

# A Decisional Space for fMRI Pattern Separation Using the Principal Component Analysis—A Comparative Study of Language Networks in Pediatric Epilepsy

Xiaozhen You,<sup>1,2</sup> Malek Adjouadi,<sup>3\*</sup> Jin Wang,<sup>3</sup> Magno R. Guillen,<sup>4</sup> Byron Bernal,<sup>4</sup> Joseph Sullivan,<sup>5</sup> Elizabeth Donner,<sup>6</sup> Bruce Bjornson,<sup>7,8</sup> Madison Berl,<sup>9</sup> and William D. Gaillard<sup>9,10,11</sup>

<sup>1</sup>Department of Biomedical Engineering, Florida International University, Miami, Florida

<sup>2</sup>Department of Psychology, Georgetown University, Washington, DC

<sup>3</sup>Department of Electrical and Computer Engineering, Florida International University, Miami, Florida

<sup>4</sup>Department of Radiology, Miami Children's Hospital, Miami, Florida

<sup>5</sup>Division of Neurology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

<sup>6</sup>Department of Neurology, Hospital for Sick Children, Toronto, Ontario, Canada

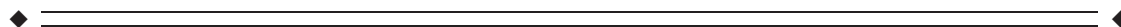
<sup>7</sup>Division of Neurology, BC Children's Hospital, Vancouver, British Columbia, Canada

<sup>8</sup>Child & Family Research Institute, Vancouver, British Columbia, Canada

<sup>9</sup>Department of Neurosciences, Children's National Medical Center, George Washington University, Washington, District of Columbia

<sup>10</sup>Department of Neurology, Georgetown University, Washington, District of Columbia

<sup>11</sup>Clinical Epilepsy Section, NINDS, NIH, Bethesda, Maryland



**Abstract:** Atypical functional magnetic resonance imaging (fMRI) language patterns may be identified by visual inspection or by region of interest (ROI)-based laterality indices (LI) but are constrained by a priori assumptions. We compared a data-driven novel application of principal component analysis (PCA) to conventional methods. We studied 122 fMRI data sets from control and localization-related epilepsy patients provided by five children's hospitals. Each subject performed an auditory description decision task. The data sets, acquired with different scanners but similar acquisition parameters, were processed through fMRIB software library to obtain 3D activation maps in standard space. A PCA analysis was applied to generate the decisional space and the data cluster into three distinct activation patterns. The classified activation maps were interpreted by (1) blinded reader rating based on predefined language patterns and (2) by language area ROI-based LI (i.e., fixed threshold vs. bootstrap

Contract grant sponsor: The American Epilepsy Society; Contract grant number: NINDS R01 NS44280; Contract grant sponsor: Children's Research Institute Avery Award, Intellectual and Developmental Disabilities Research Center at Children's National Medical Center; Contract grant number: NIH IDDC P30HD40677; Contract grant sponsor: The General Clinic Research Center; Contract grant number: NIH GCRC M01-RR13297; Contract grant sponsor: The National Science Foundation; Contract grant numbers: HRD-0833093, CNS-0959985, CNS-

1042341; Contract grant sponsor: Ware Foundation and the joint Neuro-Engineering Program with Miami Children's Hospital.

\*Correspondence to: Malek Adjouadi, College of Engineering and Computing, Florida International University, 10555 W. Flagler Street, EC 2672, Miami, FL 33174, USA. E-mail: adjouadi@fiu.edu  
Received for publication 25 July 2011; Revised 3 January 2012; Accepted 11 February 2012

DOI: 10.1002/hbm.22069

Published online 28 March 2012 in Wiley Online Library (wileyonlinelibrary.com).

approaches). The different classification results were compared through  $\kappa$  inter-rater agreement statistics. The unique decisional space classified activation maps into three clusters (a) lower intensity typical language representation, (b) higher intensity typical, as well as (c) higher intensity atypical representation. Inter-rater agreements among the three raters were excellent (Fleiss  $\kappa = 0.85$ ,  $P = 0.05$ ). There was substantial to excellent agreement between the conventional visual rating and LI methods ( $\kappa = 0.69$ – $0.82$ ,  $P = 0.05$ ). The PCA-based method yielded excellent agreement with conventional methods ( $\kappa = 0.82$ ,  $P = 0.05$ ). The automated and data-driven PCA decisional space segregates language-related activation patterns in excellent agreement with current clinical rating and ROI-based methods. *Hum Brain Mapp* 34:2330–2342, 2013. © 2012 Wiley Periodicals, Inc.

**Key words:** brain activation pattern; data-driven clustering; fMRI; epilepsy; language; lateralization indices; PCA-based decisional space; visual rating

## INTRODUCTION

Previous studies with intracarotid amobarbital (IAT), Doppler flow, and functional magnetic resonance imaging (fMRI) demonstrate that language dominance is typically left dominant, but there are known variants—bilateral or right dominance—present in both right-handed (5%) and left-handed (22%) populations [Pujol et al., 1999; Rasmussen and Milner, 1977; Springer et al., 1999; Szaflarski et al., 2002; Woods et al., 1988]. Moreover, patients with Localization-related epilepsy (LRE) reveal greater occurrence of atypical language (20–30%) based on the quantitative region of interest (ROI) analysis [Frost et al., 1999; Gaillard et al., 2007; Woermann et al., 2003] at hemispheric or regional levels [Binder et al., 1996; Gaillard et al., 2002; Ramsey et al., 2001; Spreer et al., 2002] or through visual rating (VR) [Fernandez et al., 2001; Gaillard et al., 2002, 2004] using fMRI data. This study aims to expand upon data-driven and unbiased methods to identify deviant fMRI language patterns that are not constrained by a priori assumptions including regions or thresholds used in conventional assessments.

Previously, we reported the clustering characteristics of a principal component analysis (PCA)-based decisional space in relation to intensity and lateralization dominance of fMRI language task activation patterns. The study was based on 122 children (64 control and 58 LRE patients with age range of 4.5–19 years) across five sites using EPI BOLD fMRI and an auditory description decision task [You et al., 2011]. The method identified three distinct patterns, one

atypical right dominant and two typical left dominant. One of the left dominant groups featured higher intensity activation in the left frontal lobe and was thought to represent a variant group that may have used a different cognitive strategy to perform the task. In this study, we aimed to ascertain the agreement of our PCA method with two conventional methods: (1) quantitative ROI-based laterality index (LI) rating and (2) subjective clinical VR.

Epilepsy patient populations provide a means for validating these novel segregation methods because of their known heterogeneity of language dominance either on adults [Price et al., 2006, 2005] or on older children or adolescents [Berl et al., 2005]. Such an approach could augment the knowledge gained through different PCA-based approaches reported in the literature that have either attempted to maximize variability [Suma and Murali, 2007], used for recovering signal of interest in 4-D fMRI data sets [Andersen et al., 1999; Viviani et al., 2005], applied the *k*-means method [Mbwana et al., 2009], or used the scale sub-profile model (SSM) normalization transformation [Alexander and Moeller, 1994]. In all of these approaches, the shared belief is that the PCA remains a powerful data-driven method with good scalability. The PCA can also overcome the need for a priori assumptions and the subjectivity associated with VR methods that are prone to bias.

We aimed to compare our PCA-based method, which is independent of a priori assumptions and biases inherent to ROI and visual analyses, with conventional visual and ROI rating methods among control and epilepsy groups.

## DATA AND SUBJECTS

Our institution, in partnership with pediatric epilepsy programs at five children's hospitals, built a multisite consortium and repository for pediatric epilepsy data (<http://mri-cate.fiu.edu/>) to investigate the effects of epilepsy on brain function and structure [Lahlou et al., 2006]. Procedures were followed in accordance with local institutional review board requirements; all parents gave written informed consent and children gave assent. Each data set was deidentified to guarantee patient's confidentiality.

### Abbreviations

B	bilateral
fMRI	functional magnetic resonance imaging
L	left
LI	laterality index
LRE	localization-related epilepsy
O	other
PCA	principal component analysis
R	right
ROI	region of interest

Patients were between 4.5 and 19 years old and underwent epilepsy surgery evaluation. Typically developing control subjects were right handed, native English speakers, and free of any current or past neurological or psychiatric disease; they were recruited from the community. In all, 64 control and 58 children with LRE included in this study have complete and usable clinical data [You et al., 2011]. The mean age of patients was 13.86 years, with mean age seizure onset 8.23 years (range, 1–18 years). Twenty-six patients had a left localized hemisphere focus (17 temporal and 9 extra-temporal foci); 18 had a right hemisphere focus (7 temporal foci and 11 extra-temporal foci). Three patients had bilateral independent seizure foci. Twenty-two patients had abnormal MRI: seven tumor; five mesial temporal sclerosis; four focal cortical dysplasia; one vascular malfunction; three focal gliosis; and two atrophy. Of the 45 patients with seizure etiology information, 21 had remote symptomatic etiology, 21 unknown causes, and 3 acute symptomatic. Eleven patients (out of the 54 available) had atypical dexterity (left or ambidextrous).

All participants performed an auditory description decision task-ADDT (active condition, a word definition—for example “a large gray animal is an elephant”—during which subjects pressed a button when the statement was true; control condition, reverse speech, during which subjects pressed a button on hearing an interspersed beep; for both conditions, 70% true responses and 30% foils) designed to activate both temporal (Wernicke’s area, “receptive speech”) and inferior frontal (Broca’s area, “expressive speech”) cortex [Gaillard et al., 2007]. The task required comprehension of a phrase, semantic recall, and a semantic decision via push button response. The block design paradigm consisted of 100 (TR = 3 s) or 150 (TR = 2 s) time points, with experimental and baseline periods alternating every 30 s for five cycles, totaling 5 min.

For uniform analysis of data sets collected from several sites, all data sets were exported into Neuroimaging Informatics Technology Initiative (NIFTI) format using the transversal view and radiology convention, and were registered into the standard Montreal Neurological Institute brain with  $3 \times 3 \times 3$  (mm<sup>3</sup>) voxel size and resolution of  $61 \times 73 \times 61$  (axial  $\times$  coronal  $\times$  sagittal). The fMRIB Software Library was used to perform the pre and postprocessing required for determining the 3D activation maps [Jenkinson et al., 2002; Woolrich et al., 2001], including motion correction using MCFLIRT [Jenkinson et al., 2002], brain extraction using BET [Smith, 2002]; spatial smoothing using Gaussian kernel of FWHM 8 mm, intrasubject mean-based intensity normalization, high pass temporal filtering with sigma 120.0 s, time-series statistical analysis using FMRIB’s improved linear model with local autocorrelation correction [Woolrich et al., 2001], using fMRI Expert Analysis tool to generate  $Z$  (Gaussianized  $T/F$ ) statistic images thresholded using clusters determined by  $Z > 2.3$  and a (corrected) cluster significance threshold of  $P = 0.05$  [Forman

et al., 1995; Friston et al., 1994; Worsley et al., 1992], and registration to high-resolution and standard images was carried out using FLIRT [Jenkinson et al., 2002].

## METHOD

### Conventional Methods

For quantitative assessments, we used two approaches to calculate regional LIs: (1) a fixed threshold (the threshold at which VR was performed  $Z > 2.3$ ) and (2) a bootstrap method (postulated to be more versatile than a fixed threshold as it incorporates multiple thresholds into its calculations) [Wilke and Schmithorst, 2006; Wilke and Lidzba, 2007]. There are occasions where one may be able to obtain more clinically useful assessment of language dominance by looking at several thresholds [Binder, 1996; Gaillard et al., 2002], an approach that is captured by the bootstrap method which is aimed at rendering an optimal LI value. With these measures, we then categorized the regional activation maps into a specific laterality category using these following criteria:  $LI \geq 0.2$  is deemed left;  $LI \leq -0.2$  is deemed right;  $|LI| < 0.2$  is deemed bilateral [Gaillard et al., 2002, Wilke and Schmithorst, 2006]. The regions employed can be hemispheric [Binder et al., 1996; Springer et al., 1999], or sub regions in the frontal and temporal lobes or any small region, such as specific functional Brodmann area [Gaillard et al., 2002; Ramsey et al., 2001; Spreer et al., 2002]. We defined typical language-related area Broca’s [BA: 44, 45, 47] and Wernicke’s [BA 21, 22, 39] areas as our two ROIs. These specific regions were extracted from BA templates provided by MRICro [Rorden and Brett, 2000]. LI was derived using the following equation:

$$LI = \frac{\sum V_{\text{Left}} - \sum V_{\text{Right}}}{\sum V_{\text{Left}} + \sum V_{\text{Right}}} \quad (1)$$

where  $V$  denotes the activation magnitude or voxel count. The average LI which takes into account both extent and magnitude is defined as follows:

$$LI_{\text{avg}} = (LI_{\text{extent}} + LI_{\text{magnitude}})/2 \quad (2)$$

Once the two regions’ LIs were calculated (for fixed threshold and bootstrap methods), we then used the following criteria for determining hemispheric language dominance: (1) right dominance when both Broca’s and Wernicke’s areas are right lateralized, or when one region is deemed right and the other bilateral or with no activity (NA); (2) left dominance if both regions are left, or one region is left and the other is bilateral or with NA; (3) bilateral if both regions are bilateral, or one is bilateral and the other is with NA, or one is left and the other is right [Fernandez et al., 2001; Gaillard et al., 2002, 2004]. For visual and ROI ratings, right activation and bilateral

activation are considered to be atypical language dominance patterns.

Furthermore, we used a comparable strategy for the VR system to categorize each subject's activation map into similar dominance patterns. Instead of a pre-extracted ROI, the language-related areas were visually identified. An Access-based tool was used such that all images could be scored to one of the predetermined set of language network patterns representing the differing combinations of activations in canonical frontal and temporal regions [Gaillard et al., 2004]. Each subject's activation map was first thresholded ( $Z > 2.3$  with  $P = 0.05$  cluster corrected) for visualization clarity and then overlaid on top of the brain template. Three skilled readers, who were blinded to subject identity, then categorized the laterality of activation in language network-related areas, and each rater provided the level of confidence in the rating (confidence scale was from 1 to 5: 5 being most confident, 1 least (i.e., not) confident of rating). Raters also made relevant comments on their observations when there was concern about noise or null activation [Guillen et al., 2009].

### The PCA-Based Decisional Space

Though PCA is commonly used as a dimensional reduction process, it plays an important role in multivariate statistical analysis as well. PCA has been reported as the core analysis of SSM after the initial normalization transformation in the context of modeling regional patterns of brain function [Alexander and Moeller, 1994]. In this study, normalization of intensity is purposely avoided to allow for the examination of any effect the original features embedded in the data sets may have, which may serve as a means to gauge intersubject variability.

Based on the concept of subject loading, we performed the PCA on the given 122 fMRI activation maps foregoing altogether regional masking and intensity normalization as previously reported [You et al., 2011]. 3D data were arranged into a 2D matrix where each subject's data constitute a specific column. An eigensystem was then generated. Based on the relationship among the top eigenvectors, general lateralization, intensity difference, as well as the dendrogram of the Euclidian distance matrix of the PCA, the top two eigenvectors of the PCA-based decisional space were used to delineate three primary clusters (the first as major group left dominant, the second featured higher intensity levels, and the third with right dominant activation). The undecided cases were then projected onto a new decisional space based on the PCA of only those data sets that initially were identified as belonging to the three primary clusters. By using the modified-Euclidean distance method, the undecided cases were then classified in the new decisional space into one of the three primary clusters initially determined, using unique mathematically derived PCA-based thresholds.

### Statistical Analysis

First, we assessed inter-rater agreement. Inter-rater agreements among the three raters were analyzed using Fleiss  $\kappa$  coefficients to evaluate the rating difficulty and consistency among different raters and different level of clinical experiences [Fleiss, 1971]. The interpretation from the  $\kappa$  value follows:  $>0.81$  as excellent agreement,  $0.61-0.80$  as good agreement,  $0.41-0.60$  as moderate agreement,  $0.21-0.40$  as fair agreement, and  $<0.20$  as poor agreement. Next, we compared a single rater (most senior) to the fixed threshold LI and the bootstrap LI using  $\chi^2$ . Classifications, including mismatches, between visual and LI rating were identified and plotted by the two ROIs (Broca's and Wernicke's areas). Finally, we compared the visual rater, and both LI methods to PCA analysis. Agreement between the conventional methods and the PCA results was quantified using Kappa coefficient ( $k$ ) [Viera and Garrett, 2005]. Owing to the characteristics of the three groups identified through the PCA decisional space, we compared the language dominance in terms of right dominant (corresponding to Group 3) and left dominant (corresponding to Groups 1 and 2).


















To assess the stability of our PCA decisional space, we examined whether classification results differ if a decisional space included only the LRE patients. Then we generated 200 synthetic data sets, 100 left lateralized and 100 right lateralized, and classified them based on the current PCA decisional space initially derived from the 122 real subjects to assess the sensitivity and specificity.

## RESULTS

### Summary of Previous Results on the PCA Decisional Space

After performing the PCA on the 122 fMRI activation maps, the eigenvalues of the first two eigenvectors (PCA spatial patterns) alone make up 80% of the sum of all eigenvalues and reveal the regions of interest, a bipolar value of anterior (Broca) and posterior (Wernicke) clusters. As a result, fMRI activation maps can later be masked with language areas (Broca's and Wernicke's) to calculate the ROI-based LI. Moreover, the zero line in the second eigenvector axis tends to separate typical and atypical language groups, according to the general definition derived from overall LI. Furthermore, the zero line of the first eigenvector was deemed sufficient at separating any of the following two groups: higher intensity typical vs. atypical, lower intensity typical vs. atypical, as well as higher intensity typical vs. lower intensity typical. The dendrogram of the subjects' loading matrix identified three major clusters within these 122 subjects [You et al., 2011]. After the selection of primary cluster member, as well as applying the Euclidean Distance method, three subgroups of the 122 subjects were identified with distinct group maps: typical left dominant (Group 1), typical left dominant with higher intensity (Group 2), and atypical right dominant (Group 3).

**TABLE I.** The pattern of distribution as identified by the three raters. [Color table can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

Left	Rater 1			Rater 2			Rater 3		
	Control	LRE	Total	Control	LRE	Total	Control	LRE	Total
	37	30	67	38	31	69	39	33	72
	1	3	4	2	4	6	1	4	5
	5	3	8	7	2	9	4	2	6
	1	1	2	1	1	2	0	1	1
	6	5	11	4	5	9	4	4	8
<b>Bilateral</b>									
	3	1	4	3	1	4	5	1	6
	0	0	0	0	0	0	0	0	0
	1	2	3	0	1	1	1	0	1
	1	0	1	0	0	0	0	0	0
	2	0	2	1	1	2	1	0	1
<b>Right</b>									
	0	0	0	0	0	0	0	1	1
	1	2	3	1	2	3	0	2	2
	0	1	1	0	1	1	1	1	2
	0	2	2	1	2	3	0	2	2
	0	5	5	0	5	5	1	4	5
<b>Other</b>									
	0	0	0	0	0	0	1	0	1
	6	3	9	6	2	8	6	2	9

**Categorization Through Conventional Methods**

The inter-rater agreement for VR among the three raters was 0.81 (Fleiss  $\kappa$   $P = 0.05$ ) (Table I). There was lower rat-

ing confidence in the cases where there was rater disagreement (unpaired two sample  $t$ -test,  $t = -3.61$ ,  $P < 0.001$ ,  $n1 = 86$  [concordant cases, mean 2.7],  $n2 = 36$  [discordant cases, mean 2.3]). Moreover, 52% of the confidence ratings

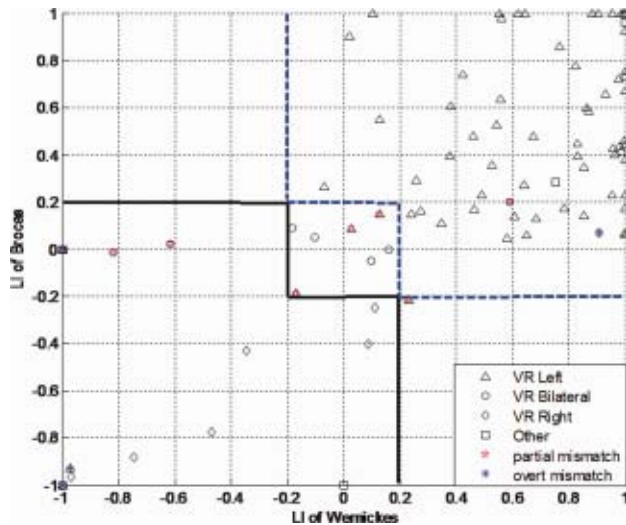


**TABLE II. Categorization results between VR and LI**

		VR (Rater 3)			
		Left	Right	Bilateral	Other
LI (fixed)	Left	83	1	1	3
	Right	3	11	2	3
	Bilateral	5	0	5	0
	Other	1	0	0	4
LI (bootstrap)	Left	84	3	5	6
	Right	3	8	0	3
	Bilateral	5	1	3	1

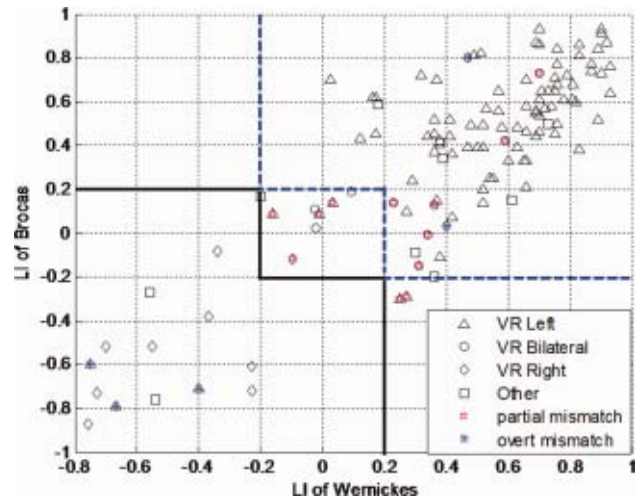
from the most senior reader (Rater 3) were a “3” or better compared to 26% for the other readers. Therefore, the categorization results from the senior reader’s rating were selected for further comparison with other categorization methods.

Rater 3 exhibited substantial to excellent inter-rater agreement with fixed LI ( $\kappa$  0.64 with all four categories and 0.69 without the “other” category; 0.82 with only left and right categories), but moderate to substantial inter-rater agreement with bootstrap LI (0.51 without the



**Figure 1.**

Distribution of visual rating results in fixed thresholded LI of Broca’s and Wernicke’s language areas. Solid black line delineates the border between right dominant and bilateral and the blue dash line is the border between bilateral and right dominant defined at fixed threshold LI ( $Z > 2.3$ ). VR standards for visual rating, different legends show the characteristics of results of visual rating by Rater 3. Red pentagram shows the partial mismatch R/B, B/R, L/B, or B/L between VR/LI. Blue asterisks are the overt mismatch cases, where R/L or L/R between VR/LI. The null activation in fixed thresholded LI is not plotted. Some symbols overlap. B, bilateral; R, right; L, left; VR, visual rating; LI, laterality indices. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]



**Figure 2.**

Distribution of visual rating results in bootstrap LI of Broca’s and Wernicke’s language areas. Solid black line delineates the border between right dominant and bilateral and the blue dash line is the border between bilateral and right dominant defined by bootstrap LI. VR standards for visual rating, different legends show the characteristics of results of visual rating by Rater 3. Red pentagram shows the partial mismatch R/B, B/R, L/B, or B/L between VR/LI. Blue asterisks are the overt mismatch cases: R/L or L/R between VR/LI. B, bilateral; R, right; L, left; VR, visual rating; LI, laterality indices. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

“other” category; 0.69 with only left and right categories). No significant differences were found in the distribution of the categorizations rendered by VR of Rater 3 as compared to either of the LI methods (Table II).

Mismatch between Rater 3’s VR and LIs (fixed thresholded and bootstrap, Figs. 1 and 2) are plotted in a 2D plane with axis of LIs of Broca’s and Wernicke’s to discern sources of disagreements.

Instances of either overt or partial disagreement were uncommon, and had either lower confidence score from Rater 3 VR or higher standard deviation for bootstrap LI.

Between the fixed threshold LI and the VR, there were four cases of overt disagreement; rater confidence for all four cases were poor ( $<2.5$ ,  $t = -1$ ,  $P < 0.2$ ); there were eight cases of partial disagreement; again, the rater confidence was poor ( $<2.5$ ,  $t = -3.7$ ,  $P = 0.003$ ); there were six cases rated visually as “other” but classified as left ( $n = 3$ ) or right ( $n = 3$ ) by LI; there was one case rated visually as “left” but classified as “other” by LI. There was lower rating confidence in the circumstances where the raters exhibited some form of disagreement ( $n = 19$ ) with LI than those that were in agreement ( $n = 103$ ) (unpaired two sample  $t$ -test,  $t = -2.9253$ ,  $P = 0.0021$ ).

Between the bootstrap LI and the VR, there were six cases of overt disagreement and 11 cases of partial disagreement. For cases of partial disagreement, the rater

**TABLE III. Categorization results between conventional methods and PCA**

PCA	Rater1			Rater2				Rater3				LI (Fixed threshold)				LI (Bootstrap)			
	L	R	B	O	L	R	B	O	L	R	B	O	L	R	B	O	L	R	B
	Group 1 (L)	75	3	8	9	78	4	5	8	75	4	6	10	71	11	8	5	79	7
Group 2 (L)	17	0	1	0	17	0	1	0	17	0	1	0	17	0	1	0	18	0	0
Group 3 (R)	0	8	1	0	0	8	1	0	0	8	1	0	0	8	1	0	1	7	1

confidence was poor ( $<2.5$ ,  $t = -2.1$ ,  $P = 0.03$ ); there were 10 cases rated visually as “other” but classified as left ( $n = 6$ ) or right ( $n = 3$ ) or bilateral ( $n = 1$ ) by bootstrap LI. There was lower rating confidence and greater variability of bootstrap LI values in the cases where the raters exhibited some form of disagreement ( $n = 27$ ) than those that were in agreement ( $n = 95$ ) (unpaired two sample  $t$ -test, VR confidence:  $t = -2.1$ ,  $P < 0.02$ ; LI standard deviation:  $t = 2.1$ ,  $P < 0.02$ ).

**Assessing Measurements with the PCA Clustering Results**

The PCA analysis placed patients with visual/ROI bilateral ratings in to either left dominant groups (1 and 2) or right dominant group (3); they are automatically considered as partial mismatches (Table III). All three raters exhibited substantial to excellent inter-rater agreement with PCA in left and right categories ( $\kappa$  0.78-0.83). ROI LI showed moderate inter-rater agreement with PCA (fixed

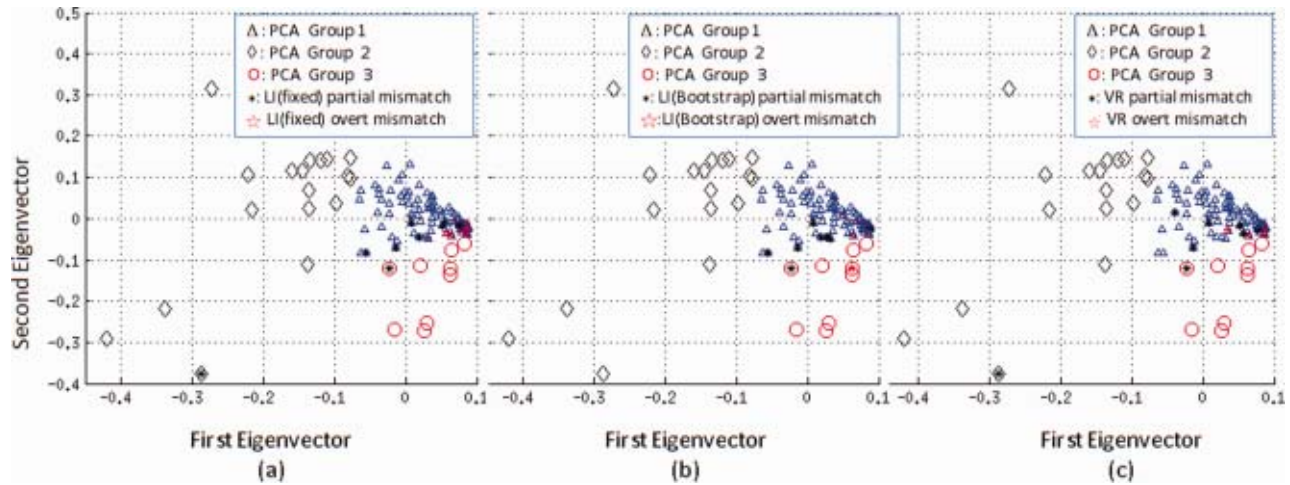
LI:  $\kappa$  0.54; bootstrap LI:  $\kappa$  0.60). The overt and partial mismatches cases between PCA and conventional methods (fixed threshold LI, bootstrap LI, and Rater 3’s VR) are shown in the PCA first two eigenvectors’ space in Figure 3.

Next, we performed a comparison of conventional methods, combining the concordant results of LI and VR as shown in Table IV and Figure 4. Intermethod agreements between PCA and conventional methods were excellent in the left and right categorization ( $\kappa = 0.82$ ,  $P = 0.05$ ).

Appendix A presents the postsurgical data for LRE participants who had epilepsy surgery.

**DISCUSSION**

A data-driven method that identifies activation patterns in heterogeneous populations is useful to facilitate timely and unbiased analysis, especially when data sets are large. Our language fMRI consortium of pediatric epilepsy and control populations aims to use these analyses to enhance



**Figure 3.**

Distribution of discordant cases in the PCA top two eigenvector space. Partial disagreement is shown in black asterisk and overt disagreement in red pentagons. There are only partial disagreements for members in Group 2 (left dominant with higher activity) and Group 3 (right dominant). (a) Between fixed threshold LI and PCA, there are totally 11 overt mismatch cases in Group

1 (left dominant with lower activity than Group 2 or Group 3). (b) Between bootstrap LI and PCA, there are total eight overt mismatch cases in Group 1 (c) Between VR and PCA, there are in total four overt mismatch cases in Group 1. VR, visual rating; LI, laterality indices. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

**TABLE IV. Categorization results between combined conventional methods and PCA**

PCA	VR and LI (fixed threshold)			
	L	R	B	O
Group 1(L)	66	3	3	4
Group 2(L)	17	0	1	0
Group 3 (R)	0	8	1	0

our understanding of activation maps in relation to clinical variables. We developed a PCA-based decisional space which achieved automatic classification of different language network activation patterns segregated into three groups: left dominant-high intensity, left dominant-low intensity, and an atypical right dominant group. Comparing the PCA method with the traditional methods of ROI-based LI and visual clinical rating, we found excellent but not complete agreement. These findings fundamentally validate the PCA decision-making process. Research methods of classification/segregation of data into subgroups can render results that are informative from a neurobiological perspective. In addition, by demonstrating that the PCA method has strong agreement with traditional clinical methods, PCA has promise as a useful clinical tool.

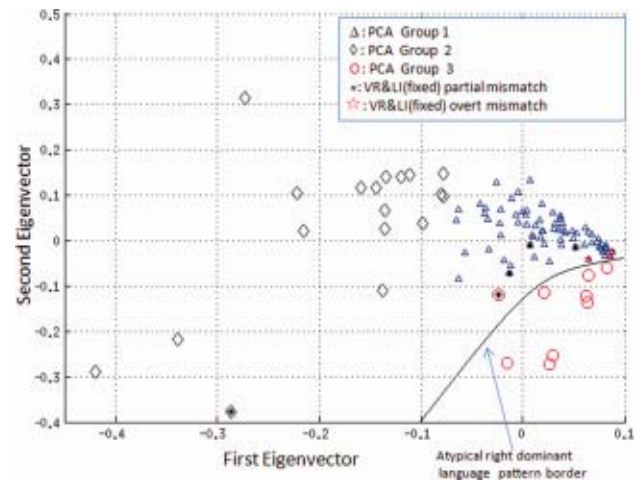
ROI methods are quantitative but are based on a priori assumptions—type of ROI, location of ROI, threshold used—and can be time consuming. The PCA method we employed is automated, data driven, uniform, and easy to use. We do not claim that the proposed method is better than other segregation methods; rather, we suggest that these methods can be applied to patient populations on group and at individual levels. One of the distinguishing factors of the PCA approach is that patients and controls are found in all three groups, an observation that emphasizes normal and pathological variability rather than strict division between controls and patients. The distribution of language patterns obtained in this study is similar to prior studies of language dominance using other techniques including transcranial-Doppler, transcranial magnetic stimulation, or the IAT [Gaillard et al., 2002, 2004; Khedr et al., 2002; Knecht et al., 2000; Kurthen et al., 1994; Rasmussen and Milner, 1977; Risse et al., 1997; Springer et al., 1999; Woermann et al., 2003; Woods et al., 1988; Wyllie et al., 1991].

Although the agreements across rating systems and methods were substantial, discrepancies did occur. The inter-rater agreement and agreement with ROI are similar to the previous reports [Fernandez et al., 2001; Gaillard et al., 2002]. Most differences are of partial disparity rather than overt disagreement [Gaillard et al., 2002]. Through a systematic target-oriented rating system [Gaillard et al., 2002], senior readers were able to rate language pattern in excellent agreement. Not surprisingly, cases where there was disagreement among three raters had lower confidence than those cases with complete agreement. Simi-

larly, among the three raters, the most senior rater had an overall higher confidence score, and had better agreement with the LI. These observations emphasize the need for skill in VR and highlight the limits of subjective rating systems, suggesting the need for caution when less confident of rating. An additional reason for the discrepancies is that the VR was performed at a fixed threshold.

Moreover, a LI calculated based on voxel counting may neglect important information of the activation intensity, whereas the voxel summation method is sensitive to statistical outliers [Price et al., 2005; Wilke and Lidzba, 2007; Wilke and Schmithorst, 2006]. In this study, the averaged LI which takes both factors into account was used to compare PCA clustering results. The LI is also affected by the defined ROI activation. Apart from the subjectivity of VR that might account for the disagreements between the PCA and the VR, the differences may also be explained by the approach taken in thresholding: PCA uses only raw intensity values, whereas the VR is based on postprocessed thresholded images and judgment of activation relevance. For example, a weak activator's thresholded image may be dominated by the peaks (maxima), whereas the raw data utilized by the PCA method account for all the activated voxels (extent).

Furthermore, PCA is sensitive to intensity difference and does not cluster based on lateralization, rather the eigenvectors which we have presumed relate primarily to this feature; hence, some cases may segregate differently from LI and VR. Though we did not have complete



**Figure 4.**

Distribution of mismatch cases between conventional methods (VR and LI) and PCA in the first two eigenvectors (subjects loading). Partial disagreement is shown in black asterisk and overt disagreement in red pentagrams. There are only partial disagreements for members in Group 2 (left dominant with higher activity) and Group 3 (right dominant). Overt disagreements are only found in Group 1. Note that the disagreements are along borders. VR, visual rating; LI, laterality indices. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]



agreement between these methods, PCA properly identified the strong right dominant group and was supported by both VR and the ROI-based LI. All of the (rare) overt disagreements between the traditional and the PCA methods occurred only in Group 1. Disagreements, when present, occurred mostly along the borders of PCA-derived clusters. Recall, Group 1 is the left dominant lower intensity group and Group 3 has the feature of right dominant higher intensity. Therefore, when the subject has low intensity and right dominance, overt disagreement with conventional methods may occur if low-intensity features overrule the right lateralized feature and PCA would cluster the subject into Group 1 instead of Group 3. As differences with PCA are found along borders of clusters; refinement of clustering algorithms may be able to improve patient segregation. The disagreements do not mean that any one method is always correct [Gaillard et al., 2004; Hunter et al., 1999].

Without comparison to invasive methods—such as Wada testing, electrocortical mapping, or postresection outcomes—it may not be possible to determine which approach is more accurate in discordant cases between conventional and PCA methods. Furthermore, fMRI information changes clinical practice [Medina et al., 2005] as it is used to plan surgery and to minimize postoperative language deficits beyond what one would expect by operating on a temporal lobe dominant for language [Sabsevitz et al., 2003]. No patient in our series had clinically apparent postoperative aphasia but postoperative outcomes were not rigorously evaluated. Our goal is to compare traditional fMRI rating techniques to data-driven methods. Conventional ratings have excellent but not complete agreement with invasive methods [Gaillard et al., 2002; Pouratian et al., 2002].

Furthermore, fMRI category decision task hemispheric laterality indices predict postoperative performance on the Boston Naming task (Binder group). Unfortunately, there are no similar postoperative outcome studies of fMRI language tasks and analysis methods in children. Ideally, we would like to have systematic invasive method results (Wada test, electrocorticography) for comparison, especially in those cases where our analysis strategies disagree. However, Wada tests are no longer routinely performed on children at most pediatric epilepsy centers. When these tests are performed—typically when there is concern about integrity of noninvasive mapping—they proceed under circumstances that make interpretation and comparison problematic. Finally, there are clear limitations to the use of either Wada test or electrocorticography as “gold standards” as both, like fMRI, may be flawed and under various circumstances yield faulty information [Gaillard et al., 2011a].

Epilepsy has complex effects of brain function, which may include atypical levels/extent of activation on fMRI as we reported in our previous study, identifying the higher intensity left dominant group. Recent studies do not find a direct effect of regional hypometabolism or hypoperfusion on the ability of patients to mount a BOLD response [Duke

et al., in press; Gaillard et al., 2011b]. More importantly, through the use of the PCA on patients only, we found that the PCA decisional space remained stable and the algorithm worked well at clustering the patients, and because of the less skewed distribution of the three primary clusters, the clustering results conformed better to the conventional methods (Appendix B). We also showed that the sensitivity and specificity of our PCA decisional space are acceptable and thus useful clinically after applying the decisional space on the synthetic data sets (Appendix C).

### Limitations

The PCA method has several limitations, which may limit its clinical utility and application at the single subject (individual patient) level. First, the PCA decision space depends on the population that is sampled; as new subjects are included in a group sample, the decisional space will potentially change unless the sampled population includes all possible language patterns. As a consequence, one needs to establish such a decision space from a fairly large representative sample before it can be used to evaluate individual subjects. As we have over 100 subjects in our group, this method can be a basis of learning the language processing features of the normal and pediatric epilepsy populations. Despite our large sample size, the relatively low number of children with atypical language lateralization in the LRE and control groups may make it difficult for the decision space derived from the PCA analysis to include all possible language patterns. As our consortium grows in number and the database increases in size, new subjects can be projected into the decisional space which is subjected to computational learning. As a consequence, the process may be refined by potentially identifying new subgroups and subsets, and thus improve clinical applicability. Second, the classification based on the PCA-derived clusters did not separate different bilateral language patterns, which potentially limits single subject clinical utility. Third, PCA places greater emphasis on intensity of activation rather than spatial extent; it is unclear which factor is more important for language dominance

To overcome these limitations; first, we expect the consortium to grow and thus provide larger representative language pattern samples. Second, it may be possible to use synthetic activation maps to generate a multiclass classifier to refine cluster border classification.

### CONCLUSION

The PCA-based method presented here employs cluster tools that may help the assessment of very large data sets and may be applied on an individual basis based on features gleaned from a large database. The agreements found between PCA and conventional methods (LI and VR) suggest that such methods may be used clinically.

Our PCA approach segregated patients and controls with overall strong agreement with conventional methods. The differences between methods were all found along the borders of PCA clusters. Adjustments in PCA clustering should be amenable to learning and modeling to avoid pitfalls and improve clinical utility.

## REFERENCES

- Alexander G, Moeller J (1994): Application of the scaled subprofile model to functional imaging in neuropsychiatric disorders: A principal component approach to modeling brain function in disease. *Human Brain Mapp* 2:79–94.
- Andersen AH, Gash DM, Avison MJ (1999): Principal component analysis of the dynamic response measured by fMRI: A generalized linear systems framework. *Magn Reson Imaging* 17: 795–815.
- Berl MM, Balsamo LM, Xu B, Moore EN, Weinstein SL, Conry JA, Pearl PL, Sachs BC, Grandin CB, Frattali C, Ritter FJ, Sato S, Theodore WH, Gaillard WD (2005): Seizure focus affects regional language networks assessed by fMRI. *Neurology* 65:1604–1611.
- Binder JR, Swanson SJ, Hammeke TA, Morris GL, Mueller WM, Fischer M, Benbadis S, Frost JA, Rao SM, Houghton VM (1996): Determination of language dominance using functional MRI: A comparison with the Wada test. *Neurology* 46:978–984.
- Duke ES, Appel S, Martinez AR, Khan OI, Dustin I, Reeves-Tyer P, Stokum J, Berl MM, Sato S, Gaillard WD, Theodore WH. Cerebral blood flow and auditory language activation in temporal lobe epilepsy. *Epilepsia* (In Press).
- Fernandez G, de Greiff A, von Oertzen J, Reuber M, Lun S, Klaver P, Ruhlmann J, Reul J, Elger CE (2001): Language mapping in less than 15 minutes: Real-time functional MRI during routine clinical investigation. *Neuroimage* 14:585–594.
- Fleiss, J (1971): Measuring nominal scale agreement among many raters. *Psychol Bull* 76:378–382.
- Forman S, Cohen J, Fitzgerald M, Eddy W, Mintun M, Noll D (1995): Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magn Reson Med* 33:636–647.
- Friston K, Worsley J, Frackowiak R, Mazziotta J, Evans A (1994): Assessing the significance of focal activations using their spatial extent. *Hum Brain Map* 1:210–220.
- Frost JA, Binder JR, Springer JA, Hammeke TA, Bellgowan PS, Rao SM, Cox RW (1999): Language processing is strongly left lateralized in both sexes. Evidence from functional MRI. *Brain* 122:199–208.
- Gaillard WD, Balsamo L, Xu B, Grandin CB, Braniecki SH, Papero PH, Weinstein S, Conry J, Pearl PL, Sachs B, Sato S, Jabbari B, Vezina LG, Frattali C, Theodore WH (2002): Language dominance in partial epilepsy patients identified with an fMRI reading task. *Neurology* 59:256–265.
- Gaillard WD, Balsamo L, Xu B, McKinney C, Papero PH, Weinstein S, Conry J, Pearl PL, Sachs B, Sato S, Vezina LG, Frattali C, Theodore WH (2004): fMRI language task panel improves determination of language dominance. *Neurology* 63: 1403–1408.
- Gaillard WD, Berl MM, Moore EN, Ritzl EK, Rosenberger LR, Weinstein SL, Conry JA, Pearl PL, Ritter FF, Sato S, Vezina LG, Vaidya CJ, Wiggs E, Fratalli C, Risse G, Ratner NB, Gioia G, Theodore WH (2007): Atypical language in lesional and nonlesional complex partial epilepsy. *Neurology* 69:1761–1771.
- Gaillard WD, Cross JH, S Duncan JS, Stefan H, Theodore WH (2011a): Epilepsy imaging study guideline criteria: commentary on diagnostic testing study guidelines and practice parameters. *Epilepsia* 52:1239–1245.
- Gaillard WD, Duke ES, Berl MM, Ritzl EK, Miranda S, Liew C, Fingersh A, Dustin I, Sato S, Theodore WH (2011b). fMRI language dominance and FDG-PET hypometabolism. *Neurology* 76:1322–1329.
- Guillen MR, Adjouadi M, Bernal B, Ayala M, Barreto A, Rishe N, Lizarraga G, You X, Gaillard W (2009): A knowledge-based database system for visual rating of fMRI activation patterns for brain language networks. In: *The Fifth Richard Tapia Celebration of Diversity in Computing Conference: Intellect, Initiatives, Insight, and Innovations, TAPIA '09*, New York, NY: ACM. pp 1–6.
- Hunter KE, Blaxton TA, Bookheimer SY, Figlozzi C, Gaillard WD, Grandin C, Anyanwu A, Theodore WH (1999): (15)O water positron emission tomography in language localization: A study comparing positron emission tomography visual and computerized region of interest analysis with the Wada test. *Ann Neurol* 45:662–665.
- Jenkinson M, Bannister P, Brady M, Smith S (2002): Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17:825–841.
- Khedr EM, Hamed E, Said A, Basahi J (2002): Handedness and language cerebral lateralization. *Eur J Appl Physiol* 87:469–473.
- Knecht S, Drager B, Deppe M, Bobe L, Lohmann H, Floel A, Ringelstein EB, Henningsen H (2000): Handedness and hemispheric language dominance in healthy humans. *Brain* 123:2512–2518.
- Kurthen M, Helmstaedter C, Linke DB, Hufnagel A, Elger CE, Schramm J (1994): Quantitative and qualitative evaluation of patterns of cerebral language dominance. An amobarbital study. *Brain Lang* 46:536–564.
- Lahlou M, Guillen MR, Adjouadi M, Gaillard WD (2006): An online web-based repository site of fMRI medical images and clinical data for childhood epilepsy. In: *The 11th World Congress on Internet in Medicine, Mednet, Ontario, Canada*, pp 120–127.
- Mbwana J, Berl MM, Ritzl EK, Rosenberger L, Mayo J, Weinstein S, Conry JA, Pearl PL, Shamim S, Moore EN, Sato S, Vezina LG, Theodore WH, Gaillard WD (2009): Limitations to plasticity of language network reorganization in localization related epilepsy. *Brain* 132:347–656.
- Medina LS, Bernal B, Dunoyer C, Cervantes L, Rodriguez M, Pacheco E, Jayakar P, Morrison G, Ragheb J, Altman NR (2005): Seizure disorders: Functional MR imaging for diagnostic evaluation and surgical treatment—Prospective study. *Radiology* 236:247–253.
- Pouratian N, Bookheimer SY, Rex DE, Martin NA, Toga AW (2002): Utility of preoperative functional magnetic resonance imaging for identifying language cortices in patients with vascular malformations. *J Neurosurg* 97:21–32.
- Price CJ, Devlin JT, Moore CJ, Morton C, Laird AR (2005): Meta-analyses of object naming: Effect of baseline. *Hum Brain Mapp* 25:70–82.
- Price CJ, Crinion J, Friston KJ (2006): Design and analysis of fMRI studies with neurologically impaired patients. *J Magn Reson Imaging* 23:816–826.
- Pujol J, Deus J, Losilla JM, Capdevila A (1999): Cerebral lateralization of language in normal left-handed people studied by functional MRI. *Neurology* 52:1038–1043.
- Ramsey NF, Sommer IE, Rutten GJ, Kahn RS (2001): Combined analysis of language tasks in fMRI improves assessment of hemispheric dominance for language functions in individual subjects. *Neuroimage* 13:719–733.

Rasmussen T, Milner B (1977): The role of early left-brain injury in determining lateralization of cerebral speech functions. *Ann NY Acad Sci* 299:355–369.

Risse GL, Gates JR, Fangman MC (1997): A reconsideration of bilateral language representation based on the intracarotid amobarbital procedure. *Brain Cogn* 33:118–132.

Rorden, C., Brett, M (2000): Stereotaxic display of brain lesions. *Behav Neurol* 12:191–200.

Sabsevitz DS, Swanson SJ, Hammeke TA, Spanaki MV, Possing ET, Morris GL 3rd, Mueller WM, Binder JR (2003): Use of preoperative functional neuroimaging to predict language deficits from epilepsy surgery. *Neurology* 60:1788–1792.

Smith SM (2002): Fast robust automated brain extraction. *Hum Brain Mapp* 17:143–155.

Spreer J, Arnold S, Quiske A, Wohlfarth R, Ziyeh S, Altenmuller D, Herpers M, Kassubek J, Klisch J, Steinhoff BJ, Honegger J, Schulze-Bonhage A, Schumacher M (2002): Determination of hemisphere dominance for language: Comparison of frontal and temporal fMRI activation with intracarotid amygdala testing. *Neuroradiology* 44:467–474.

Springer JA, Binder JR, Hammeke TA, Swanson SJ, Frost JA, Bellgowan PS, Brewer CC, Perry HM, Morris GL, Mueller WM (1999): Language dominance in neurologically normal and epilepsy subjects: A functional MRI study. *Brain* 122:2033–2046.

Suma H, Murali S (2007): Principal component analysis and classification of fmri activation maps. *Int J Comput Netw Secur* 7:235–242.

Szaflarski JP, Binder JR, Possing ET, McKiernan KA, Ward BD, Hammeke TA (2002): Language lateralization in left-handed and ambidextrous people: fMRI data. *Neurology* 59:238–244.

Viera AJ, Garrett JM (2005): Understanding inter-observer agreement: The kappa statistic. *Fam Med* 37:360–363.

Viviani R, Gron G, Spitzer M (2005): Functional principal component analysis of fMRI data. *Hum Brain Mapp* 24:109–129.

Wilke M, Lidzba K (2007): Li-tool: A new toolbox to assess lateralization in functional MR-data. *J Neurosci Methods* 163:128–136.

Wilke M, Schmithorst VJ (2006): A combined bootstrap/histogram analysis approach for computing a lateralization index from neuroimaging data. *Neuroimage* 33:522–530.

Woermann FG, Jokeit H, Luerding R, Freitag H, Schulz R, Guertler S, Okujava M, Wolf P, Tuxhorn I, Ebner A (2003): Language lateralization by Wada test and fMRI in 100 patients with epilepsy. *Neurology* 61:699–701.

Woods RP, Dodrill CB, Ojemann GA (1988): Brain injury, handedness, and speech lateralization in a series of amobarbital studies. *Ann Neurol* 23:510–518.

Woolrich MW, Ripley BD, Brady M, Smith SM (2001): Temporal autocorrelation in univariate linear modeling of FMRI data. *Neuroimage* 14:1370–1386.

Worsley K, Evans A, Marrett S, Neelin P (1992): A three-dimensional statistical analysis for CBF activation studies in human brain. *J Cerebr Blood Flow Metab* 12:900.

Wyllie E, Naugle R, Chelune G, Luders H, Morris H, Skibinski C (1991): Intracarotid amobarbital procedure: II. Lateralizing value in evaluation for temporal lobectomy. *Epilepsia* 32:865–869.

You X, Adjouadi M, Guillen MR, Ayala M, Barreto A, Rishe N, Sullivan J, Dlugos D, VanMeter J, Morris D, Donner E, Bjornson B, Smith ML, Bernal B, Berl M, Gaillard WD (2011): Sub-patterns of language network reorganization in pediatric localization related epilepsy: A multisite study. *Human Brain Mapp* 32:784–799.

**TABLE A.I. Postsurgical data for LRE participants**

Hemisphere	Lobe/Region	Number	Disagreements
Bilateral	Frontal	1	
Left	Frontal	3	
	Temporal	22	4 (partial; PCA-L Group1)
Right	Occipital	1	1 (partial; PCA-L Group1)
	Hemispherectomy	1	
	Anatomical	1	
	Hemispherectomy		
	Frontal	5	1 (overt; PCA-L Group1; path multifocal Palmini type IIA FCD)
	Temporal	6	2 (partial; PCA-L Group1)
	Parietal	1	
Hemispherectomy		1	1 (partial; PCA-L Group 1; dysplasia)
	Strip/Grid	1	
	Rolandic	1	

**APPENDIX A: POSTSURGICAL DATA FOR LRE PARTICIPANTS**

Postsurgical data for LRE participants reveal that out of 58 LRE participants, 14 subjects did not have surgery. The 44 patients had surgery on different regions of the brain as summarized in Table A.I, including 28 on left hemisphere, 15 on right hemisphere, and 1 on bilateral frontal. For three cases with some form of disagreement (two overt and one partial), surgical data are not available. For the ones we do have data, the only case of overt disagreement had surgery on the right frontal lobe, and it is of path multifocal Palmini type IIA FCD. Among the eight partial disagreements, six cases had surgery on the temporal lobe (two on the right and four on the left), one on the left occipital lobe, and one had right hemispherectomy (who was diagnosed with dysplasia). No patients had gross postoperative aphasia on bedside testing. We do not have formal neuropsychological language testing on the large majority of patients.

Postoperative outcome measures would be another source of information to examine the cases of disagreement, but this is also flawed. The objective of preoperative fMRI is to avoid postoperative naming deficits beyond what one would expect on operating on a dominant hemisphere [Sabsevitz, 2003]. Findings from fMRI change practice [Medina, 2005], for example either by not pursuing surgery, use of invasive mapping, or altered resection based on fMRI data. Part of the motivation of this project is to gather the data and develop the methods necessary to fund a larger scale outcome project with all the technical challenges that ensue as there are no outcome studies using fMRI in children with epilepsy.

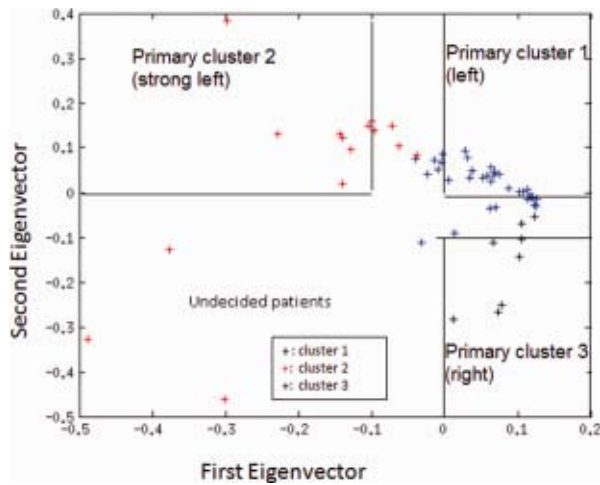


Figure B1.

New PCA decisional space using only the patient population. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

### APPENDIX B: STABILITY OF THE PCA DECISIONAL SPACE

The results given in this appendix are to demonstrate the stability of the PCA decisional space derived using both patient and control populations, by using this same decisional space but only on the patient population. Thus after applying a similar PCA process on the LRE patients only, the newly generated space, as shown in Figure B.1, differed slightly from the above-reported decisional space which included the control population.

Using the same primary clusters as before, the results summarized in Table B.I show two fewer overt disagreements between PCA and the conventional methods. Three patients changed from classification as left (“1”) to right (“3”), yielding greater agreement with either LI or visual rating; one patient changed from strong left (“2”) to left (“1”) which maintained the agreement with conventional ratings.

TABLE B.I. Comparison of different PCA space—all subjects vs. patients only<sup>a</sup>

PCA cluster	All subjects	LI fixed	LI bootstrap	Rater 1	Rater 2	Rater 3	PCA cluster Patients only
1	3	3	3	3	3	0	3
2	1	1	1	1	1	1	1
1	3	1	0	0	0	3	3
1	3	2	3	3	3	3	3

<sup>a</sup>PCA: left (“1”), right (“3”), strong left (“2”); LI and VR: left (“1”), right (“3”), bilateral (“2”), and others (“0”).

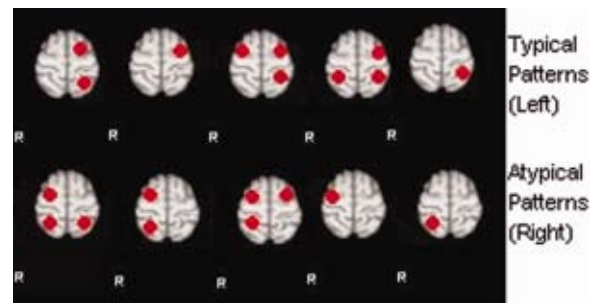


Figure C1.

Standard synthetic activation patterns. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

### APPENDIX C: SYNTHETIC DATA GENERATION AND CLASSIFICATION

Synthetic activation patterns were generated based on the masks on Broca’s and Wernicke’s areas in our standard space (61 × 73 × 61). According to the combinations of the different activations in the left and right Broca areas, and left and right Wernicke areas, we generated 10 standard patterns as shown in Figure C.1. In accordance with the definition of typical and atypical activations, the top five patterns were regarded as typical (left dominant), and the bottom five were considered atypical (right dominant).

Based on the different characteristics of these activation patterns, activations in the volume are generated randomly in accordance to the following rules: (i) each activation satisfies the Gaussian distribution; (ii) the center of activation must be within the ROI; (iii) the covariance matrix is randomly generated, and (iv) the number of activations is randomly chosen from 1 to 10.

The procedure for generating synthetic data consists of the following steps:

1. Generate related parameters following the set rules.
2. Resample the original volume size into 244 × 292 × 244.
3. Generate randomly points in the resampled volume (2,500 points in this case are generated for each activation pattern).
4. Accumulate the number of points on each voxel as the value of that voxel.

TABLE C.I. Synthetic data classification results<sup>a</sup>

Synthetic category	Cluster 1 (–)	Cluster 2 (–)	Cluster 3 (+)
Left (typical)–	84	1	15
Right(atypical)+	1	0	99

<sup>a</sup>sensitivity = 99%, specificity = 85%, accuracy = 92%, and precision = 86.8%.



5. Resample the volume back to its original size  $61 \times 73 \times 61$ .
6. Apply the mask to obtain the synthetic data.

Following these steps, 100 left dominant patterns and 100 right dominant patterns were generated for the classification process. After generating the synthetic data, based on the intensity of the real data, we first normalize it between 0 and 1, and then multiply by a random number centered at 4 with variance of 2. Then similarly to the clustering steps in our previous study, we project each synthetic data  $\Phi_{\text{new}}$  using

the relation  $U^T \Phi_{\text{new}}$  onto the primary clusters ( $U$ —the eigenvectors from our real subjects' PCA space). Then use the same distance method to classify the synthetic data.

After projecting the synthetic data onto the PCA decisional space, the classification results obtained are as summarized in Table C.I. Right dominant cases are considered atypical (positive), and left dominant cases are considered typical (negative). In these results, the reason sensitivity is relatively lower is owing to the original real data sample which had a much smaller amount of atypical (right dominant) cases in our initial PCA decisional space.