learning may have, for this reason, generalized better to identification than to discrimination test performance.

To conclude on phonetic learning results, our finding that the ability to distinguish this difficult contrast is not permanently lost during development is consistent with results of electrophysiological studies, which have shown that native English speakers pre-attentively perceive the difference between the dental and retroflex sounds (Rivera-Gaxiola et al, 2000a, Rivera-Gaxiola et al, 2000b), despite not showing accurate stimulus detection in a behavioral task (Rivera-Gaxiola et al, 2000a). It has been shown that behavioral training of two slightly different speech stimuli in adults results in a significant change in the duration and magnitude of the MMN cortical potential (Kraus et al, '95), and that this physiological change precedes behavioral discrimination improvements (Tremblay et al, '98, see Kraus & Cheour, 2000). Based on this latter finding, we would predict that with training, subjects in the two studies by Rivera-Gaxiola and colleagues would have learned to accurately detect the sounds of this contrast.

In order to elucidate the mechanism of phonetic learning in adulthood, we also tested the hypothesis that individual differences in phonetic learning are a function of a more general ability to process rapidly changing sounds, regardless of whether or not they are linguistic. Results from behavioral, functional imaging, and dichotic listening studies have suggested that the ability to process speech sounds is related to the ability to parse rapidly changing acoustic information (Tallal et al,'93, Benasich & Tallal, '96, Belin et al, '98, Johnsrude et al, '97). Our results extend these findings, and show that not only baseline phonetic perception but also *potential for change* in phonetic perception is related to low-level psychoacoustic temporal processing. Specifically, we found an association between post-training discrimination performance for the Hindi and for the rapidly changing non-linguistic stimuli, and a dissociation between performance on these two stimulus types and the steady-state tonal stimuli. Identification results were less clear; performance was not only related between the Hindi and the rapidly changing sounds, but also between the Hindi and the tonal stimuli. Partial correlations, however, showed that there was a significant relationship between performance on the Hindi and on the non-linguistic stimuli, even when controlling for the variability shared by performance on the tonal stimuli. The weaker but significant relationship between performance on the Hindi and tonal stimuli may be a function of individual differences in general attentional and motivational factors, which are likely to influence performance on different stimulus types in a similar manner across subjects.

In sum, our results support the idea that there is no permanent loss for the ability to hear "fragile" non-native phonetic contrasts. Additionally, our results extend those of previous work showing that the ability to process speech sounds depends on the ability to track rapidly changing acoustic information (Tallal et al,'93, Benasich & Tallal, '96, Belin et al, '98, Johnsrude et al, '97), and support the influence of psychoacoustic factors in phonetic *learning*.

B. Imaging study on functional plasticity related to phonetic learning

Subjects were scanned using fMRI before and after a two-week period of phonetic training with the Hindi dental-retroflex contrast. We found that behaviorally, training resulted in overall improvement in identification performance, and that there were large individual differences in performance across subjects. Imaging results suggested that training with this non-native phonetic contrast results in the recruitment of areas similar to ones involved during identification of a native contrast, including the left frontal operculum/insula and inferior frontal gyrus, as well as bilateral superior temporal regions. The right frontal operculum/insula (BA 45) and inferior frontal gyrus (BA 44) were also activated in both the native and non-native conditions, although they had not been predicted. Activity was also detected in the left superior parietal region in the native condition, whereas it was not detected in the non-native condition. Results of compound comparisons, however, showed that this region, as well as the left supramarginal region (marginally) are recruited more strongly after relative to before training in the non-native condition. Previous functional imaging studies on speech perception have shown involvement of left frontal regions in and around Broca's area, of left-sided temporoparietal regions including the supramarginal and angular gyri, and of superior temporal regions bilaterally (Zatorre et al, '92, '96, Démonet et al, '92, '94, Binder et al, '96, '97, '00, Paulesu et al, '93, Petersen et al, '88, Fiez et al, '95, Burton et al, '00). The posited roles of these various regions in phonetic processing will be discussed below.

Learning-related functional plasticity

We also found that neural activity in certain brain regions appear to be modulated by the degree of success in learning such that in more successful learners, there was *less* neural activity bilaterally in the frontal opercular/insular regions as well as in the left middle temporal gyrus, while there was also *more* activity in the left and right angular gyri. The strong positive correlation between learning and activation in the angular gyri supports the idea that "good learners" recruit these more posterior temporoparietal regions relatively more than do "poor learners". The posited role of this area will be discussed below. Conversely, the former results suggest that good learners engage frontal opercular regions and the left middle temporal regions to a lesser extent than do poor learners. These results are based on correlations, and do not inform us about the direction of causality, if any, between differences across subjects in brain activation and in performance. We did, however, speculate about the mechanism of this function-behavior relationship, and suggested that in "learners", processing in these areas may be more efficient and/or require less processing time than in "non-learners". This interpretation is consistent with the results of two studies which have shown that prefrontal regions are more active in tasks requiring "top-down", effortful processing relative to during the performance of more “bottom-up”, automatized or practiced tasks. We suggested that “poor learners” in our study performed the identification task in a more top-down, effortful manner, and that “good learners” performed the task in a more automatic, “bottom-up” manner.

Relevant to our findings are those of work by Chee and colleagues (2001), who evaluated the effect of relative language proficiency on brain function in two groups of bilingual subjects. One group spoke English more proficiently than they spoke Mandarin, and the relative proficiency for these languages was reversed for the second group. These two groups were included in order to control for inherent differences in difficulty of one language relative to the other. Subjects performed a semantic task on visually presented words and characters. It was found that behaviorally, greater language proficiency corresponded to shorter reaction times and to greater accuracy in task performance. Imaging results showed that in the more proficient language, there was relatively less activity in the left frontal operculum (BA 45, location: -29, 24, 4) and parietal (BA7: -26, -64, 45) regions. Conversely, the least proficient language was associated with more activation in the left and right frontal opercula (FO) (left: -31, 23, 4 & right: 30, 24, 6). Another group similarly showed marginally more prefrontal activation in the less proficient language of bilinguals during a picture naming task (Hernandez et al, 2000). Chee and colleagues attributed their finding of longer reaction times and of greater frontal activation in the less proficient language to the greater cognitive effort needed to process words in the less familiar language. They suggested that words in the less familiar language may have less well-tuned neural representations, and may therefore require greater neuronal activity. Our findings of more activation in frontal opercular regions in poor learners may similarly reflect greater cognitive effort exerted by individuals who were not successful at acquiring the contrast during training.

Frontal activations

Distinct frontal subregions

We found two distinct frontal activations; one in BA 44, in what we will refer to as the inferior frontal gyrus (IFG), and the other in BA 45, in what we will refer to as the frontal operculum/insula (FO). Previous functional imaging studies examining phonetic perception have more typically shown activation in regions that converge around the location of our “IFG”, BA 44 activation (Zatorre et al, ’92, ’96, Burton et al, ’00, Démonet et al, ’92, ’94). This region has been suggested to be involved in relating the speech signal to articulation during the active performance of tasks involving phonetic segmentation and/or working memory. For example, Zatorre and colleagues (’92, ’96) have proposed that during the performance of a receptive phonetic task, this IFG region is involved when phonetic segments must be extracted and manipulated in relating phonetic information to articulation. It has also been proposed that a left frontal area closer to the IFG location (location: -46, 2, 16) subserves subvocal rehearsal in verbal working memory (Paulesu et al, ’93). Démonet and colleagues (’94) interpreted results from one of their studies in a manner consistent with this; they found that when phoneme monitoring tasks involved perceptual ambiguity, there was activation in the IFG. They attributed this to a greater reliance on verbal rehearsal strategies in working memory during the ambiguous relative to the unambiguous phoneme monitoring conditions.

In contrast, a smaller proportion of previous studies have shown activation at locations similar to our second FO (BA 45) activation, a location more traditionally associated with Broca’s area (Fiez et al, ’95, Chee et al, ’01). This region is also thought to be involved in aspects of articulatory recoding during phonological processing. Fiez and Petersen (’98), in a review of nine neuroimaging studies on word reading, found convergence of results in areas including the left frontal operculum [L 44/45/insula: -41 +/- 9, 8 +/- 5, 6 +/- 4], at locations very similar to ours. Of interest were findings by several groups (Herbster et al, ’97, Fiez et al, unpublished data, Rumsey et al, ’97), that this area was activated when subjects read low frequency exception words and when they read pronounceable non-words, but not when they read low-frequency consistent words. It was suggested that reading of the former two classes of words may activate phonological processing (the ‘assembled” route, where the word is “sounded out” by translation of letters to sounds), whereas that the latter type of word may involve the “direct route” for reading, in which the entire word form is translated into phonological representations.

The “FO” and “IFG” regions likely represent separate functional subregions within what is known as Broca’s area (Zatorre et al, ’96). Our results, for example, showed that activity in the IFG increases overall after relative to before training but that activity in this region is not modulated by learning, whereas that activity in the FO region appears to be both recruited by task performance and modulated by learning. These findings highlight the complexities of differentiating the roles of different frontal subregions, and are consistent with previous work which has provided clear evidence that there are functionally distinct subregions within the left inferior prefrontal cortex (Buckner et al, '95, see Fiez, '97).

Right frontal activations

As described in the discussion of Study 2, we did not expect right frontal activations. This result is, however, consistent with those of several previous studies. Fiez and colleagues ('95) found bilateral frontal opercular activation at locations very close to those in our group subtractions [left: -37, 16, 8 and right: 41, 22, 4], while subjects performed an auditory detection task upon words, syllables, and tone sequences that incorporated rapid temporal changes. Chee and colleagues ('01) also, as reported above, showed greater activation bilaterally in frontal opercular regions, also at similar locations to our findings, during a semantic task in the least proficient language of two languages in bilingual individuals (left: -31, 23, 4 & right: 30, 24, 6). Lastly, Mazoyer and colleagues ('93) performed a PET study in which subjects listened to sentences and to words in the native language, to speech in an unknown language, to semantically anomalous sentences, and to sentences with pseudo-words pronounced like in their native language. They found that only when listening to the sentences with pseudo-words was there an additional activation in the right inferior frontal gyrus. The right hemisphere has been implicated in functions related to certain aspects of pitch perception (Zatorre et al, '92, Zatorre, '88, Johnsrude et al, 2000). It is possible that during the performance of certain language tasks, the right frontal region is recruited under certain stimulus and task conditions during which speech stimuli are processed partly in an acoustic mode, thereby recruiting right frontal regions.

Parietal activations

As noted above, the left superior parietal region was recruited in the group comparison during identification of the native but not of the non-native contrast. A compound comparison, however, showed that the left superior parietal region was significantly more active, and that the left supramarginal region was marginally more active, after relative to before training, suggesting that these areas are recruited when newly learned speech sounds are processed. These regions were not detected in the correlational analysis, suggesting that activity in these areas is not modulated by learning. The same analysis, however, revealed that activity in the left and right angular regions is modulated by learning such that it is more active in learners relative to in non-learners. It is possible that, similar to in frontal regions, the angular, supramarginal, and superior parietal regions of the parietal lobe subserve different aspects of phonological processing. A number of functional imaging studies have implicated inferior parietal regions including the supramarginal and angular gyri in aspects of phonological processing (Démonet et al, '94, Zatorre et al, '92, '96, Binder et al, '96, '97, Petersen et al, '88, Paulesu et al, '93). Paulesu and colleagues ('93) have suggested that the left supramarginal gyrus is the substrate underlying the phonological store component of verbal working memory. Results of clinical studies have shown abnormal function (Rumsey et al, '99) and structure (Duara et al, '91) of angular regions, as well as abnormal functional connectivity between the left angular and temporal regions (Horwitz et al, '98) in individuals with dyslexia.

Based on these results, and on the fact that the parietal lobe is one of two higher-order association areas in the brain (Pandya & Seltzer, '82, Selzer & Pandya, '80, Mesulam et al, '77), it can be speculated that the parietal lobe is involved in integrating phonetic representations of bottom-up, sensory information arising from processing in secondary auditory regions with top-down, long-term memory templates of encoded sounds as well as their corresponding labels. In terms of the roles of posited different parietal functional subregions, it is possible for example that the supramarginal region is involved in linking processed phonetic information with a long-term memory template of previously encoded sounds, and that other parietal regions are involved in aspects of the attentional modulation of the multi-modal integration of incoming information.

Bilateral superior temporal activations

Although the precise role of the left and right secondary auditory regions is not known, it is thought that they are involved in aspects of speech processing, since they are activated by phonetic but not by tonal or noise stimuli (Zatorre et al, '92, Démonet et al, '92, '94, Binder et al, '96, '97). The degree of functional lateralization of function of these regions appears to depend on the type of speech stimuli used, and on whether or not these incorporate semantic, syntactic, or native language phonological information. For example, Binder and colleagues (2000) used fMRI to examine the BOLD signal response in temporal regions of the brain during a material non-specific detection response to words, pseudowords, and to reversed speech. Theyfound that the degree of functional asymmetry in activation was greatest for words, which contain both semantic and phonetic information. It was less for pseudowords, which are phonetically familiar but which do not contain semantic information; there was a non-significant trend towards left-greater-than-right activation in this condition. Finally, the BOLD response asymmetry was least for reversed speech, which is neither phonetically familiar nor semantically relevant. Interestingly, these asymmetry differences were due to differential right and not to left hemisphere activation. In sum, their results show that although there was a left-lateralized response to higher-level speech which incorporates semantic information, there was a bilateral temporal response to speech sounds. Our results are consistent with this latter finding, and with those of previous work (Zatorre et al, '92, Démonet et al, '92, '94, Binder et al, '96, '97).

C. Study on morphological correlates of phonetic learning

The aim of the third study was to determine whether individual differences in the ability to learn speech sounds coming from a non-native language may be in part accounted for by individual differences in brain morphology in language-related or other regions of the brain. Using voxel-based morphometry, we correlated behavioral phonetic learning measures from Study 1 with brain morphology across 59 individuals. Results showed that individuals who more quickly learn to identify the dental versus retroflex sounds during training have a more posteriorly located parieto-occipital sulcus relative to ones who learn more slowly. We suggested that the location of the sulcus may be differentially influenced by the thickness of the adjacent forceps major, an inter-hemispheric fiber tract connecting parietal and temporal regions. This tract lies anterior to the parieto-occipital sulcus, and when thicker, may result in a posterior positional shift of the sulcus.

We suggested a second interpretation, not exclusive with the first, for the parieto-occipital result; the position of the sulcus may be differentially influenced by the size of the adjacent parietal lobe. Results of correlational analyses on the volumes of the parietal lobes confirmed this prediction, and revealed that in the left hemisphere, there is less gray and more white matter in the fast relative to the slow learners. There was a similar trend of results in the right hemisphere. Taken together, these results support the idea that the difference in location of the sulcus between our groups is driven by differences in the size of parietal regions.

Separate analyses revealed that for both groups of subjects, there are overall larger parietal volumes in the left relative to the right hemisphere. There was also evidence that there is more white relative to gray matter in the left hemisphere for learners but not for nonlearners. Consistent with these volumetric results, analyses of variance on the location of the parieto-occipital sulcus revealed that for both groups, this sulcus is located more posteriorly in the left relative to the right hemisphere, and in addition, that there is a greater asymmetry in the location of the sulcus in the learners relative to in the nonlearners. These findings suggest that larger overall parietal volumes in the left hemisphere correspond to overall more posteriorly located sulcus in this hemisphere relative to the right one. Furthermore, the results suggest a differential effect of both sulcal location and of white matter volume in the learners relative to the nonlearners. Specifically, it appears that a greater asymmetry in the location of the sulcus in the learners corresponds to a greater asymmetry in the amount of parietal white matter (i.e.; more white relative to gray matter in the left hemisphere) in this group relative to in the nonlearners. This overall morphological asymmetry, as well as the differential degree of asymmetry when comparing individuals who are better able to learn new speech sounds to those individuals who are less able, may be related to the known functional asymmetry for speech-related functions in the temporo-parietal regions in and around Wernicke's area.

We found a similar pattern of results for the ability to learn rapidly changing non-linguistic stimuli, but not for the ability to learn steady-state tonal stimuli. Left and right parietal white matter volumes were larger in individuals who learned the non-linguistic stimuli rapidly relative to those who learned them slowly, even when adjusting for performance on the Hindi stimuli. These results suggest that the morphological correlates of phonetic learning in parietal regions are not specific to speech sound learning, but rather, that they are related to a more general, psychoacoustic ability to learn rapidly changing sounds. They are also consistent with behavioral results, which showed an association between phonetic learning abilities and the ability to learn rapidly changing sounds, but no association with the ability to learn steady state sounds.

We suggested that the finding of relatively more white matter in the left relative to the right hemisphere in learners may be related to greater intra- or interhemispheric connectivity between parietal, temporal, and frontal speech regions in this group relative to in the nonlearners. Differences in white matter volume can be due to greater myelination, which allows faster conduction of neural signals. Speech sound processing depends on the ability to process sounds that change on the time scale of 30-50 milliseconds. Rapid intra- or interhemispheric transfer of information between components of a language network is critical for the processing of such rapidly changing sounds. Penhune and colleagues ('96) examined the morphometry of Heschl's gyrus, and found that the left-greater-than-right asymmetry of the total volume of this gyrus can be explained by greater white but not gray matter volume in the left hemisphere. These results suggest that the well known left-hemisphere dominance for speech may be related to a larger volume of cortical connecting WM fibers, rather than to differences in the size of speech-related GM structures. In support of the idea that differences in white matter volume may be driven in part by differences in myelination, Anderson and colleagues ('99) measured myelination sheath and axon diameters of the posterior superior temporal lobes with electron microscopy. They found that the posterior superior temporal lobe white matter volume was greater on the left than on the right side, and that this asymmetry appeared to be due to thicker axonal myelination in the left relative to the right hemisphere.

Other investigations have also revealed interesting morphological findings in the white matter regions of language-related neural substrates. Paus and colleagues ('99), using structural magnetic resonance images, examined age-related increases in white matter density of fiber tracts in the human brain. They found age-related increases in white matter density in the left hemisphere along the arcuate fasciculus, and suggested that these increases represent a structural component of the corticocortical pathway mediating sensory-motor interactions between the anterior and posterior speech regions. Similarly, Moore and colleagues ('97), based on postmortem data from children ages 5 to 11 years, observed gradual maturation of axons originating in the superficial layers of the auditory cortex.

The parietal lobe

The inferior parietal lobe includes BAs 39 and 40, which roughly correspond to the angular and supramarginal gyri, respectively. Both of these regions are thought to subserve language-related functions, and as already noted, functional activation of these areas has been observed during tasks involving phonetic processing. As well, as reviewed in the general introduction of this thesis, functional and structural abnormalities in the left parietal region have been detected in dyslexia, a disorder involving specific phonological deficits.

Klingberg and colleagues (2000) have observed, using diffusion tensor magnetic resonance imaging (DTI), that in adults with reading difficulty, there are bilateral microstructural abnormalities in the white matter tracts of temporoparietal regions. In addition, they found that the degree of abnormality in the temporoparietal region of the left hemisphere was significantly correlated with reading scores within the reading-impaired adults *and* within the control group. The authors suggested that white matter microstructure may contribute to reading ability by determining the strength of communication between cortical areas involved in visual, auditory, and language processing. Similarly, our finding of a macrostructural difference in the volume of the parietal WM bilaterally may reflect differences between our groups in the pattern and/or efficiency of communication between auditory and language processing regions of the brain.

Prominent views of the neural correlates of cognitive functions are based on the idea that functional substrates underlying cognitive functioning are distributed, and that they are organized in large networks involving different parts of the brain (Mesulam, '90, Damasio, '89). There is evidence from anatomical (Mesulam et al, '77, Pandya & Seltzer, '82, Seltzer & Pandya, '80) and from single unit recording (Hyvärinen & Shelepin, '79, Leinonen et al, '79, Lynch, '80) studies in monkeys that the inferior parietal lobule contains multimodal neurons, suggesting that this region receives converging input from various regions of the brain. Studies showing abnormal functional connectivity in dyslexic individuals between angular, inferior frontal, extrastriate occipital, and temporal areas (Horwitz et al, '98), and between temporoparietal, frontal, and angular cortices (Paulesu et al, '96), suggest that the language-related deficits arise from abnormal communication between these cortical areas. Similarly, certain types of speech disorders such as conduction aphasia have been observed after damage to the left inferior parietal region; it is debated, however, whether the deficits are due to cortical dysfunction alone (eg; see Anderson et al, '99, Chertkow & Murtha, '97), or to damage to white matter tracts such as the arcuate fasciculus underlying this region (eg; Nass et al, '98). It has even been proposed that inferior parietal areas may serve as a link between frontal and posterior speech regions (Aboitiz & Garcia, '97). These findings highlight the importance of adequate information flow, subserved by the structural integrity of white matter fiber tracts, for normal language functioning.

To conclude, we have demonstrated that individual differences in parietal lobe white matter macrostructure can help to predict individual differences in the ability of normal, healthy adults to learn nonnative speech sounds. Specifically, it appears that a greater amount of white matter in this region, in particular in the left hemisphere, predicts better phonological learning abilities. This finding is the first demonstration of interindividual variability in brain morphology as a function of interindividual differences in a continuous, speech-related measure. It raises interesting questions regarding whether such morphological differences exist at birth, or whether there is some degree of experience-dependant structural plasticity during development. It is likely that a combination of both constitutional and experience-dependant factors are involved. These findings open exciting avenues for future research on morphological correlates of other aspect of human cognition and behavior.

II. Integration of findings from the three studies

Several issues need to be addressed regarding ways in which results from the three studies converge and diverge. First, it is interesting that the left and right parietal lobes were implicated in both the functional and the anatomical studies on phonetic learning. The functional study showed that there is more activity in the angular regions bilaterally in good versus in poor phonetic learners. The anatomical study showed that the left and right parietal lobes are larger in fast relative to in slow learners. It is important, however, to be cautious in drawing links between brain anatomy and brain function, since there is no clear relationship between the two (eg; anatomical connectivity does not imply that the regions in question are always functionally connected). The possible functional roles of these regions, as well as the possible relation between the anatomy of this region and language-related behavior, have already been discussed.

Second, I will discuss some results which might at first appear to diverge. Results of voxel-wise analyses did not reveal morphological difference across subjects in the parieto-occipital region when using learning measures for the rapidly changing nonlinguistic and for the tonal steady-state stimuli. Follow-up, regional morphological analyses on automatically segmented parietal lobe volumes, however, did show an association between the structural correlates of the ability to learn the Hindi and the rapidly changing nonlinguistic stimuli. We found that there is more white and less gray matter in the left parietal region in fast relative to in slow learners for both the Hindi and for the rapidly changing nonlinguistic stimuli, but not for the steady-state, tonal stimuli. There was a similar, marginally significant pattern of results for volumes in the right parietal lobe. These results suggest that morphological differences across subjects in the morphology of the parietal lobes are not speech-specific, but rather, are related to a more general ability to learn to perceive rapidly changing sounds. This is consistent with the results of the behavioral study, which showed an association between post-training performance measures for the Hindi and for the rapidly changing, nonlinguistic stimuli.

It is not clear why in Study 3, the parietal volume analyses but not the voxel-based ones showed a difference in the parieto-occipital region. It can be speculated, however, that the relationship between non-linguistic learning measures and brain morphology is more variable than that between the Hindi learning measures and brain morphology, and that therefore voxel-based analyses did not detect the former relationship while it was sensitive to the latter.

Third and last, in the anatomical study, we found that L-scores but not post- minus pre-training identification and pre-training discrimination measures were significantly related to brain morphology. It is possible that L-scores reflect a different aspect of behavioral improvement than do the post- minus pre-training slopes. For example, L-scores reflect the rapidity of learning over the course of training itself. During training, subjects received immediate feedback for their performance, and so were maybe better able to consistently focus on the critical acoustic cues across trials. Additionally, during the course of training, within each block, subjects only heard one of two synthetic exemplars of the dental and of the retroflex sounds (eg; stimulus 1 vs 7, 2 vs 6, or 3 vs 5). For this reason, within-category physical variability was controlled for. During pre- and post-training identification and discrimination tests, however, subjects heard all of the seven sounds in the dental-retroflex stimulus continuum, and therefore there was more within category physical variability during these tests. Performance may therefore be expected to be more "noisy" during these tests, in part due to greater attentional variability.

Taken together, our results, in the context of previous work, highlight ways in which an understanding of brain structure can complement knowledge about brain function. More traditional approaches to understanding relationships between the brain and behavior is to use techniques such as fMRI and PET to understand functional activation of gray matter brain structures associated with the performance of a particular task. However, further work on the structure of gray and white matter can help to understand the anatomical bases of brain function, and of functional connectivity between brain regions, both in clinical and in normal populations.

Our findings raise interesting questions about the relative influence of genetic versus experiential contributions to morphological and/or functional correlates of the ability to learn new speech sounds. It is likely that both of these factors play a role in determining behavior-related morphological differences across individuals; future studies may address these issues. Our finding that the parieto-occipital structural correlates of phonetic learning exist even when we statistically control for language experience does not support the influence of experiential factors on anatomical correlates of phonetic learning abilities. Such influences may, however, exist; our language background measure likely did not capture the complex nature of linguistic influences during language development. Previous research has shown that the well established anatomical asymmetry of the PT exist even in human newborns (Witelson & Pallie, '73), infants (Wada et al, '75), and fetuses (Chi et al, '77), supporting the influence of genetic or of epigenetic influences on the anatomy of the planum temporale. There is also evidence that a family history of left-handedness predicts PT asymmetry (Steinmetz et al, '91). These results may indirectly support the idea that constitutional influences also play a role in morphological, phonetic learning-related differences across individuals in parietal regions of the brain. A genetic or perigenetic contribution does not exclude the possible additional contribution of post-natal plastic changes related to experience. For example, it has been shown that in monozygotic twins discordant for handedness, the right handers showed leftward asymmetry in the PT, whereas the left-handers lacked the asymmetry. This suggests that ontogenetic factors or differential action of non-genetic factors within twin pairs in utero play a role in the development of structural differences across individuals related to a behavioral trait such as handedness (Steinmetz et al, '95).

In conclusion, given our findings of differences in white matter morphology in a multi-modal brain region known to be anatomically connected to frontal and temporal speech areas, it is conceivable that differences across subjects in white matter connectivity influence the functioning of language networks during the performance of language-related tasks. Our findings open exciting avenues for exploring ways in which differences in brain structure and function can predict aspects of human cognition and behaviour.

References

Aaltonen, O., Eerola, O., Lang, A., Uusipaikka, E., & Tuomainen, J. (1994). Automatic discrimination of phonetically relevant and irrelevant vowel parameters as reflected by mismatch negativity. Journal of the Acoustical Society of America, 96, 1489-1493.

Aboitiz, F. & Garcia, R. (1997). The anatomy of language revisited. Biological Research, 30(4), 171-183.

Abramson, A.S. & Lisker, L. (1970). Discriminability along the voicing continuum: Cross language tests. Proceedings of the 6th International Congress of Phonetic Sciences. Prague: Academia, pp.569-573.

Alho, K., Connolly, J.F., Cheour, M., Lehtokoski, A., Huotilainen, M., Virtanen, J., Aulanko, R., & Ilmoniemi, R. (1998). Hemispheric lateralization in preattentive processing of speech sounds. Neuroscience Letters, 258, 9-12.

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

Anderson, J.M., Gilmore, R., Roper, S., Crosson, B., Bauer, R.M., Nadeau, S., Beversdorf, D.Q., Cibula, J., Rogish, M. 3rd., Kortencamp, S., Hughes J.D., Gonzalez Rothi, L.J., & Heilman, K.M. (1999). Conduction aphasia and the arcuate fasciculus: A reexamination of the Wernicke-Geschwind model. Brain & Language, 70(1), 1-12.

Ashburber, J. & Friston, K.J. (2000). Voxel-based morphometry – the methods. NeuroImage, 11, 805-821.

Baddeley, A. (1886). Working Memory, Clarendon Press/Oxford University Press (Oxford, England, UK).

Belin, P., Zatorre, R.J., Hoge, R., Pike, B., & Evans, A.C. (1999). Event-related fMRI of the auditory cortex. NeuroImage, 10, 417-429.

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M.C., & Samson, Y. (1998). Lateralization of speech and auditory temporal processing. Journal of Cognitive Neuroscience, 10, 536-540.

Benasich, A.A. & Tallal, P. (1996). Auditory temporal processing thresholds, habituation, and recognition memory over the 1st year. Infant Behavior & Development, 19(3), 339-357.

Benson, D.F. (1967). Fluency in aphasia. Cortex, 3, 373-394.

Benson, D.F., Sheremata, W.A., Bouchard, R., Segarra, J.M., Price, D., & Geschwind, N. (1973). Conduction aphasia: a clinicopathological study. Archives of Neurology, Chicago, 28, 339-346.

Bermudez, P. & Zatorre, R.J. (2001). Sexual dimorphism in the corpus callosum: Methodological considerations in MRI morphometry. NeuroImage, 13, 1121-1130.

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.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Springer, J.A., Kaufman, J.N., & Possing, E.T. (2000). Human temporal lobe activation by speech and non-speech sounds. Cerebral Cortex, 10, 512-528.

Binder, J.R., Frost, J.A., Hammeke, T.A., Cox, R.W., Rao, S.M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. The Journal of Neuroscience, 17(1), 353-362.

Binder, J.R., Frost, J.A., Hammeke, T.A., Rao, S.M., & Cox, R.W. (1996). Function of the left planum temporale in auditory and linguistic processing. Brain, 119, 1239-1247.

Binder, J.R., Rao, S.M, Hammeke, T.A., Yetkin, F.Z., Jesmanowicz, A., Bandettini, P.A., Wong, E.C., Estkowski, L.D., Goldstein, M.D., Haughton, V.M., & Hyde, J.S. (1994). Functional magnetic resonance imaging of human auditory cortex. Annals of Neurology, 35(6), 662-672.

Blumstein, S.E., Baker, E., & Goodglass, H. (1977). Phonological factors in auditory comprehension aphasia. Neuropsychologia, 15, 19-30.

Blumstein, S. & Cooper, W. (1974). Hemispheric processing of intonation contours. Cortex, 10, 146-158.

Buckner, R.L., Raichle, M.E., & Petersen, S.E. (1995). Dissociation of human prefrontal cortical areas across different speech production tasks and gender groups. Journal of Neurophysiology, 74, 2163-2173.

Burnham, D.K. (1986). Developmental loss of speech perception: exposure to and experience with a first language. Applied Psycholinguistics, 7, 207-240.

Burnham, D.K., Earnshaw, L.J. Quinn, M.C. (1987). Perceptual development in early infancy: Problems and issues. In The development of the categorical identification of speech. McKenzie, Day, et al (Eds). Hillsdale, NJ, USA: Lawrence Erlbaum Associates. pp. 237-275.

Burton, M.W., Small, S.L., & Blumstein, S.E. (2000). The role of segmentation in phonological processing: An fMRI investigation. Journal of Cognitive Neuroscience, 12(4), 679-690.

Calfee, R., Chapman, R., & Venezky, R. (1972). How a child needs to think to learn to read. In Cognition in learning and memory. Gregg (Ed.), New York: Wiley.

Caramazza, A., Yeni-Komshian, G.N, Zurif, E.B. & Carbone, E. (1973). The acquisition of a new phonological contrast: the case of stop consonants in French-English bilinguals. Journal of the Acoustical Society of America, 54, 421-428.

Carney, A.E., Widin, G.P., & Veimeister, N.F. (1977). Noncategorical perception of stop consonants differing in VOT. Journal of the Acoustical Society of America, 62(4), 961-970.

Chee, M.W.L., Hon, N., Lee, H.L., & Soon, C.S. (2001). Relative language proficiency modulates BOLD signal change when bilinguals perform semantic judgments. NeuroImage, 13, 1155-1163.

Chee, M.W.L., Tan, E.W.L., & Thiel, T. (1999). Mandarin and English single word processing studied with functional magnetic resonance imaging. The Journal of Neuroscience, 19(8), 3050-3056.

Chertkow, H. & Murtha, S. (1997). PET activation and language (Review). Clinical Neuroscience, 4(2), 78-86.

Chomsky, N. (1965). Aspects of Theory of Syntax. Cambridge, Mass: MIT Press.

Collins, D.L., Neelin, P., Peters, T.M., & Evans, A.C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. Journal of Computer Assisted Tomography, 18, 192-205.

Collins, D.L., Holmes, C., Peters, T.M., & Evans, A.C. (1995). Automatic 3D segmentation of neuroanatomical structures from MRI. Human Brain Mapping, 3, 190-208.

Collins, D.L. & Evans, A.C. (1997). Animal: validation and applications of non-linear registration-based segmentation. Internatioal Journal of Pattern Recognition and Artificial Intelligence, 11, 1271-1294.

Collins D.L., Zijdenbos, A.P., Baar, W.F.C., & Evans, A.C (1999). ANIMAL and INSECT: Improved cortical structure segmentation. From: Proceedings of the 16th International Conference on Information Processing in Medical Imaging (IPMI).

Csépe, V. (1995). On the origin and development of the mismatch negativity. Ear & Hearing, 16, 90-103.

Damasio, A.R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. Cognition, 33, 25-62.

DeLacoste, M.C., Kirkpatrick, J.B., & Ross, E.D. (1985). Topography of the human corpus callosum. Journal of Neuropathology and Experimental Neurology, 44(6), 578-591.

Démonet, J.F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.L., Wise, R., Rascol, A., & Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. Brain, 115, 1753-1768.

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

Démonet, J.F., Price, C., Wise, R., & Frackowiak, R.S.J. (1994b). A PET study of cognitive strategies in normal subjects during language tasks: Influence of phonetic ambiguity and sequence processing on phoneme monitoring. Brain, 117, 671-682.

Duara, R., Kushch, A., Gross-Glenn, K., Barker, W., Jallad, B., Pascal, S., Loewenstein, D.A., Sheldon, J., Rabin, M., Levin, B., & Lubs, H. (1991). Neuroanatomic differences between dyslexic and normal readers on magnetic resonance imaging scans. Archives of Neurology, 48, 410-416.

Efron, R. (1963). Temporal perception, aphasia, and déjà vu. Brain, 86, 403-423.

Eimas, P.D., Siqueland, E.R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. Science, 171, 303-306.

Elias, L.J., Bulman-Fleming, M.B., & McManus, I.C. (2000). Linguistic lateralization and asymmetry in interhemispheric transmission time. Brain & Cognition, 43(1-3), 181-185.

Fiez, J.A. & Petersen, S.E. (1998). Neuroimaging studies of word reading. Proceedings of the National Academy of Sciences of the USA, 95, 914-921.

Fiez, J.A., Raichle, M.E., Miezin, F.M., & Petersen, S.E. (1995). PET studies of auditory and phonological processing: Effects of stimulus characteristics and task demands. Journal of Cognitive Neuroscience, 7(3), 357-375.

Fiez, J.A. (1997). Phonology, semantics, and the role of the left inferior prefrontal cortex. Human Brain Mapping, 5, 79-83.

Fledge, J.E. (1984). The effect of linguistic experience on Arabs' perception of the English /s/ vs /z/ contrast. Folia Linguistica, 18, 117-138.

Foundas, A.L., Leonard, C.M., Gilmore, R., Fennell, E., and Heilman, K.M. (1994). Planum temporale asymmetry and language dominance. Neuropsychologia, 32, 1225-1231.

Foundas, A.L., Leonard, C.M., & Heilman, K.M. (1995). Morphologic cerebral asymmetries and handedness. Archives of Neurology, 52, 501-508.

Frith, C.D., Friston, K.D., Liddle, P.F., & Frackowiak, R.S.J. (1991). A PET study of word finding. Neuropsychologia, 29(12), 1137-1148.

Fry, D.B., Abramson, A.S., Eimas, P.D., & Liberman, A.M. (1962). The identification and discrimination of synthetic vowels. Language and Speech , 5, 171-189.

Fujisaki, H. & Kawashima, T. (1969). On the modes and hearing mechanisms of speech perception. Annual Report of the Engineering Research Institute (Tokyo), 28, 67-73.

Fujisaki, H. & Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. Annual Report of the Engineering Research Institute (Tokyo), 29, 207-214.

Fujisaki, H. & Kawashima, T. (1971). A model for the mechanisms of speech perception: Quantitative analysis of categorical effects. Annual Report of the Engineering Research Institute (Tokyo), 30, 59-68.

Galaburda, A.M. (1988). The pathogenesis of childhood dyslexia. In F. Plum (Ed.), Language, communication and the brain. NY: Raven Press, pp. 127-137.

Galaburda, A.M., LeMay, M., Kemper, T.L., & Geschwind, N. (1978). Right-left asymmetries in the brain: Structural differences between the hemispheres may underlie cerebral dominance. Science, 199, 852-856.

Galaburda, A.M., Sherman, G.F., Rosen, G.D., Aboitiz, F., & Geschwind, N. (1985). Perceptual dyslexia: four consecutive patients with cortical abnormalities. Annals of Neurology, 18, 222-233.

Garrett, M. (1980). Levels of processing in sentence production. In: Language Production: Vol. 1 Speech and Talk. B. Butterworth (Ed), New York: Academic Press.

George, M.S., Parekh, P.I., Rosinsky, N., Ketter, T.A., Kimbrell, T.A., Heilman, K.M., Herscovitch, P., & Post, R.M. (1996). Understanding emotional prosody activates the right hemisphere regions. Archives of Neurology, 53, 665-670.

Geschwind, N. (1965). Disconnexion syndromes in animals and man. Brain, 88, 585-644.

Geschwind, N. (1970). The organization of language and the brain. Science, 170, 940-944.

Geschwind, N. (1971). Aphasia. New England Journal of Medicine, 284, 654-656.

Geschwind, N. & Levitsky, W. (1968). Human brain: Left-right asymmetries in temporal speech region. Science, 161, 186-187.

Giard, M., Perrin, E., Pernier, J., & Bouchet, P. (1990). Brain generators implicated in the processing of auditory stimulus deviance: A topographic event-related potential study. Psychophysiology, 27, 627-640.

Giedd, J.N., Blumenthal, J., Jeffries, N.O., Castellanos, F.X., Liu, H., Zijdenbos, A., Paus, T., Evans, A.C., & Rapoport, J.L. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. Nature Neuroscience, 2(10), 861-863.

Golestani, N., Zatorre, R.J., & Paus, T. (unpublished). Learning the Hindi dental-retroflex contrast: phonetic or auditory factors?

Habib, M., Robichon, F., Lévrier, O., Khalil, R., & Salamon, G. (1995). Diverging asymmetries of temporo-parietal cortical areas: a reappraisal of Geschwind/Galaburda theory. Brain and Language, 48, 238-258.

Harnad, S. (1987). Introduction. Psychophysical and cognitive aspects of categorical perception: A critical Review. From Categorical Perception: The groundwork of cognition. S. Harnad (Ed), Cambridge University Press.

Henke, W.L. (1990). MITSYN: A synergistic family of high-level languages for time signal processing. Version 8.1, Belmont, Mass.: Author.

Herbster, A.N., Mintun, M.A., Nebes, R.D. & Becker, J.T. (1997). Regional cerebral blood flow during word-and nonword reading. Human Brain Mapping, 5(2), 85-92.

Hier, D.B., LeMay, M., Rosenberger, P.B., & Perlo, V.P. (1978). Developmental dyslexia: evidence for a subgroup with a reversal of asymmetry. Archives of Neurology, 35, 90-92.

Høien, T. & Lundberg, I. (1989). A strategy for assessing problems in word recognition among dyslexics. Scandinavian Journal of Educational Research, 33, 185-201.

Horwitz, B., Rumsey, J.M. & Donohue, B.C. (1998). Functional connectivity of the angular gyrus in normal reading and dyslexia. Proceedings of the National Academy of Sciences of the USA, 95, 8939-8944.

Hynd, G.W., Semrud-Clikeman, M., Lorys, A.R., Novey, E.S., & Eliopulos, D. (1990). Brain morphology in developmental dyslexia and attention deficit disorder/hyperactivity. Archives of Neurology, 47, 919-926.

Hyvärinen, J. & Shelepin,Y. (1979). Distribution of visual and somatic functions in the parietal association area of the monkey. Brain Research, 169, 561-564.

Illes, J., Francis, W.S., Desmond, J.E., Gabrieli, J.D.E., Glover, G.H., Poldrack, R., Lee, C.J., & Wagner, A.D. (1999). Convergent cortical representation of semantic processing in bilinguals. Brain and Language, 70, 347-363.

Isaacs, E.B., Edmonds, C.J., Lucas, A., & Gadian, D.G. (2001). Calculation difficulties in children of very low birthweight: A neural correlate. Brain, 124(9), 1701-1707.

Jamieson, D.G. & Morosan, D.E. (1986). Training non-native speech contrasts: acquisition of the English /σ/ - /θ/ contrast by francophones. Perception & Psychophysics, 40(4), 205-215.

Jäncke, L., Shah, N.J., Posse, S., Grosse-Ryuken, M., & Müller-Gärtner, H.W. (1998). Intensity coding of auditory stimuli: an fMRI study. Neuropsychologia, 36, 875-883.

Jenkins, J.J. (1979). Four points to remember: A tetrahedral model of memory experiments. In: Levels of processing in human memory. L.S. Cermak & F.I.M. Craik (Eds). Hillsdale NJ: Erlbaum, pp. 429-466.

Jenkins, J.J., Strange, W., & Polka, L. (1995). Not everyone can tell a "rock" from a "lock". From Assessing individual differences in human behaviour. New concepts, methods, and findings. Lubinski & Dawis Eds, p.297-325.

Jernigan, T.L., Hesselink, J.R., Sowell, E., & Tallal, P.A. (1991). Cerebral structure on magnetic resonance imaging in language- and learning-impaired children. Archives of Neurology, 48, 539-545.

Johnson, J.S. & Newport, E.L. (1989). Critical Period Effects in Second Language Learning: The Influence of Maturational State on the Acquisition of English as a Second Language. Cognitive Psychology, 21(1), 60-99.

Johnsrude, I.S., Penhune, V.B., & Zatorre, R.J. (2000). Functional specificity in right human auditory cortex for perceiving pitch direction. Brain, 123, 155-163.

Johnsrude, I.S., Zatorre, R.J., Milner, B.A, & Evans, A.C. (1997). Left-hemisphere specialization for the processing of acoustic transients. NeuroReport, 8, 1761-1765.

Jusczyk, P.W. (1995). Language acquisition: Speech sounds and the beginning of phonology. In: Speech, Language, and Communication. J.L. Miller & P.D. Eimas (Eds), Academic Press.

Jusczyk, P.W., Pisoni, D.B., Reed, M.A., Fernald, A., & Myers, M. (1983). Infants' discrimination of the duration of rapid spectrum changes in nonspeech signals. Science, 222, 175-177.

Kimura, D. (1961). Cerebral dominance in the perception of verbal stimuli. Canadian Journal of Psychology, 15, 166-171.

Kimura, D. (1964). Left-right differences in the perception of melodies. Quarterly Journal of Experimental Psychology, 16, 355-358.

Kimura, D. (1967). Functional asymmetry of the brain in dichotic listening. Cortex, 3, 163-178.

Kimura, D. & Folb, S. (1968). Neural processing of backwards speech sounds. Science, 61, 395-396.

Klein, D. Milner, B., Zatorre, R.J., Meyer, E., & Evans, A.C. (1995). The neural substrates underlying word generation: A bilingual functional-imaging study. Proceedings of the National Academy of the USA, 92, 2899-2903.

Klein, D., Zatorre, R.J., Milner, B., Meyer, E., & Evans, A.C. (1994). Left putaminal activation when speaking a second language: Evidence from PET. NeuroReport, 5(17), 2295-2297.

Kraus, N. & Cheour, M. (2000). Speech sound representation in the brain. Audiology Neuro-Otology, 5, 140-150.

Kraus, N. & Disterhoft, J. (1982). Response plasticity of single neurons in rabbit auditory association cortex during tone-signalled learning. Brain Research, 246, 205-215.

Kraus, N., McGee, T., Carrell, T., King, C., Tremblay, K., & Nicol, T. (1995). Central auditory system plasticity associated with speech discrimination training. Journal of Cognitive Neuroscience, 7, 25-32.

Kraus, N., McGee, T., Littman, T., Nicol, T., & King, C. (1994). Encoding of acoustic change involves non-primary auditory thalamus. Journal of Neurophysiology, 72, 1270-1277.

Kuhl, P.K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. Journal of the Acoustical Society of America, 66, 1668-1679.

Kuhl, P.K. (1980). Perceptual constancy for speech sounds in early infancy. In: Child Phonology: Volume II, Perception. G. Yeni-Komshian, J. Kavanaugh, & C. Ferguson (Eds). New York: Academic Press, pp. 41-66.

Kuhl, P.K. (1983). Perception of auditory equivalence classes for speech in early infancy. Infant Behavior and Development, 6, 263-285.

Kuhl, P.K. (1985). Categorization of speech by infants. In: Neonate Cognition: Beyond the blooming, buzzing confusion. J.Mehler & R. Fox (Eds). Hillsdale, NJ: Erlbaum, pp. 231-262.

Kuhl, P.K. (2000) A new view of language acquisition. PNAS, 97(22), 11850-11857.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by six months of age. Science, 255, 606-608.

Kushch, A., Gross-Glenn, K., Jallad, B., Lubs, H., Rabin, M., Feldman, E., & Duara, R. (1993). Temporal lobe surface area measurements on MRI in normal and dyslexic readers. Neuropsychologia, 31(8), 811-821.

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

Lane, M. (1965). The motor theory of speech perception: A critical review. Psychological Review, 72, 275-309.

Larsen, J.P., Hoien, T., Lundberg, I., & Odegaard, H. (1990). MRI evaluation of the size and symmetry of the planum temporale in adolescents with developmental dyslexia. Brain and Language, 39, 289-301.

Leinonen, L., Hyvarinen J., Nyman, G., & Linnankoski, I. (1979). Functional properties of neurons in the lateral part of associative area 7 in awake monkeys. Experimental Brain Research, 34(2), 299-320.

Liberman, A.M. (1957). Some results of research on speech perception. Journal of the Acoustical Society of America, 29, 117-123.

Liberman, A.M. (1982). On finding that speech is special. American psychologist, 37, 148-167.

Liberman, A.M., Cooper, F.S., Harris, K.S., & MacNeilage, P.F. (1963). A motor theory of speech perception. In: Proceedings of the speech communication seminar. C.G.M.

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

Liberman, A.M., Harris, K.S., Hoffman, H.S., & Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology, 54, 358-368.

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

Liberman, A.M. & Mattingly, I.G. (1989). A specialization for speech perception. Science, 243, 489-494.

Fand (Ed.), 1962. Stockholm: Royal Institute of Technology, Speech Transmission Laboratory.

Liégeois-Chauvel, C., de Graaf, J.B., Laguitton, V., & Chauvel, P. (1999). Specialization of left auditory cortex in man depends on temporal coding. Cerebral Cortex, 9(5), 484-496.

Liégeois-Chauvel, C., Giraud, K., Badier, J.M., Marquis, P., & Chauvel, P. (2001). Intracerebral evoked potentials in pitch perception reveal a functional asymmetry of human auditory cortex. Annals of the New York Academy of Sciences (in press).

Lisker, L. (1985). Preliminaries to a cross-language study of category naming and discrimination. Unpublished manuscript.

Lisker, L. & Abramson, A.S. (1967). The voicing dimension: Some experiments in comparative phonetics. Proceedings of the 6th International Congress of Phonetic Sciences. Prague: Academia, pp.563-567.

Logan, J., Lively, S., & Pisoni, D. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. Journal of the Acoustical Society of America, 89, 874-886.

Loftus, W., Tramo, M., Thomas, C., Green, R., Nordgren, R., & Gazzaniga, M. (1993). Three-dimensional quantitative analysis of hemispheric asymmetry in the human superior temporal region. Cerebral Cortex, 3(4), 348-355.

Lundberg, I. (1989). Lack of phonological awareness: A critical factor in dyslexia. In Brain and Reading. Euler, Lundberg, & Lennerstrand (Eds). NY: Macmillan Co.

Lynch, J.C. (1980). The functional organization of posterior parietal association cortex. Behav Brain Sci, 3, 485-499.

MacDonald, D., Alvis, D., & Evans, A.C. (1994). Multiple surface identification and matching in magnetic resonance images. In: Proceedings of SPIE: Visualization in Biomedical Computing, 2359, pp.160-169.

Mazoyer, B.M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamon, G., Dehaene, S., Cohen, L., & Mehler, J. (1993). The cortical representation of speech. Journal of Cognitive Neuroscience, 5(4), 467-479.

McClaskey, C.L., Pisoni, D.B., & Carrell, T.D. (1983). Transfer of training of a new linguistic contrast in voicing. Perception & Psychophysics, 34(4), 323-330.

McNichol, D. (1972). A Primer of Signal Detection Theory. (Allen and Unwin, London).

Merzenich, M.M., Schreiner, C., Jenkins, W., & Wang, X. (1993). Neural mechanisms underlying temporal integration, segmentation, and input sequence representation: some implications for the origin of learning disabilities. Annals of the New York Academy of Sciences, 682, 1-22.

Mesulam, M.M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language and memory. Annals of Neurology, 28, 597-613.

Mesulam, M.M., Van Hoesen, G.W., Pandya, D.N., & Geschwind, N. (1977). Limbic and sensory connections of the inferior parietal lobule (area PG) in the rhesus monkey: A study with a new method for horseradish peroxidase histochemistry. Brain Research, 136, 393-414.

Moore, J., Guan, Y., & Wu, B.J. (1997). Assoc. Res. Otolaryngol. Abstr., 20, 28.

Morosan, D.E. & Jamieson, D.G. (1986). Efficient training of non-native speech discrimination with perceptual fading. Proceedings of the 12th International Congress on Acoustics, Canadian Acoustical Association.

Mummery, C.J., Ashburner, J., Scott, S.K., & Wise, R.J. (1999). Functional neuroimaging of speech perception in six normal and two aphasic subjects. Journal of the Acoustical Society of America, 106, 449-457.

Näätänen, R., Gaillard, A., & Mäntysali, S. (1978). Early selective attention effect on evoked potential reinterpreted. Acta Psychologia, 42, 313-329.

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M, & Paavilainen, P. (1993). Development of a memory trace for a complex sound in the human brain. Neuroreport, 4, 503-506.

Nass, R., Leventhal, F., Levine, B., Lebron, D., Maxfield, C., McCaul, P., George, A., & Allen, J. (1998). Conduction aphasia in a 3-year-old with a left posterior cortical/subcortical abscess. Brain & Language, 62(1), 70-88.

Neville, H.J. (1995). Developmental specificity in neurocognitive development in humans. In: The Cognitive Neurosciences. Gazzaniga (Ed), pp. 219-233.

Nicholls, M.E.R., Schier, M., Stough, C.K.K., & Box, A. (1999). Psychophysical and electrophysiologic support for a left hemisphere temporal processing advantage. Neuropsychiatry, Neuropsychology, & Behavioral Neurology, 12(1), 11-16.

Pandya, D.N. & Selzer, B. (1982). Intrinsic connections and architectonics of posterior parietal cortex in the rhesus monkey. Journal of Comparative Neurology, 204, 196-210.

Pantev, C., Wollbrink, A., Roberts, L.E., Engelien, A., & Lütkenhöner, B. (1999). Short-term plasticity of the human auditory cortex. Brain Research, 842, 192-199.

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

Pastore, R.E. (1976). Categorical perception: A critical re-evaluation. In: Hearing and Davis: Essays Honoring Hallowell Davis. S.K. Hirsh, D.H. Eldridge, I.J. Hirsh, & S.R. Silverman (Eds), St. Louis, MO: Washington University Press.

Pastore, R.E., Ahroon, N.A., Buffuto, K.J., Friedman, C., Puleo, J.S., & Fink, E.A. (1977). Common-factor model of categorical perception. Journal of Experimental Psychology: Human Perception and Performance, 3, 676-696.

Paulesu, E., Frith, U., Snowling, M., Gallagher, M., Morton, J., Frackowiak, R.S.J., & Frith, C.D. (1996). Is developmental dyslexia a disconnection syndrome? Evidence from PET scanning. Brain, 119, 143-157.

Paus, T., Tomaiuolo, F., Otaky, N., MacDonald, D., Petrides, M., Atlas, J., Morris, R., & Evans, A.C. (1996). Human cingulate and paracingulate sulci: pattern, variability, asymmetry, and probabilistic map. Cerebral Cortex, 6, 207-214.

Paus, T., Zijdenbos, A., Worsley, K., Collins, D.L., Blumenthal, J., Giedd, J.N., Rapoport, J.L., & Evans, A.C. (1999). Structural maturation of neural pathways in children and adolescents: in vivo study. Science, 283, 1908-1911.

Penhune, V.B., Zatorre, R.J., MacDonald, J.D., & Evans, A.C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: Probabilistic mapping and volume measurements from magnetic resonance scans. Cerebral Cortex, 6, 661-672.

Pennington, B.F., Lefly, D.L., Van Orden, G.C., Bookman, M.O., & Smith, S.D. (1987). Is phonology bypassed in normal or dyslexic development? Annals of Dyslexia, 37, 62-89.

Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M., & Raichle, M.E. (1988). Positron emission tomographic studies of the cortical anatomy of single word processing. Nature, 331, 585-589.

Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M., & Raichle, M.E. (1989). Positron emission tomography studies of the processing of single words. Journal of Cognitive Neuroscience, 1, 153-170.

Pisoni, D.B. (1973). Auditory and phonetic codes in the discrimination of consonants and vowels. Perception & Psychophysics, 13, 253-360.

Pisoni, D.B. (1975). Auditory short-term memory and vowel perception. Memory and Cognition, 3, 7-18.

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.

Plante, E., Swisher, L., Vance, R., & Rapcsak, S. (1990). MRI findings in boys with specific language impairment. Brain & Language, 40, 52-66.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. JASA, 89(6), 2961-2977.

Polka, L. & Werker, J.F. (1994). Developmental changes in the perception of non-native vowel contrasts. Journal of Experimental Psychology: Human Perception and Performance, 20, 421-435.

Priutt, 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. JASA, 87, Suppl.1, S72.

Rack, J.P., Snowling, M.J., & Olson, R.K. (1992). The nonword reading deficit in developmental dyslexia: A review. Reading Research Quarterly, 27(1), 29-53.

Raichle, M.E., Fiez, J.A., Videen, T.O., MacLeod, A.K., Pardo, J.V., Fox, P.T., & Petersen, S.E. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. Cerebral Cortex, 4, 8-26.

Ratcliff, G., Dila, C., Taylor, L., & Milner, B. (1980). The morphological asymmetry of the hemispheres and cerebral dominance for speech: A possible relationship. Brain and Language, 11, 87-98.

Recanzone, G., Schreiner, C., & Merzenich, M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. Journal of Neuroscience, 13, 87-104.

Repp, B.H. (1984). Categorical perception: Issues, methods, findings. In: Speech and Language: Advances in Basic Research and Practice, 10. N.J. Lass (Ed). New York: Academic Press, pp. 243-335.

Repp, B.H. & Liberman, A.M. (1987). Phonetic category boundaries are flexible. In Categorical Perception: The Groundwork of Cognition. S. Harnad (Ed.), Cambridge University Press.

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

Rivera-Gaxiola, M., Johnson, M.H., Csibra, G., & Karmiloff-Smith, A. (2000b). Electrophysiological correlates of category goodness. Behavioural Brain Research, 112(1-2), 1-11.

Robin, D.A., Tranel, D., & Damasio, H. (1990). Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. Brain & Language, 39(4), 539-555.

Rohen, J.W. & Yokochi, C. (1983). Color atlas of anatomy. NY; Tokyo: Igaku-Shoin, pp.84.

Ross, E.D. & Mesulam, M.M. (1979). Dominant language functions of the right hemisphere? Prosody and emotional gesturing. Archives of Neurology, 36, 144-148.

Ross, E.D., Thompson, R.D., & Yenkosky, J. (1997). Lateralization of affective prosody in brain and the callosal integration of hemispheric language functions. Brain and Language, 56, 27-54.

Rumsey, J.M., Andreason, P., Zametkin, A.J., Aquino, T., King. A.C., Hamburger, S.D., Pikus, A., Rapoport, J.L., & Cohen, R.M. (1992). Failure to activate the left temporoparietal cortex in dyslexia: An oxygen 15 positron emission tomographic study. Archives of Neurology, 49, 527-534.

Rumsey, J.M., Dorwart, R., Vermess, M., Denckla, M.B., Kruesi, M.J.P., & Rapoport, J.L. (1986). Magnetic resonance imaging of brain anatomy in severe developmental dyslexia. Archives of Neurology, 43, 1045-1046.

Rumsey, J.M., Horwitz, B., Donohue, B.C., Nace, K.L., Maisog, J.M., & Andreason, P. (1997). Phonological and orthographic components of word recognition. A PET-rCBF study. Brain, 120, 739-759.

Rumsey, J.M., Horwitz, B., Donohue, B.C., Nace, K.L., Maisog, J.M. & Andreason, P. (1999). A functional lesion in developmental dyslexia: Left angular gyral blood flow predicts severity. Brain and Language, 70, 187-204.

Samuel, A.G. & Tartter, V.C. (1986). Acoustic-phonetic issues in speech perception. Annual Review of Anthropology, 15, 247-273.

Schlaug, G., Jäncke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. Science, 267, 699-701.

Schwartz, J. & Tallal, P. (1980). Rate of acoustic change may underlie hemispheric specialization for speech perception. Science, 207, 1380-1381.

Selzer, B. & Pandya, D.N. (1980). Converging visual and somatic sensory cortical input to the intraparietal sulcus of the rhesus monkey. Brain Research, 192, 339-351.

Shankweiler, D. & Studdert-Kennedy, M. (1967). Identification of consonants and vowels presented to left and right ears. Quarterly Journal of Experimental Psychology, 19, 59-63.

Sled, J.G., Zijdenbos, A.P., & Evans, A.C. (1998). A non-parametric method for automatic correction of intensity non-uniformity in MRI data. IEEE Trans Med Imag, 17, 87-97.

Steinmetz, H., Herzog, A., Schlaug, G., Huang, Y., & Jäncke, L. (1995). Brain (a)symmetry in monozygotic twins. Cerebral Cortex, 5, 296-300.

Steinmetz, H., Rademacher, J., Jäncke, L., Huang, Y., Thron, A., & Zilles, K. (1990). Total surface of temporoparietal intrasylvian cortex. Diverging left-right asymmetries. Brain and Language, 39, 357-372.

Steinmetz, H., Volkmann, J., Jäncke, L., & Freund, H.J. (1991). Anatomical left-right asymmetry of language-related temporal cortex is different in left- and right-handers. Annals of Neurology, 29, 315-319.

Stevens, K.N. (1981). Constraints imposed by the auditory system on the properties used to classify speech sounds: Data from phonology, acoustics and psycho-acoustics. In: The Cognitive Representation of Speech. T.F. Myers, J. Laver, & J. Anderson (Eds). Amsterdam: North Holland.

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

Strange, W. & Jenkins, J. (1978). Role of linguistic of experience in perception of speech. In: Perception and Experience. R.D. Walk & H.L.Pick (Eds). Plenum, N.Y., pp. 125-169.

Strange, W., Polka, L., & Aguilar, M.C. (1989). Effects of auditory and phonetic training on Americans' discrimination of Hindi retroflex-dental contrasts. JASA, 86, Suppl.1, S101.

Studdert-Kennedy, M. & Shankweiler, D. (1970). Hemispheric specialization for speech perception. Journal of the Acoustical Society of America, 48, 579-594.

Talairach, J. & Tournoux, P. (1988). Co-Planar Stereotaxic Atlas of the Human Brain (Thieme, New York).

Talairach, J. & Tournoux, P. (1993). Referentially oriented cerebral MRI anatomy. New York: Thieme Verlag.

Tallal, P., Miller, S., & Fitch, R.H. (1993). Neurobiological basis of speech: A case for the preeminence of temporal processing. Annals of the New York Academy of Sciences, 682, 27-47.

Tallal, P. & Newcombe, F. (1978). Impairment of auditory perception and language comprehension in dysphasia. Brain & Language, 5, 13-24.

Tartter, V.C. (1981). A comparison of the identification of synthetic vowel and stop consonant stimuli with various acoustic properties. Journal of Phonetics, 9, 447-486.

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 Experimental Analysis of Behavior, 6, 1-27.

Trehub, S. (1976). The discrimination of foreign speech contrasts by infants and adults. Child Development, 47; 466-472.

Tremblay, K., Kraus, N., Carrell, T.D., & McGee, T. (1997). Central auditory system plasticity: Generalization to novel stimuli following listening training. Journal of the Acoustical Society of America, 102, 3762-3773.

Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. NeuroReport, 9(16), 3557-3560.

Vargha-Khadem, F., Watkins, K.E., Price, C.J., Ashburner, J., Alcock, K.J., Connelly, A., Frackowiak, R.S., Friston, K.J., Pembrey, M.E., Mishkin, M., Gadian, D.G., & Passingham, R.E. (1998). Neural basis of an inherited speech and language disorder. Proceedings of the National Academy of the USA, 95, 12695-12700.

Wada, J.A., Clarke, R., & Hamm, A. (1975). Cerebral hemispheric asymmetry in humans: Cortical speech zones in 100 adults and 100 infant brains. Archives of Neurology, 32, 239-245.

Waddington, M.M. (1984). Atlas of Human Intracranial Anatomy. Vermont: Rutland (Academic books).

Watkins, K.E., Paus, T., Lerch, J.P., Zijdenbos, A., Collins, D.L., Neelin, P., Taylor, J., Worsley, K.J., & Evans, A.C. (2001). Structural asymmetries in the human brain: a voxel-based statistical analysis of 142 MRI scans. Cerebral Cortex, 11, 868-877.

Weber-Fox, C.M. & Neville, H.J. (1992). Maturational constraints on cerebral specializations for language processing: ERP and behavioral evidence in bilingual speakers. Soc. Neurosci., 18(1), 335.

Werker, J.F., Gilbert, J.H.V., Humphrey, K., and Tees, R.C. (1981). Developmental aspects of cross-language speech perception. Child Development, 52, 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 of three factors in speech perception. Perception & Psychophysics, 37(1), 35-44.

Werker, J.F. & Polka, L. (1993). Developmental changes in speech perception: New challenges and new directions. Journal of Phonetics, 21 (1-2), 83-101.

Werker, J.F. & Tees, R.C. (1983). Developmental changes across childhood in the perception of nonnative speech sounds. Canadian Journal of Psychology, 37, 278-286.

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

Werker, J.F. & Tees, R.C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. JASA, 75(6), 1866-1878.

Westbury, C.F., Zatorre, R.J., & Evans, A.C. (1999). Quantifying variability in the planum temporale: a probability map. Cerebral Cortex, 9(4), 392-405.

Williams, L. (1974). Speech perception and production as a function of exposure to a second language. Unpublished Doctoral Dissertation, Harvard University.

Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. Perception & Psychophysics, 21, 289-297.

Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., & Evans, A.C. (1996). Human Brain Mapping, 4, 58-73.

Wright, I.C., McGuire, P.K., Poline, J.B., Travere, J.M., Murray, R.M., Frith, C.D., Frackowiak, R.S., & Friston, K.J. (1995). A voxel-based method for the statistical analysis of gray and white matter density applied to schizophrenia. NeuroImage, 2, 244-252.

Zatorre, R.J. (1983). Category-boundary effects and speeded sorting with a harmonic musical-interval continuum: Evidence for dual processing. Journal of Experimental Psychology: Human Perception and Performance, 9(5), 739-752.

Zatorre, R.J. (1988). Pitch perception of complex tones and human temporal lobe function. Journal of the Acoustical Society of America, 84(2), 566-572.

Zatorre, R.J. & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. Cerebral Cortex, 11, 606-618.

Zatorre, R.J. and Binder, J.R. (2000). Functional and structural imaging of the human auditory system. In: Brain Mapping: The systems. A.W.Toga and J.C. Mazziota (Eds.) Academic press. Pp 365-402.

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

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, 21-30.

Zatorre, R.J., Perry, D.W., & Beckett, C.A. (1998). Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. Proceedings of the National Academy of Sciences of the USA, 95, 3172-3177.

Zatorre, R.J. & Samson, S. (1991). Role of the right temporal neocortex in retention of pitch in auditory short-term memory. Brain, 114, 2403-2417.

Zijdenbos, A.P., Forghani, R., & Evans, A.C. (1998). Automatic quantification of MS lesions in 3D MRI brain data sets: validation of INSECT. Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. (MICCAI). 439-448.