Cortical competition during language discrimination

Jingjing Zhao [a], Hua Shu [a,*], Linjun Zhang [a], Xiaoyi Wang [a], Qiyong Gong [b], Ping Li [c,d,*]

[a] State Key Laboratory for Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China
[b] Huaxi MR Research Center (HMRRC), Department of Radiology, West China Hospital of Sichuan University, China
[c] Division of Behavioral and Cognitive Sciences, National Science Foundation, VA 22230, USA
[d] Department of Psychology, Pennsylvania State University, PA 16802, USA

ARTICLE INFO

ABSTRACT

How do human listeners differentiate one language from another? In this study we examine the contributions of acoustic and linguistic cues to successful language discrimination. In particular, we report findings that reveal patterns of cortical competition as a function of the competition between prosodic, phonological, and lexical semantic information during language discrimination. We manipulated four types of stimuli in the listening environment: synthesized speech with rhythmic information, synthesized speech with rhythmic plus intonational information, natural speech from Japanese and Italian, and natural speech from Chinese and English. Our study shows that, depending on the amount and the kind of cues available, the listener recruits different areas of the brain for the same language task. Furthermore, brain activations do not monotonically multiply as a function of the complexity of the cues available, but are the outcomes of cue competition as a function of cue validity for the discrimination task. These findings show how acoustic and linguistic cues lead to cortical competition and how cortical activities adapt to the task demand for successful information processing.

Published by Elsevier Inc.

Introduction

An important human linguistic ability is the discrimination of languages from one another in the speech environment. This language discrimination ability occurs very early in human infants. A major task facing the bilingual child from the beginning is the discrimination of the two or more languages in the learning environment. In a few short months, bilingual infants can use a variety of acoustic and linguistic cues to distinguish between native and non-native languages (Bosch and Sebastián-Gallés, 1997, 2001a, 2001b). Furthermore, successful language discrimination does not seem to depend on early experiences with multiple languages: normal monolingual adults can easily distinguish native or familiar languages from foreign, unfamiliar languages. They can also discriminate between dialects due to accents from distinct linguistic regions. In the United States, listeners can often identify speakers of southern dialects, and in China, experienced listeners can readily differentiate the seven major dialects spoken in various parts of the country. Our study here investigates the neural mechanisms underlying such language discrimination abilities.

Early language discrimination by a general perceptual mechanism

Although the goal of the current research is the understanding of language discrimination abilities in adults, it is worthwhile to point out that recent research suggests early precursors of these abilities. Insights into infant speech perception are highly relevant to the similarities and differences between child and adult language processing (see Discussion). For example, infant studies show that babies can use acoustic features of speech to distinguish between unfamiliar languages that have distinct rhythmic or prosodic structures (e.g., English vs. French, or Dutch vs. Japanese; Nazzi et al., 1998; Ramus, 2002; Ramus et al., 2000). By 4 or 5 months of age, they can further discriminate between languages that have the same rhythmic patterns as their native language (e.g., Spanish vs. Catalan or English vs. Dutch; see Bosch and Sebastián-Gallés, 1997; Nazzi et al., 2000; Nazzi and Ramus, 2003), and at this age, they can also rely on nonlinguistic cues (e.g., oral–facial movements that accompany speech) for successful language discrimination (Weikum et al., 2007). As their language experience accumulates, older infants start to use more phonological and lexical knowledge (e.g., phonotactics) than prosodic cues to separate native and foreign languages.

Such early linguistic abilities have led to the conjecture that language discrimination could rely on general perceptual or psychoacoustic processes inherent in the primate auditory system. This conjecture has received support from parallel studies of human infants and cotton-top tamarin monkeys exposed to natural and synthesized speech materials (Ramus et al., 2000; Tincoff et al., 2005;

see also Toro et al., 2003 for a study of rats in language discrimination tasks). These studies show that human infants and non-human primates have an auditory system that is well attuned to detecting basic acoustic, specifically, rhythmic differences in different languages. For example, the use of prosodic information characterizes language discrimination in both human infants and non-human primates, suggesting that the perceptual mechanism used for language discrimination is a general one, unlikely to have evolved specifically for language (Tincoff et al., 2005).

Lexical and prelexical cues for language discrimination

What cues can the human perceptual system make use of during language discrimination? Ramus and Mehler (1999) considered three 'prelexical cues' that are important for early language discrimination: phonetic repertoire, phonotactic constraints, and prosody. Phonetic repertoire refers to the set of phonemes, consonants and vowels, of a given language. For example, English has a phonemic system that contains 24 consonants and 14 vowels, Chinese 21 consonants and 10 vowels, and Spanish 18 consonants and 5 vowels. These repertoire differences imply that each language has a unique set of phonemes that are distinctive of that particular language (e.g., retroflex consonants are used in Chinese but not English or Spanish). As infants accumulate linguistic experience with the specific language or languages, they can use phonetic repertoire as a reliable cue to discriminate languages. In addition to language-specific phonemes, human languages also differ in terms of the structural constraints that they place on the distribution or co-occurrence of phonemes, referred to as phonotactics. For example, consonant clusters (/fl/, /pr/, /str/, etc) are commonly used in forming English words, but they are not allowed in Chinese. Phonotactic information thus provides a reliable cue to older infants for language discrimination. Of particular interest among the 'prelexical cues' discussed by Ramus and Mehler (1999) is prosody, since the processing of prosody appears to tap into a general perceptual or psychoacoustic mechanism used by both humans and non-human primates.

Prosody can collectively refer to suprasegmental features in the running speech, including most notably rhythm, stress, and intonation. With respect to rhythm, linguists have traditionally classified the world's languages into three rhythm classes: stress-timed languages such as English and other Germanic languages, syllable-timed languages such as French and other Romance languages, and mora-timed languages such as Japanese (see Ramus et al., 1999). While the exact nature of the different rhythm types is a matter of linguistic debate (Ramus et al., 1999, 2000), it seems that both infants and adult listeners can reliably use rhythmic information to distinguish languages of different rhythm classes. In addition to rhythm, intonation also provides important prosodic information for listeners to use during language discrimination. Intonation on the phrasal or sentential level often indicates grammatical or pragmatic differences, while on the word level it may distinguish lexical meanings (e.g., lexical tones in Chinese and Thai). Although lexical tone is characteristic of many of the world's languages and the perception of tone has generated much research attention (see Gandour, 2006; Jongman et al., 2006 for reviews), the role of tonal information has not been directly examined for language discrimination (but see Zhang, 2006).

Phonetic repertoire, phonotactics, rhythm and intonation are all parts of the 'prelexical' information. Although they differ in whether the information is segmental (phonemes, phonotactics) or supra-segmental (rhythm, intonation), they contrast with lexical knowledge that contains the semantic and conceptual representation of words. Clearly, both prelexical information and lexical semantic knowledge jointly contribute to language discrimination for the normal adult listener. When the infant has not gained a great deal of lexical knowledge, however, she has to rely more on prelexical cues to

discriminate languages. When an adult is dealing with two unfamiliar languages he can also only use prelexical information. But as the infant's linguistic experience enriches or when the adult is handling familiar languages (native and second languages), lexical semantic information will start to play a very important, sometimes determining role in language discrimination, because this type of information makes meaning differences and hence are the most reliable. Indeed, the use of lexical knowledge might be automatic, since listeners cannot inhibit lexical information that has both high 'cue validity' and high 'cue reliability' for the processing task at hand, in the terms of the Competition Model of Bates and MacWhinney (1982, 1987, 1989) (See Discussion for further details).

Neural mechanisms of language discrimination

How do prelexical (prosodic and phonological) and lexical (semantic) cues compete and interact during language discrimination? How is the competition of various cues reflected in cortical activities? Answers to these questions are limited, since the research so far has focused on prelexical information, in particular, prosodic cues such as rhythm and intonation. Moreover, previous research has not been concerned with the neural correlates of language discrimination. The current study seeks to understand the nature and the neural mechanisms underlying the use of both pre-lexical information and lexical semantic knowledge during language discrimination.

Although there has been little neural evidence regarding the use and interaction of prosodic cues and lexical knowledge in the language discrimination context, the brain functions of prosodic, phonological, and lexical semantic processing have become increasingly clear from a large number of studies that employ other cognitive linguistic tasks (see Price, 2000; Démonet et al., 2005; Gandour, 2006; Hagoort, 2005; Stowe et al., 2005; Vigneau et al., 2006; Wong et al., 2004; Zhang and Wang, 2007, for reviews). These studies have revealed that in adults, phonological processing is associated with neural activities in the left hemisphere in inferior frontal, superior temporal, and supramarginal areas (BA44/22/40). In particular, the left superior temporal gyrus is related to the identification and storage of phonological input and phonemic features, whereas the left inferior frontal gyrus appears to handle phonological rules and other speech analysis tasks. By contrast, processing of intonation and tones is associated with acti-vities in the right anterior superior temporal gyrus (BA22), insula, middle frontal gyrus (BA9), and the right inferior frontal gyrus (operculus, BA44). In contrast to the extensive research in intonation/ tone processing, few studies have examined the neural activities associated with the processing of rhythm, except Riecker et al. (2002) who found neural activities in the right superior temporal gyrus and Broca's homologue in the right hemisphere in a syllable sequence generation task (/pa/-/pa/-/pa/).

Unlike the processing of phonological and prosodic information, in adults, lexical semantic processing is associated with widespread brain activities most clearly in the left hemisphere, including the inferior and middle temporal gyri, inferior parietal, inferior prefrontal, and even superior temporal areas (Démonet et al., 1992, 2005; Gold et al., 2006; Gabrieli et al., 1998; Price, 2000). In particular, the left inferior temporal gyrus (BA20) and the left middle temporal gyrus (BA21) are related to the representation of word meaning and conceptual knowledge during lexical and sentence comprehension, whereas the left inferior parietal and angular gyrus (BA39) are related to semantic integration. In addition, the left inferior frontal gyrus (BA45/47) is responsible for the retrieval and selection of word meanings during tasks such as semantic categorization, word association, and word generation (see Gabrieli et al., 1998 and Vigneau et al., 2006 for reviews and meta-analyses).

Most of the above mentioned neuroimaging studies of prelexical cues are based on phonemic or prosodic perception and generation

tasks, for example, by asking listeners to identify whether two phonemes are the same or belong to the same category in an AX or ABX task (Liebenthal et al., 2005) or to judge whether the current and preceding sentences share the same prosodic pattern (Meyer et al., 2002, 2004). Moreover, the brain regions implicated for different linguistic functions are revealed by different studies that employ different processing tasks at different linguistic levels. These studies did not focus on how the various brain regions may be involved in the *same* processing task such as language discrimination, and how cortical regions may be competing and interacting with each other. Our study attempts to fill these gaps, that is, to understand the neural mechanisms underlying the use of various types of cues for language discrimination, including prelexical (prosodic and phonological) and lexical (semantic) information. One important goal of the study is to identify how the various cues at different linguistics levels compete and interact in the adult brain during the discrimination of languages. We wanted to achieve this goal by manipulating four levels of prelexical and lexical information in a uniform language discrimination task in which behavioral and fMRI data are collected. In particular, we examine language discrimination under four hierarchically organized conditions: (1) only rhythmic information is available, where listeners receive resynthesized speech in which the consonants and vowels of Chinese and English sentences are replaced by /s/ and /a/, respectively, and the intonation of the sentence is flattened; (2) rhythmic plus intonation information is available, where listeners receive the same resynthesized speech with intonation; (3) listeners hear normal sentences from Italian and Japanese, two unfamiliar languages that have different prosodic and phonological structures; and (4) listeners hear normal sentences from Chinese (native language) and English (second language), familiar languages that include lexical semantic information as well as prosodic and phonological information from (1)–(3). We compare the accuracy and response times of discrimination and the BOLD signals elicited from native Chinese listeners under these four stimulus conditions.

Methods

Participants

Eighteen participants (twelve females) took part in this study. They were between 20 and 26 years of age (mean age: 22.3). All participants were native speakers of Mandarin Chinese. They reported no speech or hearing problems, and no history of neurological or psychiatric abnormalities. They all learned English as their second language since middle school (around 12 to 13 years of age), but did not learn any Italian or Japanese. All had passed the CET (College English Test) at Level 4 (highest level is 6), a standardized English test adopted by the Chinese Ministry of Education, which includes a listening component. The participants were all right-handed according to a modified Chinese version of the Edinburgh Handedness Inventory (Oldfield, 1971). They were asked to complete the Informed Consent form approved by Beijing Normal University before the scanning, and were paid for their participation in the study.

Materials

Four types of stimuli were used, each in a different test condition: normal sentences from Chinese and English (familiar languages), normal sentences from Italian and Japanese (unfamiliar languages), sentences with intonation and syllabic rhythm (resynthesized from the Chinese and English sentences; see details below), and sentences with only syllabic rhythm (resynthesized from the Chinese and English sentences). The sentences of the four languages (Chinese, English, Italian, and Japanese) were short news-like declarative statements originally in English, and were adapted from those used by Nazzi et al. (1998) and Ramus and Mehler (1999). Some words were deleted and others added so that the number of syllables in each sentence is exactly fifteen. The English sentences were then translated into the other three languages, and the number of syllables across sentences in the other languages was kept on average at fifteen (see Appendix). Sentences were read aloud in a sound attenuated room by female native speakers of each language, four speakers per language, and five sentences per speaker. The sentences were recorded at a sampling rate of 44.1 kHz and digitized at 16 kHz. Average duration of each sentence was about 2.7 s, with 2660 ms (SD=80) for English, 2681 ms (SD=80) for Italian, 2695 ms (SD=55) for Chinese, and 2676 ms (SD=55) for Japanese. The sentence materials were matched in terms of average intensity (70 dB) and average pitch (205 Hz) using the Praat software (http://www.fon.hum.uva.nl/praat/).

To manipulate the test materials so that the acoustic cues at the two prosodic levels (rhythm and intonation) can be separately or jointly preserved in the stimuli, we used the following procedure to resynthesize the Chinese and English sentences. Our procedure followed closely those used by previous studies of language discrimination (e.g., Ramus and Mehler, 1999).

(1) Each phoneme was marked as a consonant or a vowel;
(2) The fundamental frequency (F0) was extracted using the Bliss software (http://www.mertus.org/Bliss/index.html);
(3) The duration and F0 value of each phoneme were fed into the MBROLA software (Dutoit et al., 1996) to perform the synthesis, and the synthesized phonemes were based on the pronunciation of a male French speaker from the diaphone database of MBROLA. By this resynthesis process, each consonant in the original Chinese and English source sentences was replaced with /s/ and each vowel with /a/, and the F0 values of the phonemes as well as the syllabic tones and intonations in the sentence were preserved.
(4) For the rhythm-only stimuli, the intonation of the source sentences was flattened, by which all sentences were resynthesized with a constant F0 at 205.7 Hz, which was the approximate mean F0 value of the Chinese source sentences.

Procedure

Participants performed a forced-choice AX discrimination task, in which they judged whether sentence X came from the same language of sentence A. The inter-stimulus interval (ISI) between A and X was 1200 ms. A pure tone of 1 kHz lasting for 100 ms was added before each trial to signal that the next two sentences were to form the AX discrimination pair in the trial. The interval between the tone and sentence A was also 100 ms. Thus, a trial consisted of a pure tone, the two sentences with ISI, and an inter-trial interval. A variable TOA design was used, in which stimuli were presented at different trial onset asynchronies (TOA=10, 12, 14, 16 s; mean was 11 s). Presentation of the stimuli was controlled by the E-Prime 1.1 software (MacWhinney et al., 2001; Psychology Software Tools Inc., 2005).

Each participant went through four sessions of the experiment, with each session including one of the four types of stimuli. Three different orders of presentation were derived so that the order in which participants received the stimulus types could be counterbalanced. Each session contained 40 trials (40 sentence pairs, 80 sentences), in which 20 were the 'same' trials and 20 were 'different' trials. The order of presentation of the 40 trials was randomized within each session. Participants were instructed to respond as quickly and as accurately as they could when they heard sentence X. Both discrimination accuracy and reaction times (RTs) were recorded for each participant. RTs were measured from the onset of X. Before formal scanning, subjects practiced outside the scanner on a subset of test sentences not used for the given subjects to become familiar with the experimental procedure.
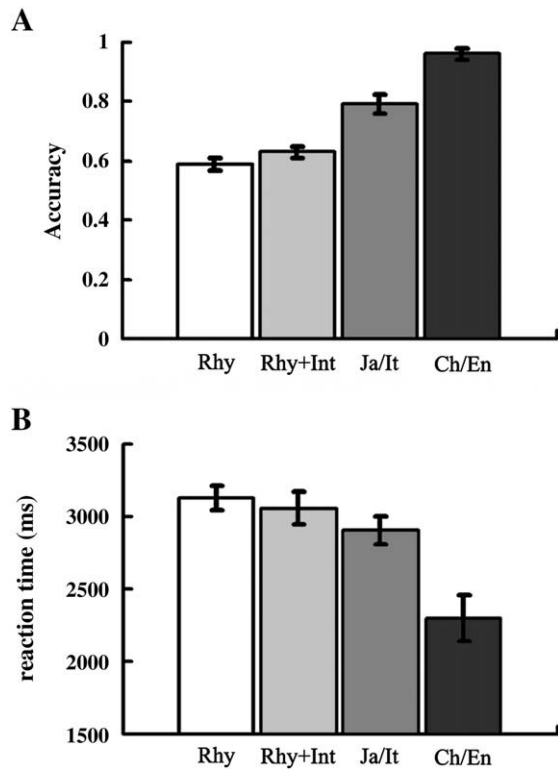
Fig. 1. Discrimination accuracy (A) and reaction times (B) under the four language discrimination conditions (rhy = rhythm; rhy+int = rhythm+intonation; Ja/It = Japanese/Italian; Ch/En = Chinese/English; same in Figs. 2 and 5). Error bars indicate standard error of the mean.

Image acquisition

Images were acquired on a 3-Tesla GE Signa scanner equipped with an fMRI Silent Scan Audio System (SS3100, AVOTEC) in the Southwestern China Medical School, Sichuan University. Participants heard the spoken stimuli on both ears with a comfortable, equal, volume. Each participant's head was aligned to the center of the magnetic field. They were instructed to relax, keep their eyes closed, and refrain from moving their heads.

A gradient-echo, echo planar imaging (EPI) sequence was used to acquire the fMRI data. The functional images were reconstructed from 26 axial slices, with the following parameters: thickness=4 mm, FOV=240 mm^2, matrix=64×64, TR/TE/FA=2000 ms/30 ms/90°. High-resolution anatomical images (156 slices) of the entire brain were obtained by a 3-D gradient-echo sequence (SPGR) after the functional images were acquired, with the following parameters: matrix=320×256, thickness=1 mm, TR/TE/FA=8.46 ms/3.336 ms/25°. A total of 236 volumes was acquired for each session, and the session lasted about 8 min. The entire experiment with four sessions took about 40 min, including about 6 min of anatomical scanning.

Data analysis

The imaging data were preprocessed and statistically analyzed with AFNI (Cox, 1996). The first seven scans were excluded from data processing to minimize the transit effects of hemodynamic responses. Functional images were corrected for head motion by aligning all volumes to the eighth volume using a six-parameter rigid-body transformation (Cox and Jesmanowicz, 1999). Statistical maps were spatially smoothed with a 6-mm FWHM Gaussian kernel.

The preprocessed images were submitted to individual GLM-based analyses to estimate the individual statistical *t*-maps for each condition. Voxel-wise scaling coefficients were output at 2-s interval ranging from 2 to 14 s poststimulus, and were then converted to percent signal change by dividing the coefficient from each voxel by the mean for that voxel across the entire experiment. Individual anatomical images and *t*-maps were both coregistered to the standard Talairach and Tournoux (1988) space. All images were resampled to 3×3×3 mm^3 voxels. In a random effects analysis, Talairach-warped maps of percent signal change (mean of 4 to 10 s) for each stimulus type and each participant were entered into a two-way, mixed-factor ANOVA with stimulus type as fixed factor and participant as random factor.
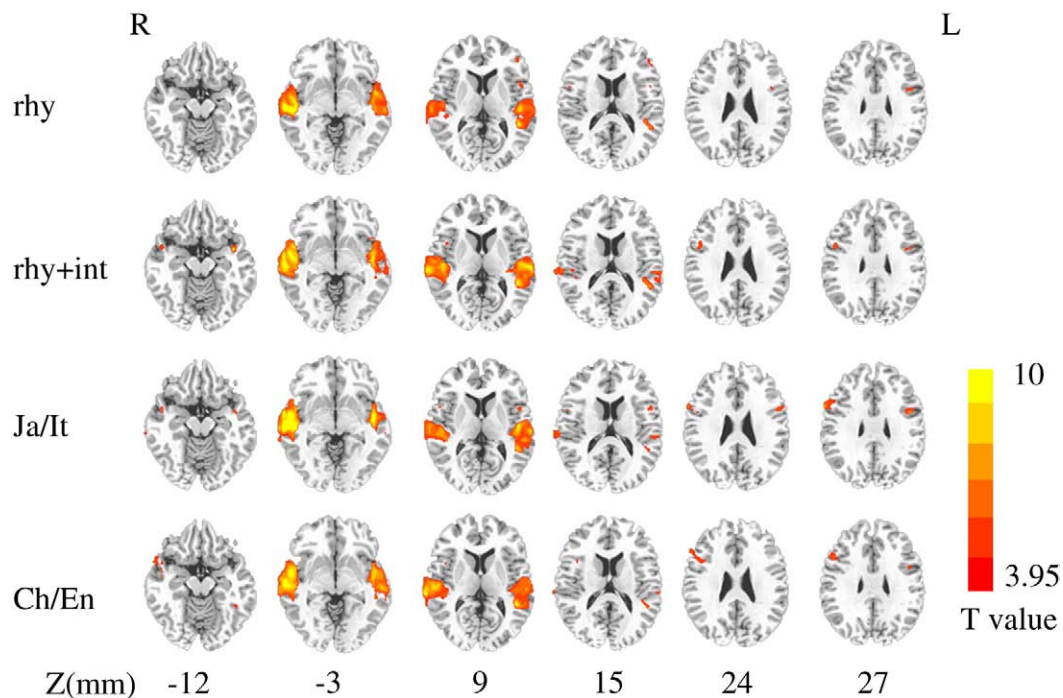

Fig. 2. Group activation maps for the four stimulus conditions, thresholded at a voxel-wise *p*<0.001 (*t*>3.95), uncorrected. Cluster level activated volume ≥270 mm^3. Axial slices at z=−12, −3, 9, 15, 24, 27, left to right.

A group statistical map was created by contrasting individual *t*-maps against a constant value of 0 for each condition. The group maps were thresholded at a voxel level of $t > 3.95$ ($p < 0.001$) with an automated cluster detection through AFNI (minimum volume = 270 mm³). The spatial extent correction for planned comparisons corresponds to an overall $\alpha < 0.05$ of FWE rate, as calculated with AlphaSim (http://afni.nimh.nih.gov), for all intracranial voxels in the imaged volume (the corresponding correction for data thresholded at $p < 0.05$ was 378 mm³). We focused on areas of the temporal lobe, frontal lobe, inferior parietal lobe, supramarginal gyrus, angular gyrus in our analyses. Brain regions of cerebellum, occipital cortex, primary motor-sensory cortex, and supplemental motor areas were not reported, as these regions are shown in previous studies to be uncorrelated with speech processing. The inferior frontal gyrus (opercularis), which was more active for the Japanese/Italian stimuli, and the inferior and middle temporal gyri, which were more active for the Chinese/English stimuli (see Results section), were chosen as regions of interest (ROIs) in our analysis. We selected the peak activation coordinates from the cluster of the contrast analysis as the center of each ROI. The two ROIs are uniform in size, with a sphere of 7 mm in radius. Time course of the hemodynamic responses from 2 to 14 s within the ROI was generated. Time-series plots were constructed for each participant first and then averaged to create a group time-series graph.

Results

Behavioral results

Fig. 1 presents the discrimination accuracy (A) and reaction times (RT, B) for all stimulus types. A one-way repeated-measures ANOVA showed a significant main effect for stimulus type, on accuracy [$F(3, 51) = 73.466$, $p < 0.01$], and on RT [$F(3, 51) = 28.252$, $p < 0.01$]. Post hoc Bonferroni tests revealed significant differences in accuracy between all comparison pairs except between the rhythm-only sentences and the rhythm + intonation sentences. The results indicate that listeners' discrimination accuracy increased with the type and amount of information available, with rhythm plus intonation eliciting better discrimination than rhythm alone (though not statistically significant), natural languages (Italian/Japanese) eliciting better discrimination than resynthesized speech that contains only prosodic information, and familiar languages (Chinese/English) eliciting better discrimination than unfamiliar languages (Italian/Japanese). Post hoc Bonferroni tests also revealed significant differences in RT between the Chinese/English sentences and the other three sentence types (but no differences within the other sentence types), indicating that listeners are faster at discriminating between familiar languages than at discriminating between unfamiliar languages or between speech signals that contain only prosodic information.

Table 1
Areas of significant activation in planned comparisons, thresholded at voxel level $p < 0.05$, cluster level activated volume ≥ 378 mm³

Anatomical region	Brodmann areas	Activated volume (mm³)	Maximum intensity[a]			Local maximum (*t* value)
			x	*y*	*z*	
Rhythm + intonation > Rhythm						
Right STG	22	6615	50	−8	−1	4.334
Left STG	41	5670	−53	−17	6	4.898
Left STG	22	2214	−53	−41	15	3.650
Right STG	42	1188	68	−29	12	3.306
Right STS	22/21	486	68	−23	−1	2.864
Left IFG (oper)	44	378	−56	11	9	2.836
Japanese/Italian > Rhythm + intonation						
Right IFG (oper)	9	1809	44	5	33	3.325
Left STG	22	1674	−50	−14	3	4.141
Right MTG	21	1566	56	−20	−7	4.049
Left IFG	9/44/45	945	−35	14	24	2.930
Right MTG	21	918	68	−29	−10	3.312
Chinese/English > Japanese/Italian						
Left ITG	20	2862	−47	−38	−10	5.008
Right ITG	20	810	62	−47	−13	3.027
Left ITG	20	513	−41	−14	−25	2.941
Japanese/Italian > Chinese/English						
Left STG	22	1242	−47	−14	9	4.035
Left IFG (oper)	44	594	−44	5	12	3.316
Right IFG (oper)	9	567	59	8	24	4.901

[a] Note: coordinates are in Talairach and Tournoux (1988) space, where the voxel with the maximum intensity for the cluster lies.

Imaging results

Group activation maps

Fig. 2 presents the group activation maps, summarizing activation clusters for each of the four stimulus conditions during language discrimination. The cluster level activated volume was set at 270 mm³ or greater. BOLD activations during the language discrimination task were generally widespread, but tended to concentrate on the temporal cortex (centering in the peri-sylvian and midline regions). Other areas of activation included the frontal region, the primary motor-sensory cortex, the insula, and the cerebellum. Activations in the temporal lobe were bilateral, and showed no clear lateralization differences.

Table 1 presents a summary of all activation clusters (except cerebellum, occipital lobe, primary motor-sensory cortex, and supplemental motor areas) in the planned comparisons, which were all significant at $p < 0.05$. Below we focus on the planned comparisons between the following stimulus types: (1) rhythm + intonation versus rhythm, (2) Japanese/Italian versus rhythm + intonation, and (3) Chinese/English versus Japanese/Italian.
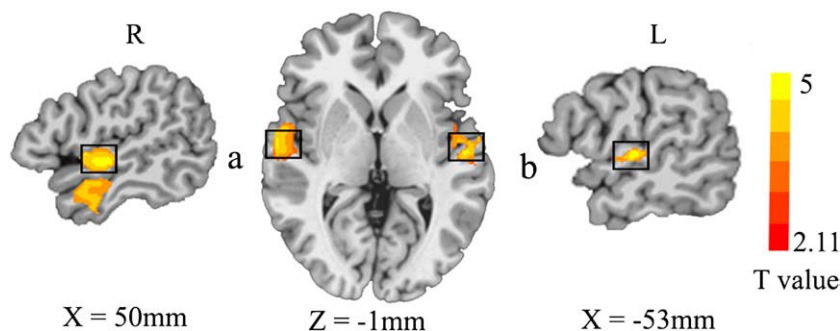

Fig. 3. Activation maps for the contrast between rhythm + intonation versus rhythm-only sentences, thresholded at a voxel-wise $p < 0.05$ ($t > 2.11$), uncorrected. Cluster level activated volume ≥ 378 mm³. (a) Right STG (50, −8, −1); (b) left STG (−53, −17, 6), including activation in the left temporal pole.
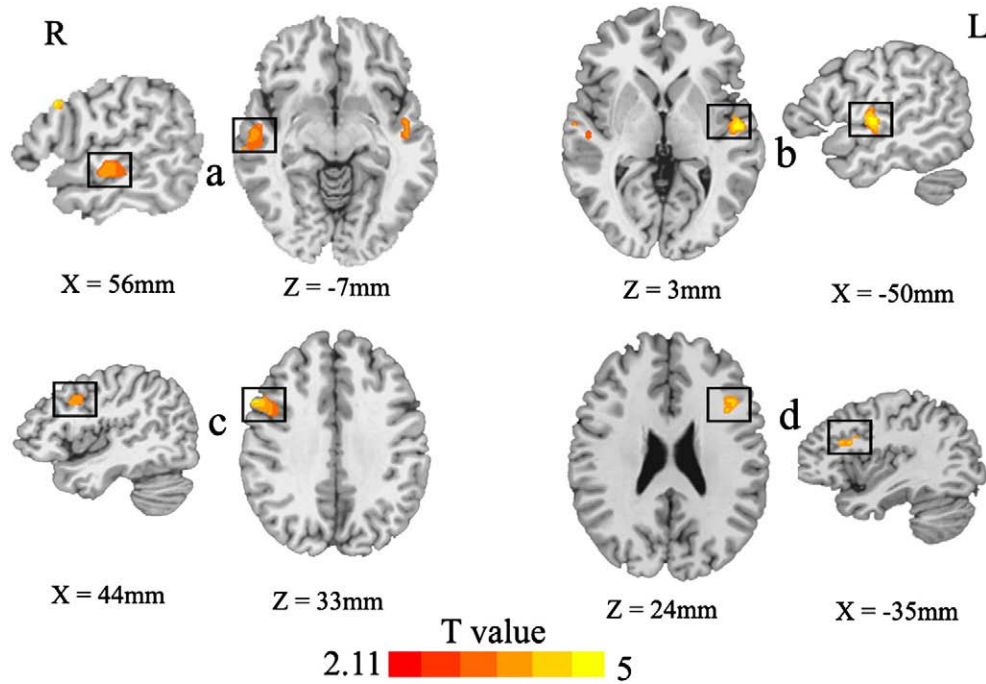
Fig. 4. Activation maps for the contrast between Japanese/Italian versus rhythm+intonation sentences, thresholded at a voxel-wise $p<0.05$ ($t>2.11$), uncorrected. Cluster level activated volume ≥ 378 mm^3. (a) Right MTG (56, −20, −7); (b) left STG (−50, −14, 3); (c) right IFG (44, 5, 33); (d) left IFG (−35, 14, 24).

Rhythm+intonation versus rhythm

Whether rhythm+intonation cues elicited more areas of activation than rhythm alone was investigated by contrasting rhythm versus rhythm+intonation stimulus conditions. No areas of activation were found for the rhythm-only condition, but significantly more areas of activation were found for the rhythm+intonation condition in this comparison, as shown in Fig. 3. The areas that were activated more for the rhythm+intonation condition were in the right superior temporal gyrus (BA22/42) and the left superior temporal gyrus (BA41/22). These additional areas of neural activation are consistent with findings in the literature regarding their role in the processing of intonation (Brown et al., 2004; Gandour et al., 2003, 2004; Hesling et al., 2005; Meyer et al., 2002, 2004; Plante et al., 2002). The largest of these activations (6615 mm^3) was in the right superior

temporal gyrus, consistent with the role of the STG in the right hemisphere for the processing of intonational/tonal information for novel languages (Gandour et al., 2003, 2004).

Japanese/Italian versus rhythm+intonation

As shown in Fig. 4, more areas of activation for the Japanese/Italian stimuli compared with the rhythm+intonation stimuli were found in the left inferior frontal gyrus (BA9/44/45), left superior temporal gyrus (BA22), right middle temporal gyrus (BA21), and right inferior frontal gyrus (BA9). The activation of the inferior frontal gyrus indicates that more phonological analysis was involved in the processing of natural though unfamiliar languages, compared with the processing of synthesized speech that had only prosodic information. No areas
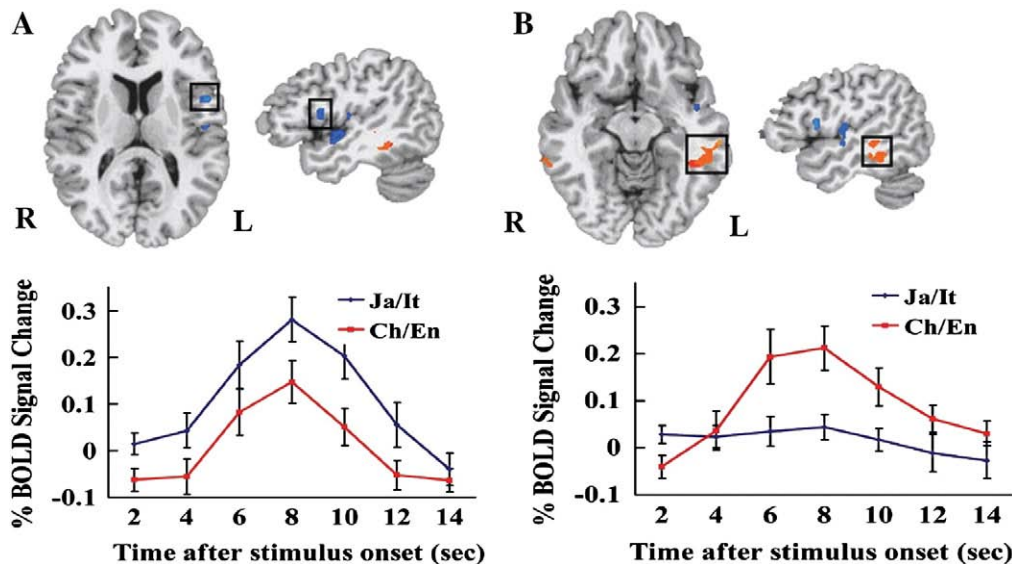

Fig. 5. Clusters with significant activation differences and their time courses for the Japanese/Italian and Chinese/English stimuli ($p<0.05$, uncorrected). Cluster level activated volume ≥ 378 mm^3. (A) The cluster of left IFG (oper) was more active in the Japanese/Italian condition. Time course of activation for that cluster centered at (−44, 5, 12), shown in the lower left panel. (B) The cluster for left ITG was more active in the Chinese/English condition. Time course of activation for that cluster centered at (−47, −38, −10), shown in the lower right panel. Error bars indicate standard error of the mean.

Table 2
Areas of significant activation in a post-hoc comparison, thresholded at voxel level $p < 0.05$, cluster level activated volume ≥ 378 mm³

Anatomical region	Brodmann areas	Activated volume (mm³)	Maximum intensity[a]			Local maximum (t value)
			x	y	z	
Rhythm+intonation>Chinese/English						
Left MTG/STG/SMG	21/22/40	999	−47	−47	18	4.059
Right IFG (oper)	44/45	567	59	11	18	3.455
Left IFG (obi)/insula	47/13	540	−26	20	−7	3.204
Right IFG (tri)/insula	45/13	432	35	29	15	3.683
Left MFG/IFG (tri)	9/46	378	−41	32	27	3.777
Chinese/English>Rhythm+intonation						
Left MTG/ITG	20/21	2565	−47	−41	−4	4.067
Right ITG/MTG	20/21	1134	56	−44	−7	3.161
Right STG/MTG	22/21	945	59	−5	−7	3.953
Left IFG (tri)	45	405	−38	17	24	3.147

[a] Note: coordinates are in Talairach and Tournoux (1988) space, where the voxel with the maximum intensity for the cluster lies.

were found to be more activated for the rhythm+intonation stimuli than for the Japanese/Italian sentences.

Chinese/English versus Japanese/Italian

Six areas showed significant differences between the Chinese/English stimuli and the Japanese/Italian stimuli, with three areas showing more activation for the Chinese/English sentences than the Japanese/Italian sentences and three areas showing the reverse. The largest activation for the Chinese/English condition compared with the Japanese/Italian condition was centered in the inferior temporal gyrus (BA20), with more activation in the left hemisphere than in the right hemisphere. The areas that showed more activation for the Japanese/Italian sentences than for the Chinese/English sentences were the left superior temporal gyrus (BA22), left inferior frontal gyrus (opercularis, BA44), and right inferior frontal gyrus (opercularis, BA9). As discussed in the Introduction, the left inferior or middle temporal gyrus is related to semantic representation and verbal knowledge, whereas the left inferior frontal gyrus and the superior temporal gyrus are associated with phonological and/or prosodic processing. The contrast between the Chinese/English sentences (familiar languages) and the Japanese/Italian sentences (unfamiliar languages) shows that the different regions relevant to the processing task were activated for discriminating the relevant kinds of languages. Moreover, when lexical semantic cues are available and semantic analysis is possible, as in the Chinese/English case, the brain areas associated with phonological or prosodic processing become less activated, as compared with the situation (e.g., unfamiliar languages) in which language discrimination has to rely only on phonological or prosodic cues.

ROI analyses

The different activation patterns for the Chinese/English condition versus the Japanese/Italian condition led us to a Region-of-Interest (ROI) analysis. Given the role of the left inferior temporal gyrus (BA20) and the left inferior frontal gyrus (opercularis, BA44) in speech and language processing, we set these two regions as our ROIs and generated the time-course data for them.

A repeated-measures within-subject ANOVA was run with the mean cluster-wise percent signal change as the dependent variable, and time point and stimulus type as independent variables. As shown in Fig. 5, this analysis revealed a significant main effect of stimulus type in the left inferior frontal gyrus [$F(1, 17) = 7.028$, $p < 0.05$] and the left inferior temporal gyrus [$F(1, 17) = 6.93$, $p < 0.05$]. Fig. 5 also shows an interaction between brain region and stimulus type, and this was confirmed in a second ANOVA in which time point, stimulus type, and brain region were treated as independent variables. The interaction of brain region with stimulus type [$F(1, 17) = 32.498$, $p < 0.01$] was due to

that in the left IFG, the Japanese/Italian sentences elicited more peak activation than the Chinese/English sentences whereas in the left ITG, the reverse was true.[1]

Discussion

How do human speakers differentiate one language from another? This question arises not only when we are faced with familiar languages (discriminating our native language from non-native languages) but also when we are trying to distinguish between different dialects or unfamiliarly sounding languages. Previous studies have identified the role of prosodic and phonological cues in this process, but little research has been done to examine the interaction between these cues and how human language discrimination relies on lexical semantic knowledge in real time. In addition, few studies have investigated the neural mechanisms underlying the discrimination of familiar and unfamiliar languages. Here we attempt to fill these gaps by identifying the nature of the interaction between prosodic cues, phonological cues, and lexical knowledge and the neural correlates of this interaction.

In this study we have manipulated four types of 'linguistic' stimuli (resynthesized speech and familiar or unfamiliar natural languages) in an fMRI study. In the resynthesized speech conditions, listeners received only rhythmic information or only rhythmic plus intonation information of the sentences, and they have to rely on these prosodic features to discriminate between the relevant 'languages'. In the unfamiliar language condition, they received Italian or Japanese sentences, stimuli from foreign languages with which they have no prior experience, and they would need to use phonological cues such as phonetic repertoire or phonotactic rules, in addition to prosodic cues, to figure out the differences between the languages. Finally, in the familiar language condition, listeners received Chinese and English sentences, stimuli from their native language or second language that they have had extensive experiences with, and they could use lexical semantic knowledge of the words and sentences, in addition to prosodic and phonological cues, to distinguish between these languages. These conditions progressively include more information for the listeners to use, and each higher level encompasses all the cues at a lower level, forming a hierarchically organized structure that listeners can explore at different degrees of complexity for language discrimination.

Our study reveals that as the complexity of the available information increases, the discrimination accuracy, not surprisingly, also increases. Moreover, depending on the cues available, the listener recruits different brain regions, independently or jointly, to handle the processing task. For example, when only lower-level acoustic cues are available, the brain regions most relevant to the processing of prosody become active, in particular, the superior temporal gyrus and the temporal pole (Fig. 3); when phonological cues are available in addition to prosodic cues, the brain regions associated with phonological processing become highly active, in particular, the inferior frontal gyrus and the left superior temporal gyrus, in addition to areas that are responsible for prosodic processing (Fig. 4). Finally, when lexical semantic cues are available, listeners rely more exclusively on brain regions that are related to semantic processing, for example, the inferior temporal gyrus, to discriminate between familiar languages (Fig. 5).

Most important of these patterns is that the brain activations do not simply multiply monotonically as a function of the complexity of the information involved, but depend rather on the validity of the relevant cues during cue competition for language discrimination. In fact, the use of lexical semantic knowledge for the discrimination of familiar languages (Chinese and English) involves fewer brain regions

---

[1] Note that although we speak mainly of the left ITG in the interaction, the peak activation involves a wide area that borders the MTG area (BA21) as well.

than the use of phonological and prosodic cues for the discrimination of unfamiliar languages (Italian and Japanese). This pattern is consistent with findings in the literature that learners can automatically access and use lexical semantic knowledge for language discrimination (Leavers and Burley, 2001; Lorch and Meara, 1995), in a manner in which meanings of words directly impinge on processing whenever such information is available.

Patterns of cortical competition discussed above led us to further consider the theoretical implications of our results. More than 25 years ago Bates and MacWhinney (1982) proposed the Competition Model, according to which different linguistic (grammatical, semantic, and pragmatic) cues compete with each other in language acquisition and language processing (see later formulations in Bates and MacWhinney, 1987, 1989). Although the original model had been applied mainly to explain sentence comprehension and production with respect to the assignment of thematic roles ('who does what to whom'), the model has also been used to examine other linguistic domains (MacWhinney, 1989, 2004). In particular, two critical constructs of the model, cue validity — how valid the relevant cue is in the given language for predicting the agenthood or objecthood, and cue reliability — how reliable the relevant cue can be used in the prediction, have clear implications here for our understanding of how prosodic cues, phonological cues, and lexical semantic knowledge compete in the discrimination of languages. For example, lexical tones are highly valid cues for the discrimination of Chinese versus other languages, while phonotactic patterns are reliable cues for the discrimination of English versus other languages.

Cue competition occurs at various linguistic levels. One general finding from our study is that higher-level cues lead to better discrimination accuracy and faster discrimination speed than lower-level cues. In the terms of the Competition Model, the cues that have higher validities and reliabilities will lead to better and faster language comprehension (Bates and MacWhinney, 1989; Li et al., 1993). Lexical semantic knowledge is the most valid and reliable cue for discriminating familiar languages, and if the listener can make use of this knowledge, discrimination will be most successful. With regard to cortical competition, lexical knowledge clearly plays a more dominant role. While we see some degree of cortical competition with prelexical cues (e.g., between prosodic and phonological cues), their effects for language discrimination are for the most part additive (from prosody to phonology) with respect to the magnitude of cortical activation. By contrast, the competition between lexical and prelexical cues elicits stronger cortical competition: when both lexical and prelexical cues are available for language discrimination, the neural region associated with lexical semantic processing (e.g., inferior temporal gyrus) is more strongly activated and at the same time, the regions associated with prelexical processing become correspondingly less activated. In other words, higher-level cues are more favorably selected in the competition process, as the competitive outcome leads to successful language discrimination more effectively and reliably.

This cortical competition view, while consistent with existing data from our behavioral and fMRI analyses, may be subject to an alternative explanation.[2] One possibility is that the familiarity of the test stimuli could lead to different neural patterns of activation. For example, the greater activation in the inferior frontal gyrus in the Japanese/Italian condition, as compared with the Chinese/English condition (Fig. 5), might be due to the more controlled, effortful processing of the Japanese and Italian sentence materials because these materials were unfamiliar to the participants. Familiarity could be one important variable in lexical and sentence processing, as previously demonstrated by researchers in other contexts. In the current context, however, we must note that familiarity inherently involves two types of effect: familiarity with the form of a word (due to its lexicality, phonological and phonotactic constraints), and familiarity with the meaning of a word (Cordier and Le, 2005). Our results indicate that both types of familiarity effect may be important for language discrimination. For example, the different activation patterns for the Chinese/English versus Japanese/Italian conditions are probably due to both familiarity with form and familiarity with meaning, since the Japanese and Italian, in contrast to the Chinese and English, sentences convey no semantic information and contain unfamiliar phonological forms for the listener. To further distinguish effects of familiarity from semantic processing in our data, we have additionally compared the activation patterns for the Chinese/English sentences with those obtained in the rhythm+intonation condition (Table 2), showing that our results are best accounted for by the competitive interactions between semantic, phonological, and prosodic cues during processing.[3]

The strong reliance on lexical semantic knowledge does not mean that other cues are not used for language discrimination. It simply shows that the listener distributes more cognitive resources to the analysis of those cues that are highly valid and more reliable for the processing task at hand. In fact, in the case of Chinese/English language discrimination where lexical semantic processing is dominant, we could also see activations in other areas such as the left middle temporal gyrus, the right superior temporal sulcus and middle temporal gyrus.[4] These other areas are implicated in the processing of prosodic information, and most likely in this case, the processing of tonal information in Chinese (see also the contrast analysis above between rhythm+intonation and rhythm-only conditions).[5] It is also clear that the processing of higher-level cues (lexical semantic) cannot proceed alone by itself and must also rely on lower-level, prelexical cues and the interaction between lexical and prelexical cues. Fig. 5A shows clearly that even when the listener is able to make use of lexical semantic information in the processing of Chinese/English materials, the inferior frontal gyrus remains active, only that it is less active than in the situation in which the listener has no semantic information to rely on (e.g., in the processing of Japanese/Italian materials). After all, when we listen to our first and second languages, semantic processing and integration must rely on the interactive, and often reciprocal processes that occur between higher-level and lower-level information sources. As a result, the brain must also utilize, often in parallel, areas that are responsible for the processing of semantic, phonological, and prosodic cues in the language being analyzed.

Language discrimination occurs very early, in newborns and infants, as discussed earlier. Infants and young children rely on prelexical cues to do so, whereas adults rely on a variety of prelexical

---

[2] We are grateful to an anonymous reviewer for pointing out this alternative explanation.

[3] In this separate comparison we found that the superior temporal and inferior frontal areas – areas that are responsible for prosodic and phonological processing – show reduced activations when semantic information is available (i.e., in the processing of Chinese and English sentences), as compared with the processing of materials in which only prosodic information (i.e., rhythm and intonation) is available (see Table 2 for details). This is strong evidence that familiarity alone cannot explain the cortical patterns in our data.

[4] These activation patterns are not shown in the contrast analysis reported here. Interested readers can request these data from the corresponding authors via email.

[5] The rhythm+intonation versus rhythm-only stimuli did not differ significantly in accuracy or speed of discrimination, although the former involved more cortical regions in the processing. It is somewhat surprising, though, that tonal information from Chinese has not played an even stronger role in the language discrimination task, given the significance of tones that has been demonstrated for language perception in Chinese (e.g., Gandour, 2006; Wang, 2002). This might be due to the nature of the synthesized speech in the rhythm+intonation condition, in which the monotonic syllabic structure (/sa/-/sa/-/sa/) provides no lexical significance to the items being processed. The activation of the right instead of the left superior temporal gyrus also indicates that tones and intonation are processed similarly by our listeners in the same 'non-linguistic', 'non-native' way, according to the formulations of Gandour (2006), Wang et al. (2003), and Zhang and Wang (2007).

and more importantly, lexical cues for the task. Our study shows that adults recruit complimentary and competing neural structures depending on the amount and the kind of the acoustic and linguistic cues available. Although we do not yet know when children start to treat lexical semantic knowledge as seriously and as reliably as adults do for language discrimination, our study suggests that adults do not use the prelexical, particularly the prosodic, cues as fully and as effectively as infants and children do, partly because adults already have in place a large knowledge base of lexical semantic information.

This child–adult difference in the reliance on prelexical cues versus lexical cues, in our view, has significant implications for understanding why language acquisition in adulthood is fundamentally different from that in infancy or young childhood. In infancy through young childhood, the learner builds up language from the ground up, acquiring and using lower-level information such as prosodic and phonetic cues first, segmental and phonological cues second, and lexical semantic information last. In contrast, in adulthood the learner faces the language task through exposure and experiences with prelexical and lexical, suprasegmental and segmental, and syntactic and semantic cues all at once. This view is consistent also with explanations of the age of acquisition effect discussed in the literature: early learned words have a strong auditory basis, drawing on the learner's phonetic and phonological knowledge, whereas late learned words involve more effortful processes such as semantic retrieval (see Hernandez and Li, 2007 for review). Our findings here suggest that cortical competition might take very different forms in infants, children, and adults for language discrimination, and consequently, in the learning of a new language.

Acknowledgments

Appendix. Natural language materials used for the language discrimination task

English stimuli

The next local elections will take place during the winter.
A hurricane was announced this afternoon on the TV.
The candidates will meet this weekend for a special debate.
This season promises to be a very exciting one.
Writers have been attracted by the life in the capital.
My grandparents' neighbour is the most charming person I know.
The art gallery in this street was opened only last week.
Most local stores close very early on Sunday afternoons.
Having a big car is not something that I would recommend.
The council has decided to renovate the old center.
Much more money will be needed to make this project successful.
The supermarket was closed due to economic problems.
The rebuilding of the town started quickly after the flood.
There is a market twice a week on the square of the village.
Getting the job is difficult without a doctor's degree.
The local train left the station more than five minutes ago.
The government is planning a reform of the tax system.
The boy got up quite early in order to watch the sunrise.

The recent rainfall has caused severe damage in the valleys.
The man is a famous specialist in plastic surgery.

Chinese stimuli

Xia yi ci difang xuanju hui zai jinnian dongji juxing.
Dianshitai xiawu fabu le taifeng dao lai de jingbao.
Houxuanren zhoumo jiang hui juxing yi ci tebie bianlun.
Zhe hui shi ge rang ren ganjue shifen xingfeng de saiji.
Zhe xie zuojiamen bei shoudu de shenghuo xiyin zhu le.
Wo zufumu de linjun shi zu juyou meili de ren.
Zhe tiao jiedao shang de meishu hualang shangzhou cai kaiye.
Dangdi daduoshu shangdian zhouri xiawu guanmen hen zao.
Wo bing bu shi jianyi dajia dou goumai da de qiche.
Lishihui yijing jueding zhengxiu jiu de huodong zhongxin.
Xu yao touru geng duo de qian lai shi gai xiangmu chenggong.
Jingji fangmian de yuanyin zhishi chaoshi beipo guanbi.
Hongzai guo hou hen kuai jiu kaizhan le chengshi de chongjian.
Xiangcun de guangchang shang mei zhou dou hui you liang ci jishi.
Meiyou boshi xueli jiang hen nan de dao zhe fen zhongzuo.
Dangdi de huoche wu fen duo zhong qian kai chu le chezhan.
Zhengfu zhengzai jihua dui shuishou zhidu jinxing gaige.
Na ge nanhai weile kan richu er qi de feichang zao.
Jiang yu gei shangu diqu dai lai le juda de sunshi.
Na ge ren shi feichang zhuming de zhengxing shoushu zhuanjia.

Italian stimuli

Le elezioni locali si terranno in inverno.
Il T.G. ha annunciato l'arrivo del tifone.
I candidati si vedranno per un dibattito.
Questa musica ti emozionera' tantissimo.
La capitale affascina molto gli scrittori.
Il vicino dei mie nonni e' molto attraente.
La galleria d'arte era aperta gia' da giorni.
Sabato i negozi locali chiudono presto.
Io non disapprovo chi compra le macchine grandi.
Il consiglio rinnovera' il centro ricreativo
Ci occorrono molti soldi per questo progetto.
Il supermercato chiuse per gravi problemi.
La piena causo' la ricostruzione della citta'.
In citta'si fa il mercato ogni settimana.
Per ottenere quel lavoro ci vuole una laurea..
Il treno e' partito da qui cinque minuti fa.
Il governo sta' pianificando una riforma.
Per vedere l'alba quel ragazzo si alzo' presto.
Le pioggie portarono molti danni nelle valle.
Quell'uomo e' un famoso chirurgo plastico.

Japanese stimuli

Jikai no senkyo ha fuyu ni okonawareru.
Gogo no terebi ni yoruto taihu ga kuru.
Rikkouhosha ha shuumatu touronkai wo okonau.
Kore ha totemo omoshiroi shiai dearu.
Sakka ha toshi no seikatu ni miserareta.
Sofubo no rinjin ha miryokuteki dearu.
Ano garou ha kaigyou shitabakari dearu.
Soko ha nichiyou gogo hayaku heiten suru.
Oogatasha kounyuu wo teian suru nodehanai.
Rijikai ha shisetu no tatekae wo kimeta.
Kore no seikou niha sarani kane ga iru.
Keizai mondai de supa ha heiten wo shiirareru.
Kouzuigo suguni machi ga tatenaosareta.
Nouson de shuu nikai ichiba ga hirakareru.
Hakasegou ga naito konoshoku ha e nikui.

Genchi no kisha ha gofunmae ni shuppatu shita.
Seifu ha zeiseikaikaku wo keikakuchuu desu.
Anoko ha hayaokisite hinode o mita.
Ame de teichi ga kanari higai wo uketa.
Ano hito ha yuumeina seikeigekai dearu.

References

Bates, E., MacWhinney, B., 1982. Functionalist approaches to grammar. In: Wanner, E., Gleitman, L. (Eds.), Child Language: the State of the Art. Cambridge University Press, New York, pp. 173–218.

Bates, E., MacWhinney, B., 1987. Competition, variation and language learning. In: MacWhinney, B. (Ed.), Mechanisms of Language Acquisition. Erlbaum, Hillsdale, NJ, pp. 157–194.

Bates, E., MacWhinney, B., 1989. Functionalism and the Competition Model. In: MacWhinney, B., Bates, E. (Eds.), The Crosslinguistic Study of Sentence Processing. Cambridge University Press, New York, pp. 3–76.

Bosch, L., Sebastián-Gallés, N., 1997. Native language recognition abilities in 4-month old infants from monolingual and bilingual environments. Cognition 65, 33–69.

Bosch, L., Sebastián-Gallés, N., 2001a. Early language discrimination in bilingual infants. In: Cenoz, J., Genesee, F. (Eds.), Trends in Bilingual Acquisition. John Benjamins, Amsterdam.

Bosch, L., Sebastián-Gallés, N., 2001b. Evidence of early language discrimination abilities in infants from bilingual environments. Infancy 2, 29–49.

Brown, S., Martinez, M.J., Hodges, D.A., Fox, P.T., Parsons, L.M., 2004. The song system of the human brain. Cognit. Brain Res. 20, 363–375.

Cordier, F., Le, N.Y.J., 2005. Evidence for several components of word familiarity. Behavior Research Methods 37, 528–537.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.

Cox, R.W., Jesmanowicz, A., 1999. Real-time 3D image registration for functional MRI. Magn. Reson. Med. 42, 1014–1018.

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., Thierry, G., Cardebat, D., 2005. Renewal of the neurophysiology of language: Functional neuroimaging. Physiol. Rev. 85, 49–95.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., van der Vrecken, O., 1996. The MBROLA Project: Toward a set of high-quality speech synthesizers free of use for non-commercial purposes. Proceedings of the Fourth International Conference on Spoken Language Processing (Philadelphia, PA), pp. 1393–1396.

Gabrieli, J.D.E., Poldrack, R.A., Desmond, J.E., 1998. The role of left prefrontal cortex in language and memory. Proc. Natl. Acad. Sci. U. S. A. 95, 906–913.

Gandour, J., 2006. Brain mapping of Chinese speech prosody. In: Li, P., Tan, L.H., Bates, E., Tzeng, O.J.L. (Eds.), Handbook of East Asian Psycholinguistics (Vol. 1: Chinese). Cambridge University Press.

Gandour, J., Xu, Y., Wong, D., Dzemidzic, M., Lowe, M., Li, X., Tong, Y., 2003. Neural correlates of segmental and tonal information in speech perception. Hum. Brain Mapp. 20, 185–200.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., Lowe, M., 2004. Hemispheric roles in the perception of speech prosody. NeuroImage 23, 344–357.

Gold, B., Balota, D., Jones, S., Powell, D., Smith, C., Andersen, A., 2006. Dissociation of automatic and strategic lexical-semantics: functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. J. Neurosci. 26, 6523–6532.

Hagoort, P., 2005. On Broca, brain, and binding: a new framework. Trends Cognit. Sci. 9, 416–423.

Hesling, I., Clement, S., Bordessoules, M., Allard, M., 2005. Cerebral mechanisms of prosodic integration: evidence from connected speech. NeuroImage 24, 937–947.

Hernandez, A., Li, P., 2007. Age of acquisition: its neural and computational mechanisms. Psychol. Bull. 133, 638–650.

Jongman, A., Wang, Y., Moore, C., Sereno, J., 2006. Perception and production of Chinese tones. In: Li, P., Tan, L.H., Bates, E., Tzeng, O.J.L. (Eds.), Handbook of East Asian psycholinguistics (Vol. 1: Chinese). Cambridge University Press.

Leavers, V.F., Burley, C.E., 2001. The use of cognitive processing strategies and linguistic cues for efficient automatic language identification. Lang. Sci. 23, 639–650.

Li, P., Bates, E., MacWhinney, B., 1993. Processing a language without inflections: a reaction time study of sentence interpretation in Chinese. J. Mem. Lang. 32, 169–192.

Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T., Medler, D.A., 2005. Neural substrates of phonemic perception. Cereb. Cortex 15, 1621–1631.

Lorch, M., Meara, P., 1995. Can people discriminate languages they don't know? Lang. Sci. 17, 65–71.

MacWhinney, B., 1989. Competition and connectionism. In: MacWhinney, B., Bates, E. (Eds.), The Crosslinguistic Study of Sentence Processing. Cambridge University Press, New York, pp. 422–457.

MacWhinney, B., 2004. A multiple process solution to the logical problem of language acquisition. J. Child Lang. 31, 883–914.

MacWhinney, B., James, J., Schunn, C., Li, P., Schneider, W., 2001. STEP—a system for teaching experimental psychology using E-prime. Behav. Res. Meth. Ins. C 33, 287–296.

Meyer, M., Alter, K., Friederici, A.D., Lohmann, G., von Cramon, D.Y., 2002. FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum. Brain Mapp. 17, 73–88.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A.D., von Cramon, D.Y., 2004. Brain activity varies with modulation of dynamic pitch variance in sentence melody. Brain Lang. 9, 277–289.

Nazzi, T., Ramus, F., 2003. Perception and acquisition of linguistic rhythm by infants. Speech Comm. 41, 233–243.

Nazzi, T., Bertoncini, J., Mehler, J., 1998. Language discrimination by newborns: towards an understanding of the role of rhythm. J. Exp. Psych. Hum. Percept. Perform. 24, 756–766.

Nazzi, T., Jusczyk, P.W., Johnson, E.K., 2000. Language discrimination by English-learning 5-month-olds: effects of rhythm and familiarity. J. Mem. Lang. 43 (1), 1–19.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

Plante, E., Creusere, M., Sabin, C., 2002. Dissociating sentential prosody from sentence processing: activation interacts with task demands. NeuroImage 17, 401–410.

Price, C.J., 2000. The anatomy of language: contributions from functional neuroimaging. J. Anat. 197, 335–359.

Psychology Software Tools, Inc., 2005. E-Prime (Version 1.1.). http://www.pstnet.com/products/e-prime/2005.

Ramus, F., 2002. Language discrimination by newborns: teasing apart phonotactic, rhythmic, and intonational cues. Annu. Rev. Lang. Acquis. 2, 85–115.

Ramus, F., Mehler, J., 1999. Language identification with suprasegmental cues: a study based on speech resynthesis. J. Acoust. Soc. Am. 105, 512–521.

Ramus, F., Nespor, M., Mehler, J., 1999. Correlates of linguistic rhythm in the speech signal. Cognition 73, 265–292.

Ramus, F., Hauser, M., Miller, C., Morris, D., Mehler, J., 2000. Language discrimination by human newborns and by cotton-top tamarin monkeys. Science 288, 349–351.

Riecker, A., Wildgruber, D., Dogil, G., Grodd, W., Ackermann, H., 2002. Hemispheric lateralization effects of rhythm implementation during syllable repetitions: an fMRI study. NeruoImage 16, 169–176.

Stowe, L.A., Haverkort, M., Zwarts, F., 2005. Rethinking the neurological basis of language. Lingua 115, 997–1042.

Talairach, J., Tournoux, P., 1988. Co-planar Stereotaxic Atlas of the Human Brain. Thieme, Stuttgart.

Tincoff, R., Hauser, M., Tsao, F., Mehler, J., 2005. The role of speech rhythm in language discrimination: further tests with a non-human primate. Dev. Sci. 8, 26–35.

Toro, J.M., Trobalon, J., Sebastián-Gallés, N., 2003. The use of prosodic cues in language discrimination tasks by rats. Anim. Cognit. 6, 131–136.

Vigneau, M., Beaucousin, V., Herve, P.Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., Tzourio-Mazoyer, N., 2006. Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. NeuroImage 30, 1414–1432.

Wang, Y. 2002. Perception, production, and neurophysiological processing of lexical tone by native and non-native speakers. Unpublished Doctoral Dissertation, Cornell University, Ithaca, USA.

Wang, Y., Sereno, J.A., Jongman, A., Hirsch, J., 2003. fMRI evidence for cortical modification during learning of Mandarin lexical tone. J. Cogn. Neurosci. 15, 1019–1027.

Weikum, W., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., Werker, J., 2007. Visual language discrimination in infancy. Science 316, 1159.

Wong, P.C.M., Parsons, L.M., Martinez, M., Diehl, R.L., 2004. The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. J. Neurosci. 24, 9153–9160.

Zhang, L., 2006. Rhythm in Chinese and its acquisition. Doctoral dissertation, Beijing Normal University, Beijing.

Zhang, Y., Wang, Y., 2007. Neural plasticity in speech acquisition and learning. Biling. Lang. Cognit. 10, 147–160.