NEUROIMAGE **8,** 283–301 (1998)
ARTICLE NO. NI980370

Influence of ANOVA Design and Anatomical Standardization on Statistical Mapping for PET Activation

Michio Senda,* Kenji Ishii,* Keiichi Oda,* Norihiro Sadato,† Ryuta Kawashima,‡ Motoaki Sugiura,‡ Iwao Kanno,§ Babak Ardekani,§ Satoshi Minoshima,¶ and Itaru Tatsumi‖

*Positron Medical Center, Tokyo Metropolitan Institute of Gerontology, Tokyo 173-0015, Japan; †Biomedical Imaging Research Center, Fukui Medical School, Fukui, Japan; ‡Institute of Development, Aging, and Cancer, Tohoku University, Sendai, Japan; §Department of Radiology, Research Institute for Brain and Blood Vessels, Akita, Japan; ¶Division of Nuclear Medicine, University of Michigan, Ann Arbor, Michigan; and ‖Department of Language and Cognition, Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan

Received March 20, 1998

We have created images of *z* value, error, and variation components for a PET activation study using various ANOVA designs and anatomical standardization methods. Data were acquired in four PET centers. In each center, CBF was measured on six normal male subjects under resting and covert verb generation, three times for each. The images were anatomically standardized with LINEAR transformation, SPM (Ver. 95), HBA (Karolinska/Tohoku), or MICHIGAN (Minoshima). ANOVA was performed pixel by pixel to compute *t* (and *z*) for the task main effect (Verb vs Rest) in four different designs: (i) two way (subject and task) (2W), (ii) two-way with interaction (2WI), (iii) subject considered a random factor (2WI-MX), and (iv) three-way (subject, task, and replication) (3W). A large area extending from the Broca to the left premotor cortex was activated. The localization of the highest peak depended both on the anatomical standardization and on the ANOVA design, the variation ranging 3–4 cm. Smoothing reduced the variation while erasing possible subfoci. The *z* images of 2W, 2WI, and 3W looked alike, whereas 2WI-MX presented lower peak *z* values. SPM tended to present higher *z* values than the other methods. The error was high in the gray and low in the white matter. The root mean square for the subject effect was high on the border of gray matter especially in LINEAR and HBA, revealing intersubject mismatch in the gray matter distribution. The root mean square for the subject-by-task interaction effect revealed individual variation in activation. © 1998 Academic Press

INTRODUCTION

In a PET activation analysis on a group of normal subjects, analysis of variance (ANOVA) or related techniques such as analysis of covariance (ANCOVA) are usually employed to create a statistical map of activation (*t* or *z* map) after an anatomical standardization for intersubject averaging. The process of *anatomical standardization,* also called *spatial normalization,* transforms the brain images of each subject into a standard brain, for which an atlas is provided, so that the regional cerebral blood flow (CBF) in the corresponding pixels of all subjects could be analyzed statistically.

In the case of task replications within subjects, in which CBF is measured twice or more under the same condition (task) within subject, a number of ANOVA designs are applicable with different definitions of *t* and error.

Let X_{ijk} be the CBF for subject i, task j, and replication k and X'_{ijk} represent CBF adjusted for the global value (gCBF) G_{ijk}, e.g., $X'_{ijk} = X_{ijk}/(G_{ijk}/50)$. Most previous reports of PET activation with task replications, although not explicitly stated, used a simple two-way design (Rumsey *et al.,* 1997; Grabowski *et al.,* 1996; Herbster *et al.,* 1997; Wessel *et al.,* 1997). Their models appear to be either

$$X'_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \quad (1)$$

or

$$X_{ijk} = \mu + \alpha_i + \beta_j + \delta G_{ijk} + \varepsilon_{ijk}, \quad (2)$$

in which μ is the grand mean (intercept), α_i is the subject effect, β_j is the task effect, and ϵ_{ijk} represents an independently and identically distributing normal error with mean zero and an unknown variance. The effect of gCBF is removed either by dividing the CBF values by gCBF to make X'_{ijk} or by ANCOVA treating gCBF as a covariate as shown in Eq. (2). Although Friston *et al.* (1990) reported that ANCOVA better described the relationship between regional and global CBF, subsequent studies showed that both methods worked equally well in stabilizing the variation (Arndt *et al.,* 1996; McIntosh *et al.,* 1996). In any case, recent

1053-8119/98 $25.00

investigators usually do not measure the arterial input function, and the measured global value no longer represents the actual global CBF.

Woods and colleagues proposed a three-way ANOVA for PET activation analysis with task replications and pointed out that it can theoretically reveal interactions between subject and replication as well as between task and replication (Woods, 1996; Woods *et al.*, 1996),

$$X'_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \varepsilon_{ijk}, \quad (3)$$

in which γ_k is the replication effect and $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, and $(\beta\gamma)_{jk}$ are the interaction effects of subject-by-task, subject-by-replication, and task-by-replication, respectively. (They should not be interpreted as the product of α and β, etc.)

A recent study (Poline *et al.*, 1996a) used a statistical design available in SPM (Ver. 96),

$$X_{ih} = \alpha_i + \lambda_h + \delta G_{ih} + \varepsilon_{ih}, \quad (4)$$

in which X_{ih} is the CBF for scan h of subject i, λ_h is the scan effect, and half of h's belong to task1 and the other half belong to task2. To test the significant difference between task1 and task2, they computed the t statistic of the contrast:

$$C = \sum_{h\in\text{task1}} \lambda_h - \sum_{h\in\text{task2}} \lambda_h. \quad (5)$$

The statistical design is supposed to be built based on a hypothesis by the investigator before the data are analyzed. However, the difference in the philosophy underlying each ANOVA/ANCOVA design is so subtle that few investigators seem to define their design intentionally. Some investigators do not even describe specifically which ANOVA/ANCOVA design they use. It will be a serious matter if the result depends heavily on the ANOVA design. In fact, very few reports so far have looked into how the result of an activation study depends on the ANOVA/ANCOVA design.

The error term ϵ_{ijk} reflects the variations that are not explained by the ANOVA/ANCOVA model. There has been a controversy as to whether the error is heterogeneous across the brain (Friston *et al.*, 1991) or can be treated as homogeneous (Worsley *et al.*, 1992). Ruttimann *et al.* (1997) proposed a variance stabilizing transform assuming that the error is a function of regional radioactivity. Although numerous reports that apply the PET activation technique to neuroscience have been published, few have addressed this issue or presented the error images (Arndt *et al.*, 1995). The homogeneity of error image may depend on the task paradigm or on the ANOVA/ANCOVA design. The z map and error images also depend on the method of anatomical standardization because it influences the intersubject variation of the data.

Another aspect of ANOVA is that it divides the total variation into components, each of which is defined as an effect in the ANOVA design and is explicitly stated in the equation. The subject effect (α_i) represents the between-subject variation of CBF for each pixel. It is considered to reflect the mismatch of gray matter distribution between subjects in standardized images and may indicate the pertinence of the anatomical standardization method. The subject-by-task interaction effect $(\alpha\beta)_{ij}$ represents the intersubject variation in the amount of activation that is not accounted for by the anatomical standardization. It may provide valuable information for interpretation of z images and, therefore, about the effectiveness of the anatomical standardization as well. Although few investigators have examined the variation components in detail or presented the images, it can be obtained in the course of ANOVA computation as a mean square, which is the sum of the square divided by the degree of freedom for the effect. The root of the mean square (RMS) assumes the same unit as X'_{ijk} and reflects the size of the component.

In the present study, we have analyzed PET activation data on the verb generation paradigm using four types of ANOVA designs and using four different methods of anatomical standardization (LINEAR, SPM, HBA, and MICHIGAN). We have created the z maps and error images and examined how they depend on the ANOVA design and on the anatomical standardization method. We have also created RMS images for the subject effect (RMS-S) and for the subject-by-task interaction effect (RMS-ST) and examined the effect of anatomical standardization on those variation component images. As for the adjustment for the global value, we have used the approach of $X'_{ijk} = X_{ijk}/(G_{ijk}/50)$ to simplify the equation in the present study. As stated above, whether to use this or the ANCOVA approach is not a significant issue.

This study has been conducted as part of the Japanese Multicenter PET Project.

MATERIALS AND METHODS

Data Acquisition

Data were acquired independently in four institutes in Japan under the same protocol: Tokyo Metropolitan Institute of Gerontology (Tokyo), Fukui Medical School (Fukui), Tohoku University and Japan Radioisotope Association Nishina Memorial Cyclotron Center (Sendai), and Research Institute for Brain and Blood Vessels–Akita (Akita).

In each institute, six young right-handed normal male subjects (ages 19–29) were recruited locally. All

the subjects were free from previous or current neurological disorders and were strongly right-handed as assessed by Edinburgh's handedness inventory (Oldfield, 1971).

The CBF was measured with bolus or slow bolus intravenous injection of $H_2{}^{15}O$ under the resting condition (Rest) and while the subject tried to think of as many verbs as possible that are associated with auditorily presented nouns without overtly pronouncing them (Verb), three times for each in the alternating order (RVRVRV or VRVRVR). The nouns were selected from the standard pictures of Snodgrass (Snodgrass and Vanderwart, 1980) and were given at a rate of one noun per 6 s.

The radioactivity images were acquired for 60 (Fukui and Sendai), 90 (Akita), or 120 s (Tokyo); starting at the injection (Tokyo), at 10 s after injection (Akita), or at the arrival of radioactivity in the brain (Fukui and Sendai). The injected dose was approximately 1300 (Sendai), 1000 (Tokyo), 370 (Fukui), or 250 MBq (Akita). The photon attenuation was corrected for with measured transmission data. The images were converted into CBF images with PET-ARG method (Herscovitch *et al.*, 1983; Kanno *et al.*, 1987) using measured arterial time course (Tokyo) or standard input function of the institute (Akita, Fukui, Sendai).

Tokyo and Sendai used a 2D PET scanner HEADTOME-IV (Shimadzu, Japan) (Iida *et al.*, 1989), which provided 14 slices of images with a center-to-center interval of 6.5 mm and with spatial resolution of 7.5 mm and axial resolution of 10 mm FWHM. Akita used a 3D scanner HEADTOME-V (Shimadzu, Japan) (Iida *et al.*, 1996), providing 63 slices with an interval of 3.125 mm, having an isotropic spatial resolution of 6 mm in effective FWHM in reconstructed images. Fukui used a 3D scanner, GE-ADVANCE (GE, Milwaukee, WI) (De Grado *et al.*, 1994; Lewellen *et al.*, 1996), with the interslice septa being retracted, acquiring 35 slices of images with an interslice spacing of 4.25 mm and with spatial resolution of 6 mm and axial resolution of 10-mm FWHM.

T1-weighted transaxial MRI of each subject was also acquired for anatomical reference. The center-to-center slice interval was 0.9 mm for Fukui, 1.5 mm for Sendai, and 3.0 mm for Akita and Tokyo.

The protocol has been approved by the ethics committee or its equivalent at each institute.

Anatomical Standardization

The CBF images of each subject were checked for motion between scans by the realign program in SPM (Ver. 95) and by inspection of subtracted images and were realigned if a motion of >1 mm or >1° was detected or if a streak was observed along the brain surface in the subtraction images. Significant motion was not detected for any subject imaged in Tokyo or Sendai, where customized headholders were routinely used to secure the fixation so that the already small axial field of view would not be curtailed by realignment.

The six CBF images were averaged within each subject and the average CBF image was registered to the subject's MRI using the registration method proposed by Ardekani *et al.* (1995). The resulting fusion images were inspected and the accuracy of registration was confirmed by one of the authors (K.I.), an expert in brain PET and MRI interpretation. The registration method combined with visual inspection provided an average error of <2 mm. The images were then anatomically standardized with each of the four methods: LINEAR, SPM (Ver. 95), HBA, and MICHIGAN.

In LINEAR, the anterior commissure (AC) and posterior commissure (PC) were identified and the brain size was measured on the MRI to compute the transformation parameters that would reorient the MRI to the AC–PC line and scale them to Talairach's atlas (Talairach and Szikla, 1967) with a different scaling factor for each axis (Fox *et al.*, 1985). This parameter set and that for the PET–MRI registration were used to transform the PET images into Talairach's space (Senda *et al.*, 1994). The standardized images had 25 slices with 4-mm interslice spacings and a pixel size of 2 mm.

In SPM, the AC–PC line was automatically estimated from the PET without using MRI (Friston *et al.*, 1989), and the PET images were transformed into Talairach's space (Talairach and Tournoux, 1988) using a 12-parameter affine transformation and a 6-parameter 3D quadratic deformation (Friston *et al.*, 1995). The procedure minimizes the sum of squares between the subject PET and the template PET. The standardized images had 26 slices with 4-mm interslice spacings and a pixel size of 2 mm.

In HBA, the computerized brain atlas (Roland *et al.*, 1994; Roland and Zilles, 1994) was processed interactively by a trained operator in Tohoku University with multiple linear and nonlinear transformations so that the brain contour, ventricles, and major sulci defined in the atlas would match the structures identified in each subject's MRI. The PET images that had been registered to the MRI were transformed into the atlas space with the reverse transformations. The standardized images had 38 slices with 4-mm interslice spacings and a pixel size of 2 mm.

In MICHIGAN, the subject PET image was reoriented and scaled using the AC–PC and the brain size that were automatically estimated on the PET. The cerebral cortex and the central gray matter were stretched on hundreds of predefined lines along major neuronal fiber bundles so that the profile could match that in the standard normal FDG image, which had been created based on Talairach's atlas (Talairach and Tournoux, 1988). Thin-plane spline interpolation with

trilinear resampling was employed to fill out the pixels between the matched points (Minoshima *et al.*, 1994a,b). The standardized images had 60 slices with 2.25-mm uniform voxel size.

It should be noted that MRI was used in LINEAR and in HBA, but was not used in SPM or in MICHIGAN.

Statistical Analysis

In the present study, statistical analysis was performed on the data set from each center (six scans by six subjects making $n = 36$) and not on the entire data of 24 subjects, because data from different centers are supposed to have different noise and spatial resolution due to difference in dose, scanning period, camera sensitivity, and various corrections. We also planned to look at the reproducibility of findings between centers, i.e., between subject samples or PET cameras. Therefore, each of the four data sets was analyzed with an identical procedure described below. The statistical computation was performed with locally developed programs which utilized IMSL/STAT library (IMSL Inc., Houston, TX), and the image processing in general was performed using DrView (Asahi-Kasei Joho System, Japan).

First, the gCBF value for each scan was obtained as the average CBF value within the brain area defined by >0.2 of the maximum value in each subject's mean CBF image. Each CBF image was divided by (gCBF)/50 to adjust for the gCBF fluctuations. The images were then smoothed with a 3D Gaussian filter of 10-mm FWHM. For the Fukui data and for the Akita data, smoothing was also performed using a 3D Gaussian filter of 15-mm FWHM and of 20-mm FWHM. These stronger smoothing filters were not applied to the other two data sets because they did not tolerate such a strong z-axis smoothing operation due to a short z-axis field of view.

For the set of six-subject data, ANOVA was performed pixel by pixel in the following four different designs to compute the F statistic for the task main effect (Verb vs Rest), and its square root ($=t$) was mapped. To make the comparison between t statistics with different degrees of freedom (df) easier, the t value was transformed into the equivalent z score of normal distribution. The mean difference of the task effect (ΔCBF) was also computed and mapped. In addition, the RMS-S (α) and the RMS-ST ($\alpha\beta$) as well as for the error (RMSE) were computed and were mapped to evaluate the components of variation.

As X'_{ijk} denotes the gCBF-adjusted CBF for subject i (=1–6), task j (=Rest, Verb), and replication k (=1–3), the ANOVA designs employed in the present study are described:

(1) Two-way ANOVA (2W)

$$X'_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk},\ F = \mathrm{MS}\beta/\mathrm{MSE},\ df = 29; \quad (6)$$

(2) Two-way ANOVA with interaction (2WI)

$$X'_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk},\ F = \mathrm{MS}\beta/\mathrm{MSE},\ df = 24; \quad (7)$$

(3) Two-way ANOVA with interaction, using a mixed-effects design (2WI-MX)

$$X'_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk},\ F = \mathrm{MS}\beta/\mathrm{MS}(\alpha\beta),\ df = 5; \quad (8)$$

(4) Three-way ANOVA with two-factor interactions (3W)

$$X'_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \varepsilon_{ijk},\ F = \mathrm{MS}\beta/\mathrm{MSE},\ df = 10. \quad (9)$$

The 2W, 2WI, and 3W are fixed-effects designs because all the factors ("subject," "task," and "replication") are treated as fixed factors. On the other hand, the 2WI-MX is a mixed-effects design because the "subject" is treated as a random factor while the task is treated as a fixed factor.

Strictly, the subject is considered a random factor because the subjects in the present study are not specific subjects but are sampled from a population. The subject effect reflects the intersubject variation in the population and is treated as a random variable from normal distribution with an unknown variance, whereas a fixed effect is expressed by an unknown constant. In 2WI-MX, MSβ becomes equal to MS($\alpha\beta$), not to MSE, under the null hypothesis of no task effect ($\beta_{\mathrm{Rest}} = \beta_{\mathrm{Verb}}$) (Kleinbaum *et al.*, 1988, pp. 416–477). Although this approach may not be appropriate for a small set of subjects, we investigated how this would influence the z map in the two-way ANOVA with interaction in the present study.

A similar viewpoint applies to the three-way ANOVA design. The "replication" may be considered either a random factor (three randomly selected measurements from a large number of trials) or a fixed factor (three specific trials). The latter is more acceptable from the standpoint of neuroscience because habituation would occur when a subject undergoes the same task three times. If subject is a random factor and task and replication are fixed factors in 3W, MSβ becomes equal to MS($\alpha\beta$), not to MSE, under the null hypothesis of no task effect ($\beta_{\mathrm{Rest}} = \beta_{\mathrm{Verb}}$) (Winer, 1971). Because our ANOVA design is balanced (one observation for each set of i, j, k), MSβ for 3W is equal to that for 2W (and also 2WI), and MS($\alpha\beta$) for 3W is equal to that for 2WI. (Note that MSE is different between 2W, 2WI, and 3W.) Therefore, 3W-MX with the subject being treated as a

random factor would give the same z map as the 2WI-MX.

In the present study, 2W, 2WI, and 3W are fixed-effects models with their effects being nested with each other. In such a case, significance of the additional effects in 2WI compared to 2W (i.e., subject-by-task interaction effect) can be evaluated with a partial F (Kleinbaum *et al.*, 1988, pp. 126–137). Accordingly, F (2W vs 2WI) = (difference in SSE between 2W and 2WI)/(difference in df between 2W and 2WI)/(MSE of 2WI) was computed pixel by pixel and was mapped. Similarly, F (2WI vs 3W) was computed and mapped, which is a measure of significance for the additional effects regarding the replication. Note degree of freedom is (5, 24) for F (2W vs 2WI) and (14, 10) for F (2WI vs 3W).

Evaluation of Smoothness

Smoothness should be taken into account when the above-mentioned statistics are interpreted. The smoothness represents a correlation between neighboring pixels and is essential for estimating the P value from the z value based on the stochastic model (Friston *et al.*, 1991; Worsley *et al.*, 1996a). Smoothness exists in the original PET images according to the finite spatial resolution, which depends on the camera and the reconstruction filter, and is added to by subsequent smoothing operations. Because an interpolation, having a smoothing effect, is inevitably involved in an image transformation process, smoothness may depend on the anatomical standardization method.

Smoothness is estimated from the variance of the partial derivative of the statistic image that can be treated as a Gaussian random field (Friston *et al.*, 1991). Because the activation z map satisfies the condition only under the null hypothesis of no activation, Worsley *et al.* (1992) showed that it is better to estimate the smoothness on the residual images standardized by the pooled standard deviation under the assumption of uniform error across the brain. The present study, however, does not assume uniformity of the error, but investigates its distribution. Therefore, in the present study, the residual was standardized by RMSE $\times$ $(1 - \text{leverage})^{1/2}$ for each pixel to create the standardized (studentized) residual image, which follows a Student t distribution if the ANOVA design is correct (Kleinbaum *et al.*, 1988, pp. 186–187). The standardized residual images were Gaussianized, from which the smoothness was computed. Because the smoothness is theoretically independent of the ANOVA design, it was estimated on the residual for the 3W design. Considering the wide interslice spacing, the smoothness was estimated two-dimensionally in each slice from ACPC+0 to +36 mm or within the corresponding slice range.

RESULTS

Figure 1 illustrates the mean CBF images of six subjects acquired in Fukui that have been anatomically standardized with each of the four methods (LINEAR, SPM, HBA, and MICHIGAN). The MICHIGAN images, having a different pixel size, were scaled to the other methods for display. Because of differences in the standard brain and in the interslice spacing between the standardization methods, roughly corresponding slices were selected for each method and depicted for comparison. As a result, the slice z position from AC-PC is not necessarily the same for all methods. As shown in Fig. 1, major gray matter structures were equally delineated in the mean CBF images by each method. However, the shape of the brain was different between methods according to the standard brain. The images standardized by MICHIGAN, whose standard brain had been molded most loyally after Talairach's atlas, presented compressed lateral contours along the Sylvian fissure as they are portrayed in the atlas.

When the area activated by Verb compared with Rest was examined, the z maps computed in various ways showed an activation more or less in the left frontal cortex extending from the inferior frontal gyrus (Broca's area) upward to the middle frontal and premotor cortex. An activation was also observed in the supplementary motor area (SMA), in the anterior cingulate gyrus, and in the left superior temporal cortex (Wernicke's area). The activated area was consistent among the centers, although the Fukui data presented stronger activation than the other centers.

Figure 2 illustrates the z map computed with the 2W ANOVA design from the data of six subjects acquired in Fukui and anatomically standardized with four different methods. The z images looked similar to each other in appearance although some differences are observed by a close inspection.

Figure 3 illustrates the z map computed by four different ANOVA designs (2W, 2WI, 2WI-MX, 3W) from the data of Tokyo and Fukui anatomically standardized with SPM. 2W, 2WI, and 3W presented similar z images, 3W being a little different from 2W and 2WI. On the other hand, 2WI-MX presented substantially lower z values and different z distribution from the others.

Table 1 summarizes the highest and the second highest peaks in z and CBF difference images within the activated area extending from the Broca to the left middle frontal gyrus. The exact localization of foci peak in the z images depended both on the ANOVA design and on the anatomical standardization method, and the variation ranged as much as 3–4 cm. This variation exceeds the difference in the definition of structures in the atlas each method is based on. As for the peak values, SPM tended to present higher z values than

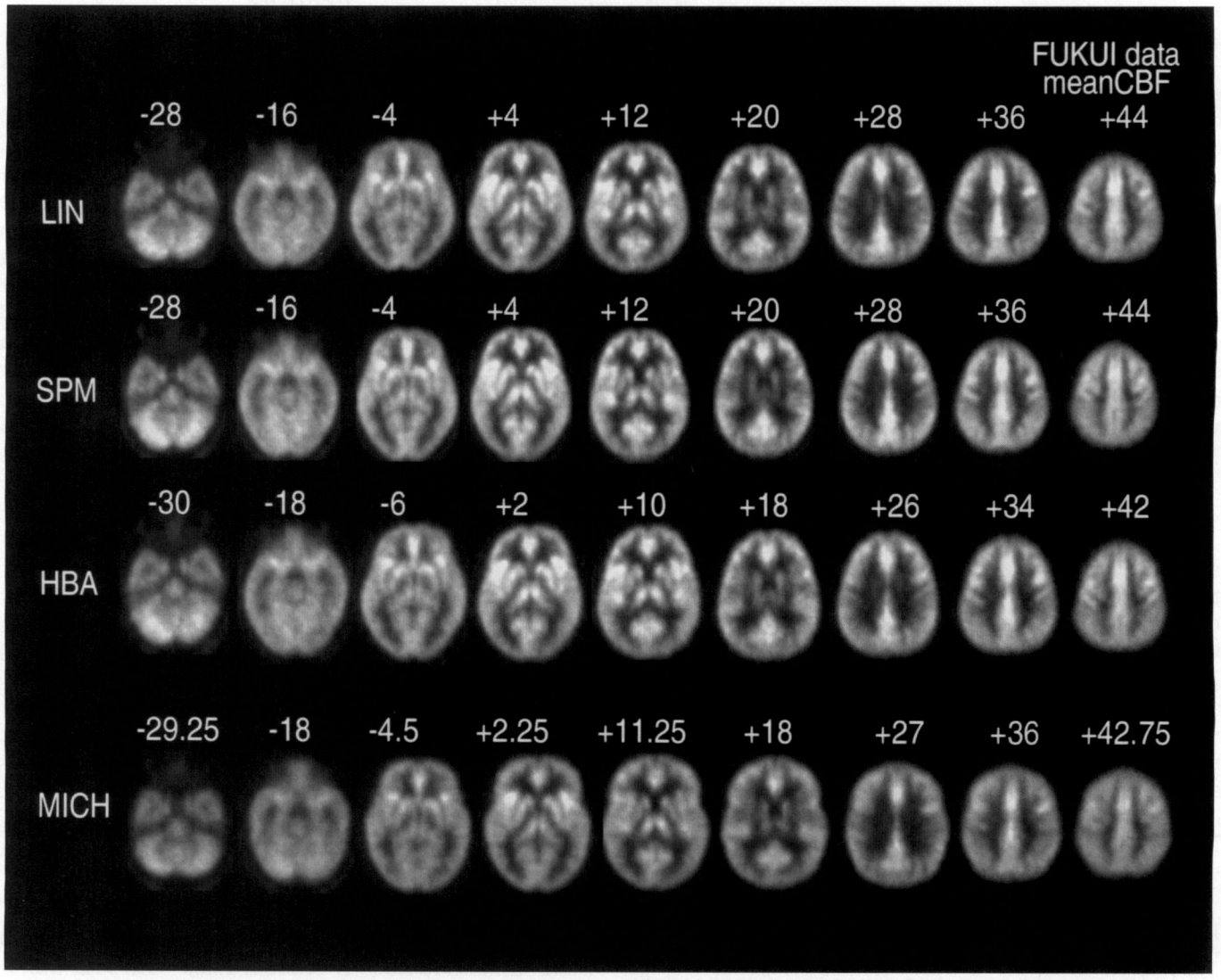

FIG. 1. Mean CBF images of six subjects acquired in Fukui that have been anatomically standardized with four different methods (LINEAR, SPM, HBA, MICHIGAN; see text for details). The slice positions are expressed in mm from AC-PC in the standard brain each method is based on. Note different appearance between standardization methods despite identical original data set.

LINEAR and HBA. The variation of the peak localization was not caused by the shift of the same peak but by the presence of several peaks within the activated area. Which subpeak became the highest depended on the standardization method and on the ANOVA design.

Figure 4 illustrates the dependence of RMSE on the ANOVA design (2W, 2WI, or 3W) that were computed from identical data (Tokyo or Fukui) with standardization by SPM. The RMSE were not homogeneous across the brain, being high in the gray matter and low in the white matter, so that the gray matter structures were delineated. Within the gray, the central and deep structures including thalamus, basal ganglia, and insula tended to present higher RMSE than the surface cortical rim.

Figure 5 shows the RMS-S and how it was influenced by the standardization method (LINEAR, SPM, HBA, and MICHIGAN). The RMS-S represents the variation explained by the subject and therefore reflects the intersubject mismatch in CBF distribution. Although they were computed from identical data of Fukui, LINEAR and HBA presented higher values than SPM and MICHIGAN in general. When the images were closely inspected, high RMS-S values were observed on the border of gray matter structures such as on the brain contour, between the cortex and the subcortical white, between the cingulate gyrus and the corpus callosum, along the Sylvian fissure, and around the thalamus and basal ganglia. In LINEAR, the occipital cortex also presented high RMS-S values.

Figure 6 illustrates the RMS-ST computed from the data of six subjects acquired in each city and anatomi-

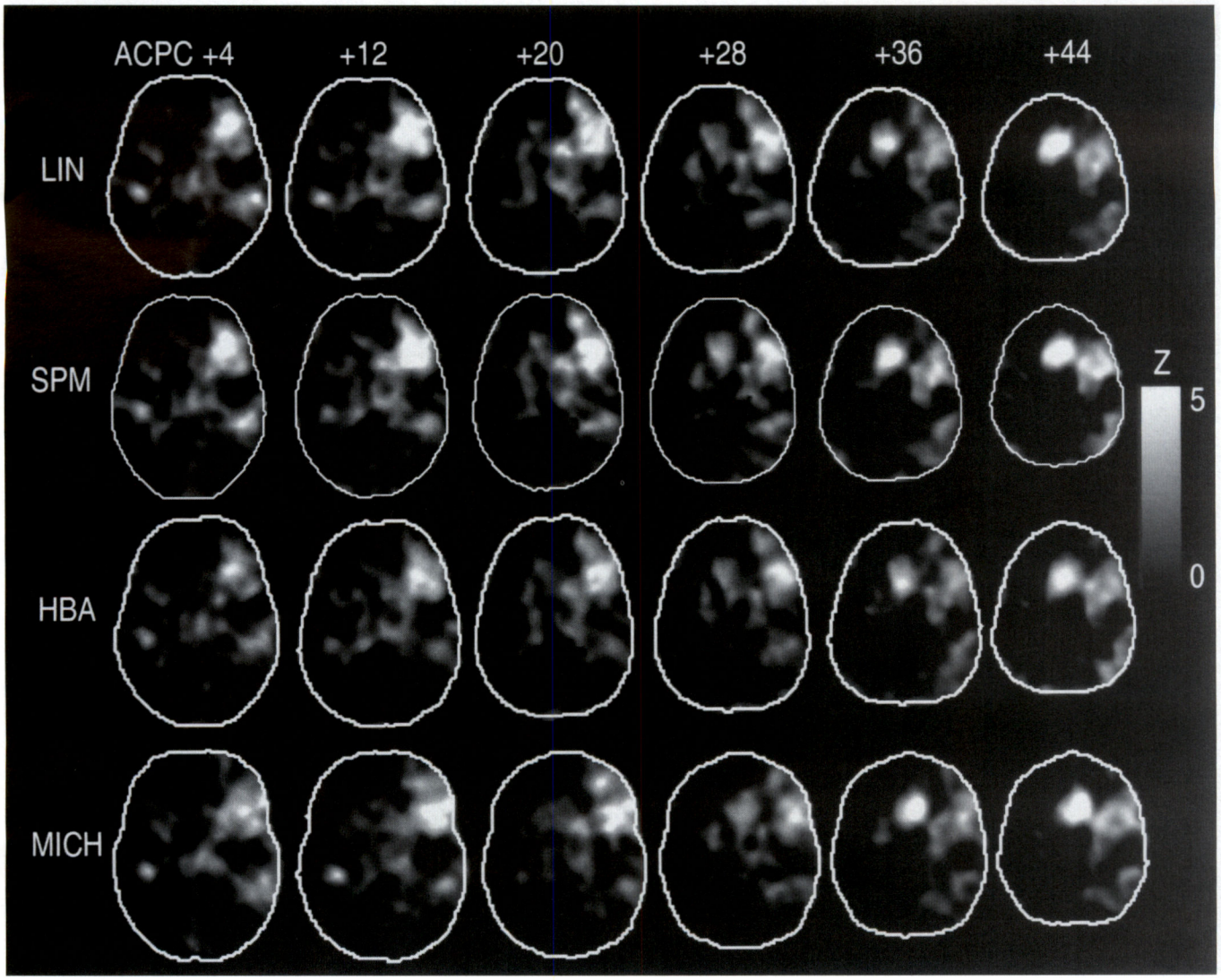

FIG. 2. The z map of verb generation vs resting based on two-way ANOVA design computed from the identical data of six subjects acquired in Fukui and anatomically standardized with four different methods (LINEAR, SPM, HBA, MICHIGAN; see text for details). Slice position labels are for LIN and SPM. For HBA and MICH, slices corresponding to those shown in Fig. 1 are depicted. The contour lines were obtained from the corresponding mean CBF images. Note difference in the peak localization between standardization methods, which is summarized in Table 1.

cally standardized with SPM. The RMS-ST, representing intersubject variation in activation, presented hot spots in the left posterior inferior frontal gyrus (Broca), anterior cingulate, and SMA. The Akita and Fukui data also showed high RMS-ST in the medial occipital cortex known as the primary visual area.

Figure 7 illustrates the F images to assess the significance of additional effects for 2W vs 2WI and for 2WI vs 3W, which were computed on the Tokyo data and on the Fukui data standardized with SPM. The F image for 2W vs 2WI looked similar to the respective RMS-ST image, the latter being the numerator of the F, and presented localized high values within or on the margin of the left frontal activated area. Hot spots were also observed in the supplementary motor area in Tokyo data and in the anterior cingulate cortex in Fukui data. On the other hand, the F image for 2WI vs 3W did not present apparent high values in or around the left frontal activation foci. Instead, high F values were noted in the occipital cortex in Fukui data and in Akita data.

To evaluate the effect of smoothing, the statistical analysis was repeated on the Fukui data and on the Akita data after smoothing the images with a Gaussian filter of 15- or 20-mm FWHM instead of 10 mm. All resultant images, including z maps, mean difference, RMSE, RMS-S, and RMS-ST, became "smoothed" in appearance when the transformed images were treated

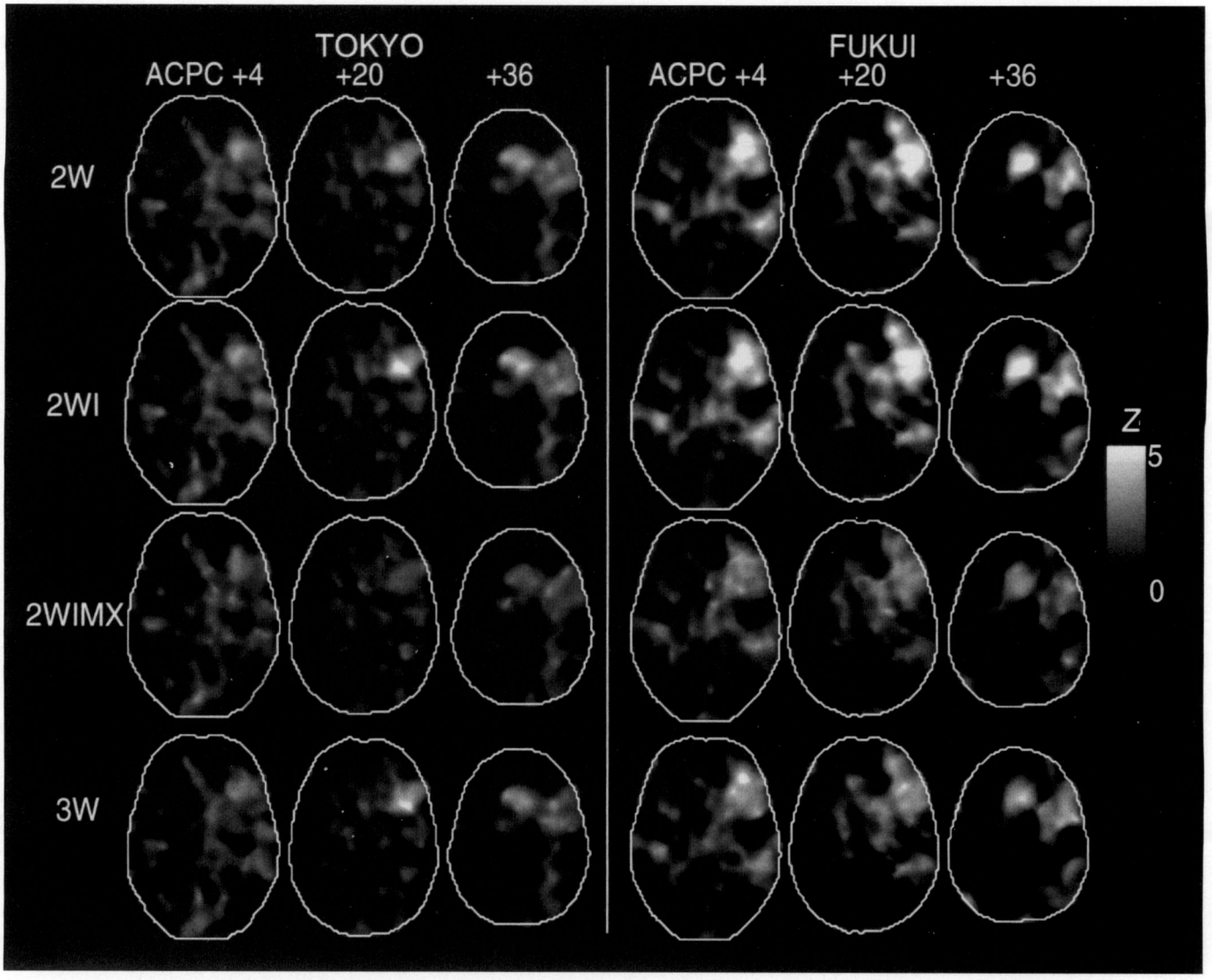

FIG. 3. The z map of verb generation vs resting based on four different ANOVA designs (2W, 2WI, 2WI-MX, 3W; see text for definitions) computed from the identical data of six subjects acquired in Tokyo (left) and Fukui (right) and anatomically standardized with SPM. The contour lines were obtained from the mean CBF images. Note difference in the peak value and localization between ANOVA designs, which is summarized in Table 1.

with stronger smoothing. In general, the values of the z maps tended to increase while those of the mean difference, RMSE, RMS-S, and RMS-ST tended to decrease with stronger smoothing. However, those images more or less retained the relative regional characteristics described so far, regardless of the degree of smoothing. The activated region covered similar cortical areas shown in Figs. 2 and 3. Table 2 summarizes the value and localization of the highest and the second highest peaks in z and CBF difference images in and around the activated area of the Fukui data with a Gaussian filter of 15- or 20-mm FWHM. The exact localization of the peak in the z images depended both on the ANOVA design and on the anatomical standardization, although the variation ranged 1–3 cm, which was smaller than the results with a 10-mm FWHM filter, shown in Table 1. The RMSE images showed the same relative heterogeneity as in Fig. 4, being high in the gray and low in the white. The RMS-S images presented relatively large values on the border of gray matter structures in the same way as in Fig. 5 although the absolute values were smaller. The different appearance in RMS-S between anatomical standardization methods were also preserved when the images were smoothed with a stronger filter. The RMS-ST images also showed findings relatively similar to those of Fig. 6.

The smoothness of the standardized residual images is expressed as effective FWHM(x) and is summarized in Table 3. The smoothness was almost the same regardless of the standardization method and regard-

TABLE 1

Dependence of Peak z Score and CBF Difference and Their Localization in and around Broca's Area upon ANOVA Design and Anatomical Standardization

ANOVA design	LINEAR				SPM				HBA				MICHIGAN			
	x	y	z	Peak	x	y	z	Peak	x	y	z	Peak	x	y	z	Peak
Tokyo data																
2W	−57	19	0	4.67	−30	18	12	4.79	−53	43	10	4.65	−26	6	20	4.19
	−31	23	12	4.34	−30	8	24	4.55	−33	35	10	4.61	−30	15	9	3.97
2WI	−53	15	4	4.86	−30	8	24	5.77	−33	35	10	4.86	−30	8	23	5.08
	−31	13	28	4.79	−30	18	12	4.80	−55	43	14	4.81	−51	17	16	4.15
2WIMX	−59	19	4	3.73	−22	8	0	3.82	−35	33	−6	4.32	−57	19	0	3.48
	−25	21	8	3.32	−26	30	8	3.51	−27	23	6	3.43	−39	17	38	3.35
3W	−57	11	0	4.68	−36	6	20	5.50	−39	25	18	5.08	−44	13	27	4.89
	−43	25	20	4.56	−38	18	16	4.93	−59	23	−6	5.01	−39	6	23	4.72
Difference	−33	21	12	6.30	−30	16	8	6.03	−31	31	6	5.96	−33	15	4	6.90
	−45	19	24	5.55	−36	6	24	4.97	−45	31	18	5.15	−48	15	18	5.83
Sendai data																
2W	−41	17	20	5.46	−26	20	4	5.82	−49	27	10	5.19	−48	26	18	5.46
	−45	23	24	5.20	−38	12	20	5.69	−41	31	18	5.10	−51	6	7	4.87
2WI	−41	17	20	5.22	−26	20	4	5.80	−43	37	18	4.94	−48	26	18	5.19
	−45	25	24	5.06	−38	14	16	5.31	−41	31	18	4.91	−51	6	7	4.67
2WIMX	−33	27	8	5.12	−32	16	0	5.22	−49	27	14	4.94	−53	24	23	4.44
	−35	21	4	4.79	−46	10	12	4.35	−33	35	2	4.72	−53	1	9	3.97
3W	−51	13	16	5.54	−24	20	4	5.19	−61	33	10	5.50	−26	17	2	4.89
	−49	27	28	4.65	−30	12	16	5.15	−49	43	26	4.61	−28	15	−2	4.87
Difference	−35	19	8	7.37	−44	12	12	7.53	−47	29	6	8.01	−46	19	18	7.62
	−47	15	12	7.32	−32	12	8	7.40	−25	57	18	4.50	−48	33	9	7.35
Fukui data																
2W	−31	35	12	6.16	−30	20	16	6.42	−29	39	6	5.38	−44	15	16	5.76
	−43	19	16	5.46	−32	6	16	5.87	−33	37	14	5.38	−37	6	18	5.74
2WI	−27	27	8	6.29	−30	18	20	6.29	−27	39	2	5.53	−37	6	20	5.60
	−31	35	8	5.87	−26	16	8	6.03	−43	37	22	5.34	−44	15	16	5.52
2WIMX	−35	37	12	4.45	−26	6	8	4.48	−47	27	34	4.87	−60	22	7	5.02
	−37	13	16	4.30	−30	22	12	4.44	−41	39	10	4.39	−57	24	11	4.96
3W	−33	35	4	5.15	−32	28	4	5.49	−37	37	30	4.73	−37	31	2	4.91
	−25	25	8	5.02	−16	6	12	5.11	−23	37	−2	4.71	−33	26	2	4.84
Difference	−43	19	12	7.92	−40	8	8	8.67	−43	29	10	8.75	−39	15	4	8.84
	−41	31	24	5.82	−32	34	12	5.17	−39	51	14	4.71	−46	10	25	7.02
Akita data																
2W	−39	17	32	5.56	−40	10	28	5.79	−41	27	26	5.41	−42	10	25	5.83
	−31	33	24	4.78	−28	14	12	4.87	−29	33	10	4.88	−30	4	27	4.91
2WI	−41	15	32	5.30	−40	12	28	5.57	−37	41	14	5.47	−44	10	27	5.70
	−43	25	28	5.22	−42	4	24	5.52	−27	35	10	5.43	−30	4	27	5.20
2WIMX	−33	15	32	4.30	−32	−2	16	3.89	−37	25	26	4.45	−51	10	23	4.45
	−37	17	28	4.15	−38	10	36	3.77	−61	29	14	3.53	−37	8	25	4.42
3W	−41	17	36	5.54	−42	0	24	5.47	−41	29	34	5.57	−44	13	29	5.88
	−47	33	24	5.16	−40	10	32	5.41	−27	29	14	5.08	−15	4	20	4.41
Difference	−39	15	36	5.53	−40	8	28	5.77	−41	25	30	5.22	−44	10	29	6.24
	−33	23	12	4.69	−32	16	8	4.66	−33	37	10	4.76	−44	24	23	5.79

Note. Analyzed from the data of six subjects acquired in Tokyo, Sendai, Fukui, and Akita. The highest and the second highest peaks are shown. Coordinates (x, y, z) are mm from AC, with positive directions toward right, anterior, and superior. Peak values are in z score for each ANOVA design and in ml/min/100 ml for CBF difference (gCBF = 50). The images were smoothed with a Gaussian of 10-mm FWHM. See text for details of ANOVA designs (2W, 2WI, 2WIMX, 3W) and anatomical standardization methods (LINEAR, SPM, HBA, MICHIGAN).

less of the data set. The effective FWHM was 10.2–11.8, 15.9–16.8, and 21.5–23.6 mm for the data that had been smoothed with a filter of 10, 15, and 20 mm, respectively, and was comparable to the respective Gaussian filter added to the original spatial resolution in quadrature (11.7-, 16.2-, and 20.9-mm FWHM for Akita and Fukui; 12.5-, 16.8-, and 21.4-mm FWHM for Sendai and Tokyo).

DISCUSSION

Foci Localization

In the present study, we used various anatomical standardization methods and various ANOVA designs on identical data sets to create z maps and compared

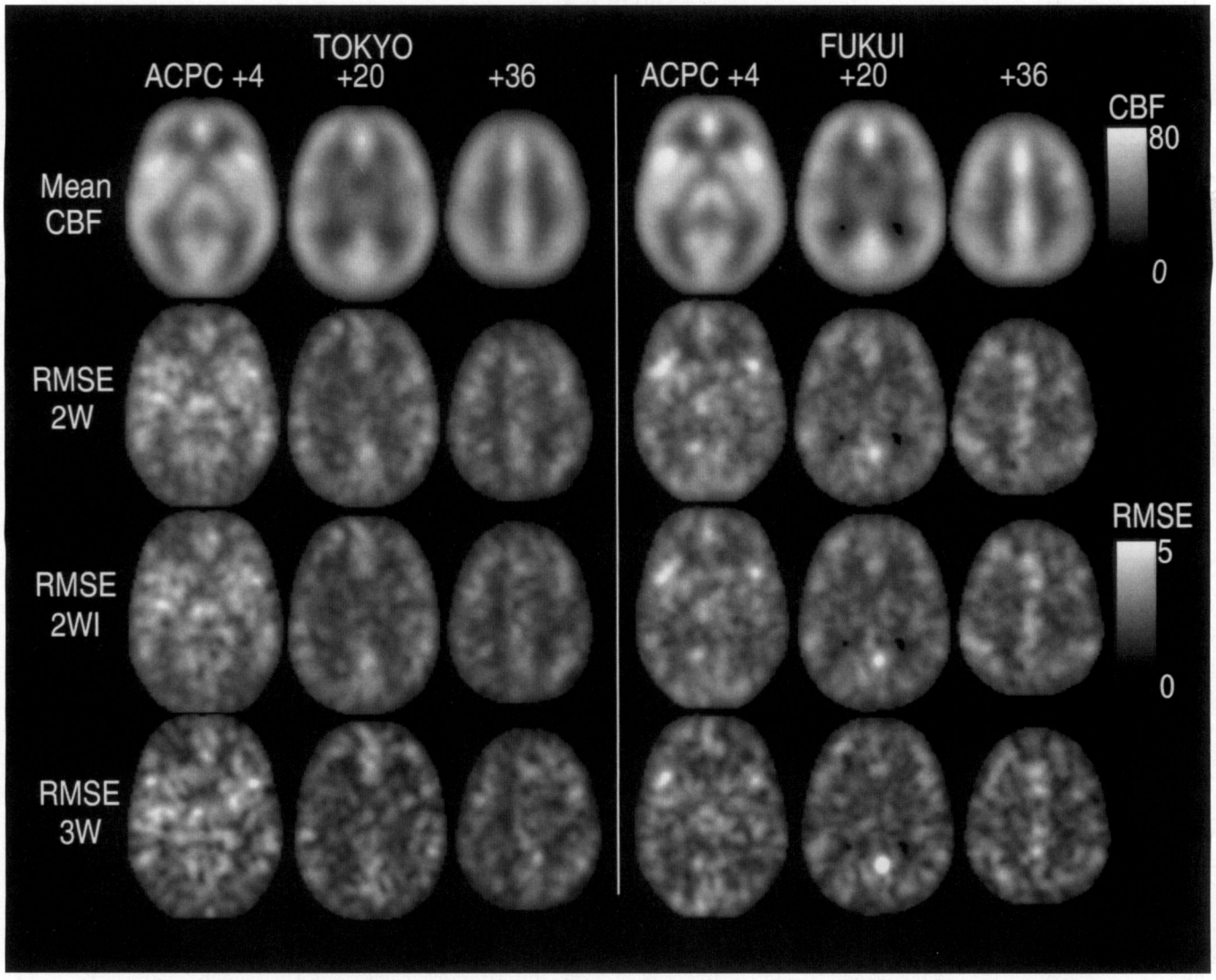

FIG. 4. Mean CBF images and root mean square for error (RMSE) based on three different ANOVA designs (2W, 2WI, 3W; see text for definitions) computed from the identical data of six subjects acquired in Tokyo (left) and Fukui (right) and anatomically standardized with SPM. The gray scale is in ml/min/100 ml when the global CBF is adjusted at 50. Note heterogeneity in RMSE, higher in the gray matter than in the white matter, and higher in the central gray than in the cortex.

the results. We used the covert verb veneration paradigm, because it has been investigated with the PET activation technique by a number of groups (Wise *et al.,* 1991; Herholz *et al.,* 1996; Warburton *et al.,* 1996; Grabowski *et al.,* 1996; Poline *et al.,* 1996a) and is known to activate language related areas evidently and, therefore, is suitable for comparison of the methods of data analysis. The activated areas visualized in the present study were, in general, consistent with those previous reports and included the left frontal cortex covering the inferior frontal gyrus (Broca's area) and extending upward to the middle frontal and premotor cortex, posteriorly to the superior temporal cortex, and medially to the operculum. To evaluate the effect of standardization method and ANOVA design on foci localization, we examined the highest and the second highest peaks within the activated area covering and surrounding Broca's area as described above. The results indicated a maximum variation of 3–4 cm in the highest peak localization depending on the standardization method and on the ANOVA design even if identical data sets were analyzed (Figs. 2 and 3 and Table 1). When the standardized images were treated with a stronger smoothing filter, the maximum variation was reduced to 1–3 cm (Table 2).

The variation in the peak z localization due to ANOVA design is caused by the regional relative F value being affected by the ANOVA design. Because the ANOVAs in the present study are of balanced design, the MSβs in the numerator for F in Eqs. (6)–(9) are

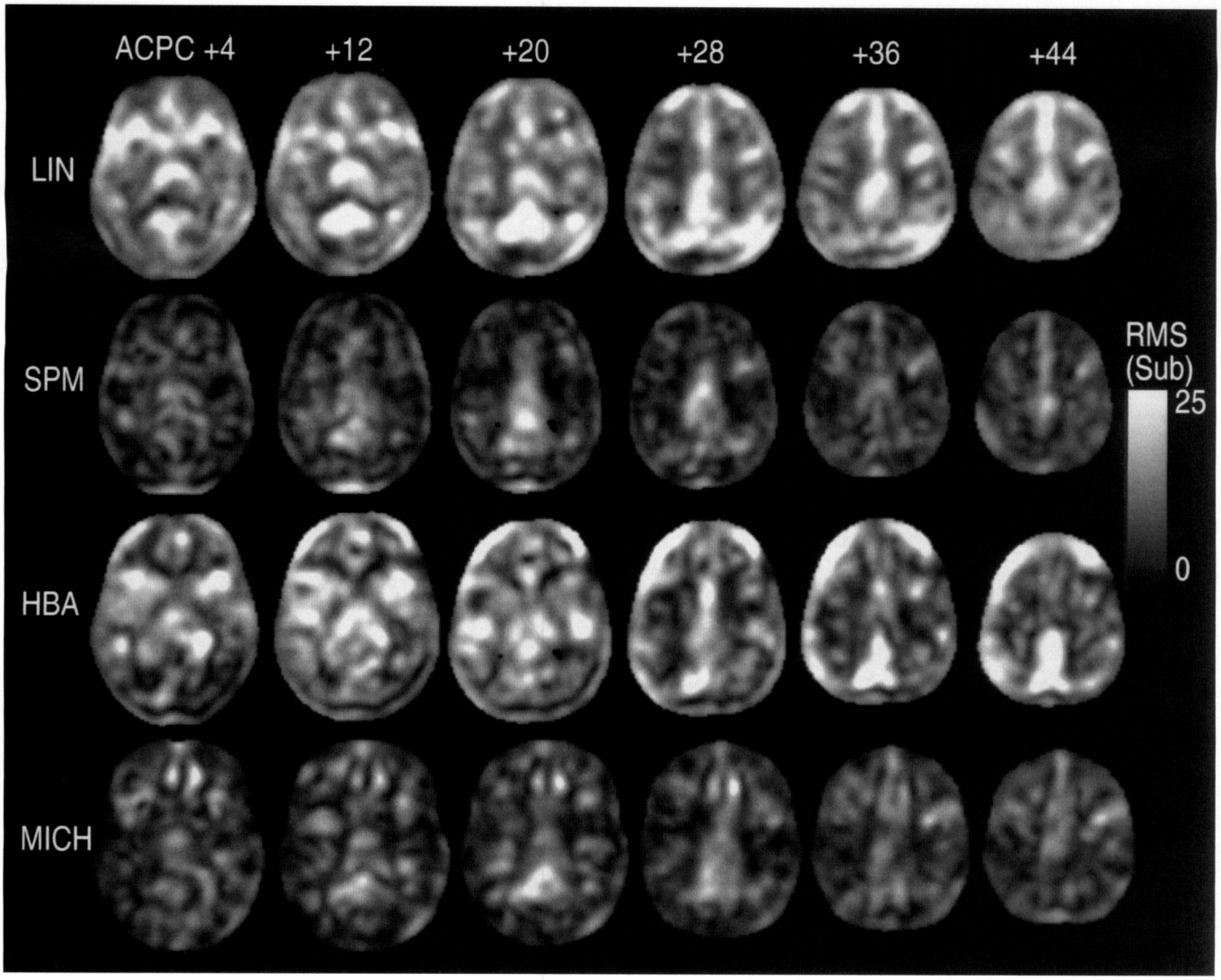

FIG. 5. Images of root mean square for the subject effect (RMS-S) computed from the identical data of six subjects acquired in Fukui and anatomically standardized with four different methods (LINEAR, SPM, HBA, and MICHIGAN; see text for details). Slice position labels are for LIN and SPM. For HBA and MICH, slices corresponding to those shown in Fig. 1 are depicted. The gray scale is in ml/min/100 ml when the global CBF is adjusted at 50. The RMS-S represents intersubject mismatch that has not been accounted for by the anatomical standardization and assumes higher values on the border of gray matter.

equal. Therefore it is the denominator, MSE or MS($\alpha\beta$), that makes F different between ANOVA designs. The MSE is not uniform across the brain and also depends on the ANOVA design as shown by the images of RMSE in Fig. 4 which are the square root of MSE. The MS($\alpha\beta$) is not uniform either, as shown by the images of RMS-ST in Fig. 6 which are the square root of MS($\alpha\beta$). Because the activated region in and around Broca's area spanned several centimeters and contained a number of potential local maxima, a small relative change in the denominator may switch the highest z point from one peak to another located centimeters away. Those local maxima may be an unstable manifestation of a single neuronal activation focus or may represent two or more significant neuronal foci within the above-mentioned area. Image smoothing reduced the number of small peaks in the numerator and increased the homogeneity of the denominator for F in Eqs. (6)–(9). However, while stronger smoothing reduced the variation and made the highest peak localization more stable, it resulted in a single peak over the entire activation area in some cases, erasing all local maxima, including possible significant subfoci.

The difference in MSE between ANOVA designs, at least partly, is due to the instability of the MSE caused by the small number of observations. In other words, the measured MSE does not precisely represent the true variance of the error term (ϵ) in the corresponding

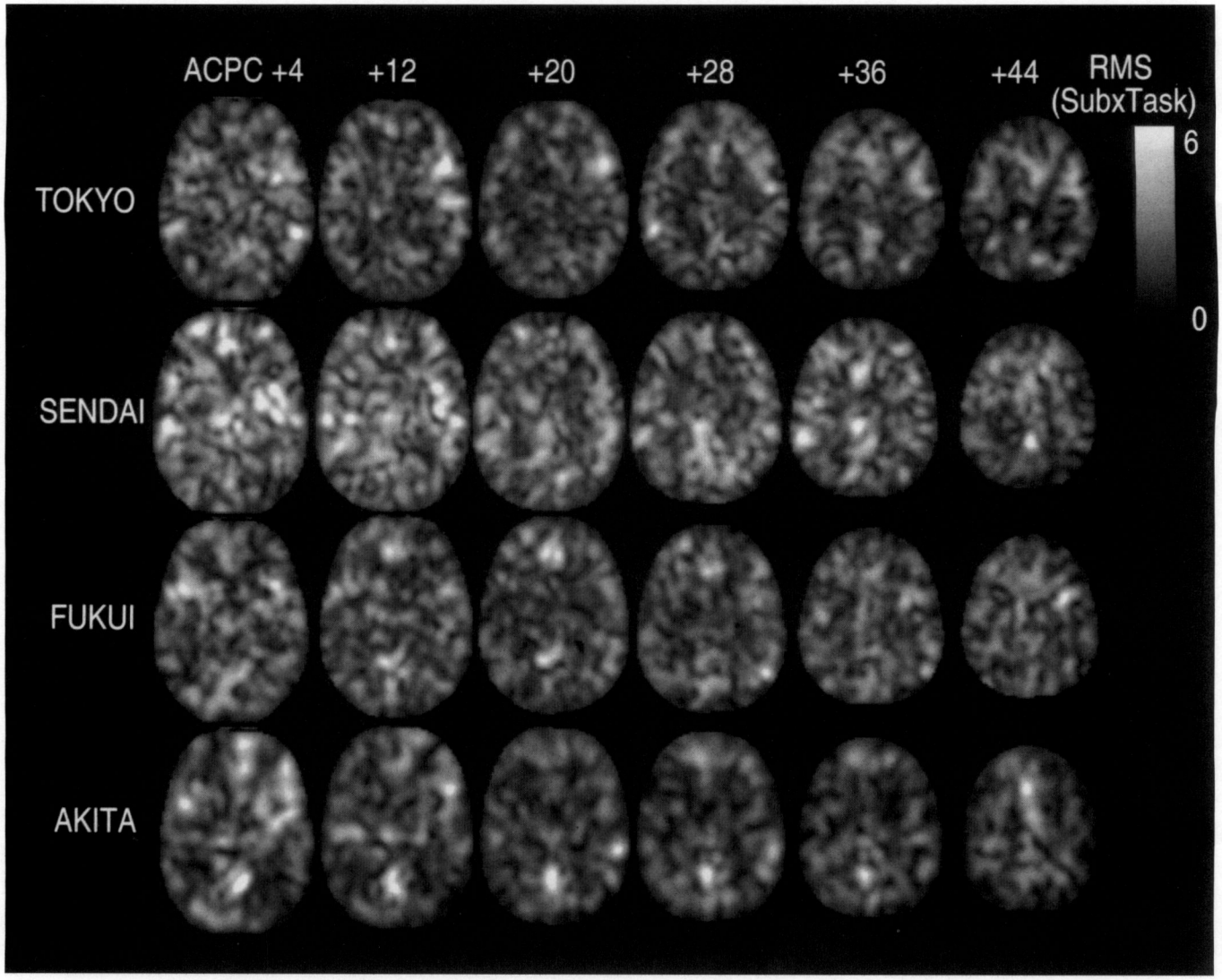

FIG. 6. Images of root mean square for the subject-by-task interaction effect (RMS-ST) computed from the data of six subjects acquired in each city and anatomically standardized with SPM. The gray scale is in ml/min/100 ml when the global CBF is adjusted at 50. The RMS-ST represents intersubject variation in activation.

Eq. (6)–(9). This interpretation is supported by the "noisy" appearance of RMSE images (Fig. 4) and by the fact that stronger smoothing reduced the difference in peak localization especially between 2W and 2WI (Table 2). However, a variation of 1–2 cm in the peak localization still remained between 3W and 2W/2WI even in case of the 20-mm FWHM smoothing. This suggests that the foci localization may actually depend on the mathematical assumptions underlying each ANOVA design that formulate the CBF change with potential factors and define the null hypothesis. Therefore, investigators are encouraged to explicitly specify the ANOVA/ANCOVA design they use for the analysis because the result may substantially depend on it.

These results have two important implications. (i) Image smoothing makes the foci localization more robust, while it sacrifices the spatial resolution and may erase possible subfoci. Determination of the smoothing level has been somewhat a matter of trial and error, and no criterion is established that convinces everyone. Some investigators do not even describe anything about smoothing operations. As proposed by recent investigations (Poline *et al.,* 1994b,c; Worsley *et al.,* 1996b), it may be appropriate to execute multiple levels of smoothing and to interpret the integrated results if the area of activation is not a single point. Another possible approach may be to smooth the denominator (MSE) more intensely than the numerator (MSβ) when computing F in Eqs. (6)–(9) (Worsley *et al.,* 1992; Holmes *et al.,* 1996). (ii) It may be insufficient to describe the activation foci only with the highest z peak. The highest peak coordinates, which are used by

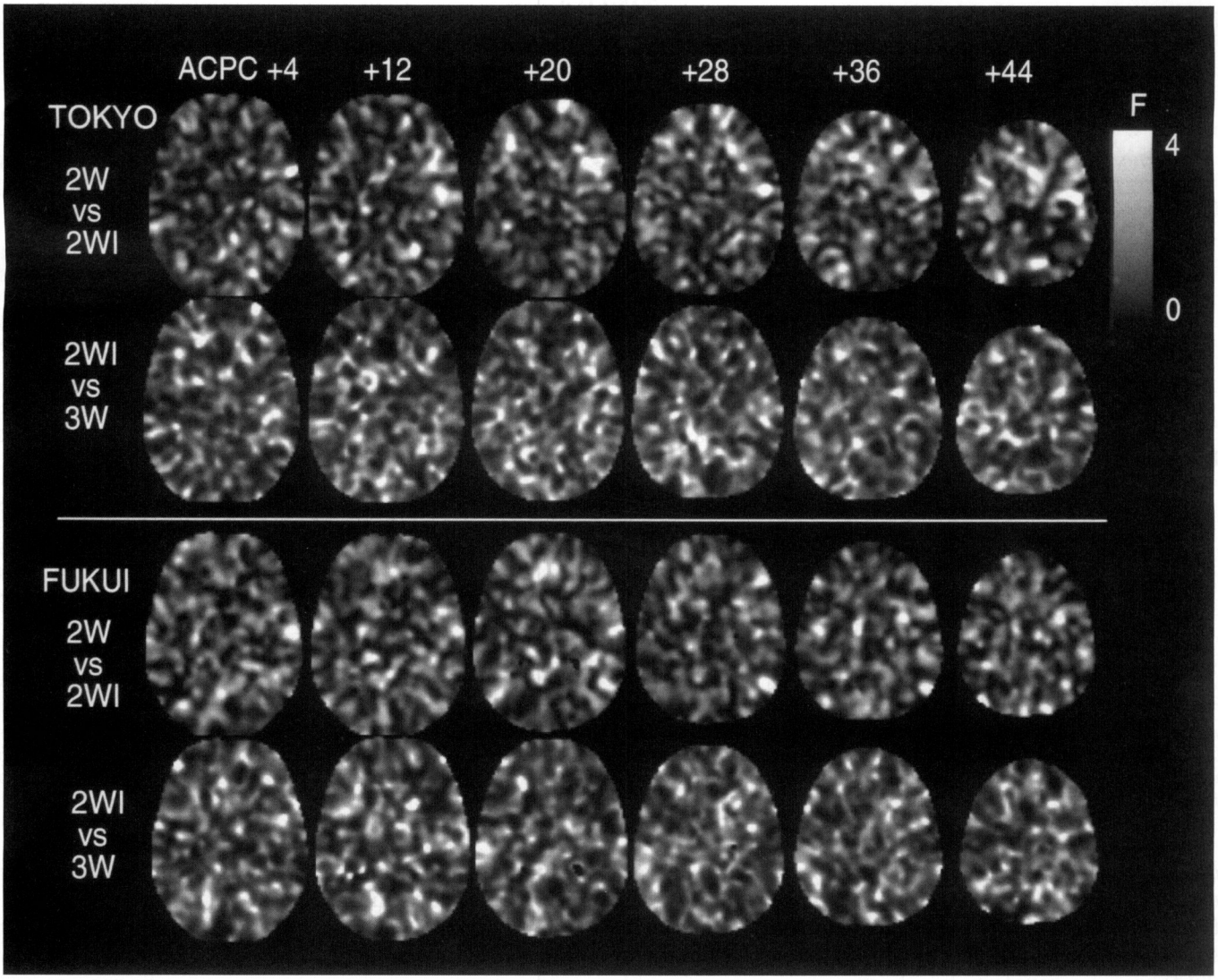

FIG. 7. Images of F value representing the significance of additional effects for 2W vs 2WI and for 2WI vs 3W, which have been computed from the data of six subjects acquired in Tokyo and in Fukui and anatomically standardized with SPM.

the majority of current investigators, heavily depend on the ANOVA design. Center of gravity is another approach, but it may depend on the region, of which it is to be computed, or on the threshold value to determine the region. A more appropriate way of describing the foci should be developed in the future. We have no reason to believe that a single point or a single spherical volume is activated by a task as complicated as the verb generation. Although concise description of the activated area is preferable, the images themselves may be used to describe the result.

The apparent variation in the peak localization due to the anatomical standardization method is caused both by the smoothing effect due to the interpolation procedure involved in the transformation and by the actual shift of the peak by the transformation. The former affects the MSE, and accordingly, the z map as discussed above. The latter comes from the fact that the coordinate system is defined in the atlas on which each method is based, and therefore, the same coordinates may correspond to a different structure on a different atlas. For example, the distance between the frontal pole to AC, AC to PC, and PC to the occipital pole along the AC–PC line is 73/25/73 mm in LINEAR, 72/24/80 mm in HBA, and 70/24/79 mm in MICHIGAN. This simply demonstrates that an identical peak in the original PET images is expressed with different y coordinates by different atlases. Furthermore, even if the same atlas is used, a different transformation algorithm may map the same pixel in the image to different coordinates in the atlas. Therefore, the anatomical standardization method must be specified when coordinates in the standardized images are described. Choice of an option and other details should also be

TABLE 2

Dependence of Peak z Score and CBF difference and Their Localization in and around Broca's Area upon ANOVA Design and Anatomical Standardization (Effect of Smoothing)

	LINEAR				SPM				HBA				MICHIGAN			
ANOVA design	x	y	z	Peak	x	y	z	Peak	x	y	z	Peak	x	y	z	Peak
Fukui data (FWHM = 15 mm)																
2W	−33	33	8	6.70	−30	18	12	6.94	−33	37	14	5.70	−57	28	0	6.94
	−17	11	20	5.84	−24	38	16	4.93	−21	21	14	4.96	−44	17	14	6.23
2WI	−27	29	8	6.77	−28	16	12	6.71	−31	35	14	5.61	−57	28	0	6.40
	−19	11	20	5.81	−18	4	16	5.59	−33	23	18	5.36	−42	17	16	5.91
2WIMX	−15	23	12	4.35	−16	14	8	4.38	−11	21	18	4.39	−60	24	11	4.78
	−35	15	16	4.18	−34	24	8	4.30	−29	59	26	4.01	−62	24	18	4.70
3W	−29	35	24	5.31	−16	8	16	5.20	−17	27	14	4.77	−15	8	16	5.08
	−17	15	16	5.30	−28	18	20	4.97	−27	37	26	4.76	−35	28	0	4.99
Difference	−41	21	16	5.93	−38	12	12	6.71	−43	31	14	6.61	−42	13	9	7.02
Fukui data (FWHM = 20 mm)																
2W	−29	21	16	6.73	−26	12	16	7.12	−31	33	14	5.73	−60	31	0	6.93
	−33	31	12	6.71									−48	10	27	6.29
2WI	−29	29	12	6.78	−26	14	12	6.79	−29	31	14	5.67	−57	28	−2	6.46
													−51	8	22	6.04
2WIMX	−13	5	36	4.36	−16	0	16	4.99	−15	15	14	4.33	−55	17	7	4.48
	−17	7	16	4.24	−34	20	12	3.95	−9	19	18	4.18	−60	28	2	4.39
3W	−25	25	16	5.05	−22	14	16	5.17	−27	33	22	4.91	−33	26	20	4.70
	−27	33	24	4.92	−48	20	4	4.59	−31	21	26	4.60	−28	15	9	4.60
Difference	−39	23	16	4.80	−36	14	12	5.35	−41	33	14	5.21	−44	15	11	5.68

Note. Analyzed from the data of six subjects acquired in Fukui. The highest and the second highest peaks are shown for z peaks. Coordinates (x, y, z) are mm from AC, with positive directions toward right, anterior, and superior. Peak values are in z score for each ANOVA design and in ml/min/100 ml for CBF difference (gCBF = 50). See text for details of ANOVA designs (2W, 2WI, 2WIMX, 3W) and anatomical standardization methods (LINEAR, SPM, HBA, MICHIGAN).

described when a number of options are available, as occurs in SPM. This discrepancy of coordinates between atlases and between transformation methods should be taken into account in an attempt to make a database of activation foci from a collection of data acquired and analyzed by different investigators (Fox and Lancaster, 1994). It should be noted that the limitation of the atlas coordinates described above occurs only in an intersubject analysis and not in an individual subject analysis, in which PET can be registered to the individual MRI.

TABLE 3

Smoothness Expressed as Effective FWHM(x) (mm) for the Residual Images of 3W-ANOVA Computed from Each Data Set Using Various Anatomical Standardization Methods and Smoothing Filters

Data set	Filter	LINEAR	SPM	HBA	MICHIGAN
Tokyo	10 mm	10.4	10.6	10.6	10.8
Sendai	10 mm	10.1	10.3	10.4	10.5
Fukui	10 mm	10.7	10.6	11.0	10.8
	15 mm	16.3	16.5	16.8	16.4
	20 mm	23.1	23.5	23.6	22.5
Akita	10 mm	11.6	11.8	11.8	11.8
	15 mm	15.9	16.8	16.1	16.1
	20 mm	21.6	23.6	22.0	21.5

The variation of 3–4 cm by the standardization observed in the present study is contrasted with that for the primary sensorimotor cortex for fingers investigated by our group (Senda *et al.*, 1993, 1994). In that study, we transformed an identical data set on vibrotactile and motor stimulation of the fingers using three different methods of anatomical standardization and found a variation of no more than 1 cm in the foci localization. This is because the area of activation by the vibrotactile and motor stimulation is smaller than the verb generation and contains only a single peak.

Peak z Values

The z value at the highest peak was similar among the 2W, 2WI, and 3W (Table 1). Stronger smoothing in general increased the peak z values due to improved signal-to-noise ratio for 2W and 2WI (Table 2). As the ANOVA model becomes complex, incorporating more effects and factors, the RMSE decreases, but the degree of freedom decreases simultaneously, resulting in similar peak z values. In this sense 3W provided no more sensitive detection of the activation foci than 2W or

2WI. The apparent absence of hot spots in the F images of the additional effects for 2WI vs 3W (Fig. 7) suggests that 3W is not necessary to explain the CBF variation of the present data in or around the activation foci.

When the peak z value was compared among the standardization methods, SPM tended to present higher z values than the other methods except for the Akita data, in which MICHIGAN presented the highest. Anatomical standardization is aimed at accounting for the individual variation in activation. Our results indicate that SPM reduced the individual variation in foci localization more efficiently than LINEAR and HBA. Because LINEAR used only linear parameters, it did not normalize the shape of the brain appropriately nor, therefore, the foci either. On the other hand, HBA used nonlinear transformation parameters and was most loyal to morphology in matching anatomical landmarks to the atlas. However, HBA provided lower z values than SPM. This suggests that utilizing the information about the sulci location may not reduce the interindividual foci variation. Herholz *et al.* (1996) examined functional anatomy of the verb generation in each subject and reported a significant individual variation in the anatomical configuration of the Broca's area as well as in the relationship between activation foci and morphology. Therefore, HBA may not have accounted for the individual anatomical variation in the Broca and surrounding areas. Even if it did, functional variation may have remained. It should be noted that HBA depends on the operator. A different result may have come out if it had been more skillfully operated on and around Broca's area.

Since the z value is closely related to the degree of smoothing, the smoothness was estimated and expressed as effective FWHM in Table 3. The smoothness was similar among the standardization methods and among the centers and was comparable to the predicted spatial resolution of the images. The slight differences may have come from computation error. The difference between the methods may also arise from how and how many times the images were interpolated in each method of anatomical standardization. The camera resolution differs among the centers and the images were realigned to correct for the interscan motion in Fukui and Akita data, which may also contribute to the slight difference between the data sets.

The present study did not handle the P values computed from the z map or the z threshold for testing the statistical significance. There are a number of theories as to how the null hypothesis of "no activation" is statistically tested (Friston *et al.*, 1991; Worsley *et al.*, 1992). Recent theories assume activation of a cluster or a region instead of a peak (Friston *et al.*, 1996; Worsley *et al.*, 1996a; Poline *et al.*, 1997). Also, statistical methods other than ANOVA are available, and some of them were evaluated by Grabowski *et al.* (1996).

Mixed-Effects Design

The subject in the ANOVA model is theoretically considered a random factor because the group of six subjects is a random sample from the population. Our interest lies not in the difference between specific subjects but in the intersubject variation of the population. Therefore 2WI-MX is a theoretically correct design. As Eqs. (8) and (7) are compared, one can realize that the task-induced mean CBF increase is evaluated against the individual variation in CBF increase [MS($\alpha\beta$)] in 2WI-MX, while it is evaluated against the experimental error (MSE) in 2WI. In the present study, 2WI-MX showed lower z values than the other designs. This indicates that the individual variation in activation is larger than the experimental error. The small degree of freedom ($df = 5$) in 2WI-MX is also responsible for the poorer outcome. Those findings suggest that this approach is less sensitive and less stable in detecting activation foci and may be inappropriate for a small number of subjects, unless individual variation in activation is very small (Woods *et al.*, 1996).

Error Images

The error images (RMSE for 2W, 2WI, 3W) were heterogeneous across the brain, being high in the gray matter and low in the white matter (Fig. 4). The error represents the variation in gCBF-adjusted CBF that is not explained by the ANOVA design and is attributed to the statistical noise, other experimental errors, and the CBF fluctuations. The propagation of Poisson statistical noise to the PET images has been analytically estimated by many investigators including Alpert *et al.* (1982) and Carson *et al.* (1993). The statistical noise is known to be higher in the center than in the periphery due mainly to poorer count statistics for the center both in the emission and in the transmission scan. The regional statistical noise is not necessarily correlated to the regional radioactivity concentration as illustrated by Palmer *et al.* (1986). Alpert *et al.* (1991) reported that the fluctuation other than the statistical noise constitutes as much as about half of the observed total variation in $C^{15}O_2$–CBF PET images. Therefore, the gray-to-white contrast in the RMSE images observed in the present study suggests larger physiological fluctuation in CBF for the gray than for the white. It may be partly caused by the larger experimental error in CBF measurement in a high flow range than in a low flow range, which is characteristic for the PET–ARG method. Within the gray, the central and deep structures tended to show higher RMSE than the surface cortical rim (Fig. 4). This may be explained by the statistical noise of PET being higher in the center than in the periphery as has been demonstrated by previous investigators.

Although the MSE is a rough estimate of the true variance of the error term [ϵ in Eqs. (6)–(9)], visualization of gray matter structures strongly suggests heterogeneity of the true error variance and that we should be cautious about the assumption of homogeneous error across the entire brain. Arndt *et al.* (1995) compared the histogram of standard error (SE, equivalent to our RMSE) with its theoretical distribution and found a discrepancy. In a practical situation of detecting statistically significant foci, however, it may be a case-by-case matter whether to compute the z value based on homogeneous or heterogeneous error distribution. As Worsley *et al.* (1992) proposed, assuming a homogeneity may be justified based on the unreliability of estimated error in each pixel for a small number of subjects. It may be practicable to segment the brain into regions having similar RMSE and to compute the pooled variance for each segmented region.

Figure 4 also revealed no apparent correlation between RMSE and activation by the task except for sporadic high RMSE regions in Broca's area and in the cingulate cortex. This suggests that the cortical RMSE is independent of the activation and that our ANOVA designs sufficiently explain regional difference in the behavior of CBF for the task paradigm. This may depend on the task paradigm because Arndt *et al.* (1995) reported larger error in activated areas than in nonactivated areas for memory and reading tasks. The sporadic high RMSE in the anterior and posterior cingulate cortex may be attributed to their higher CBF fluctuation that is not controlled by the subject condition.

Subject Effect

The RMS for the subject effect, reflecting intersubject mismatch in the gray matter distribution, was high on the brain contour and gray–white borders, which was most pronounced in LINEAR and in HBA, but was still notable in SPM and MICHIGAN, especially in the cingulate cortex (Fig. 5). This indicates that the gray matter structure is not properly normalized anatomically with LINEAR or HBA and that the normalization is still imperfect even with SPM or MICHIGAN. Since the LINEAR method transforms the brain linearly, nonlinear variation in gray matter distribution was not accounted for, resulting in a mismatch on the border of gray matter. In HBA, the brain image was transformed interactively so that the brain contour and major sulci would match their counterparts in the atlas. However, even if the contour and sulci are matched, the gray matter distribution may not be, because the relationship between the morphological landmarks observed in MRI and the gray matter distribution is not constant. The RMS-S is further enlarged where there is a large individual variation in CBF distribution even if the gray matter is matched. The accuracy of HBA is also influenced by possible PET–MRI misregistration and MRI distortion. On the other hand SPM and MICHIGAN use only PET images and directly match the CBF distribution to their templates. These factors may explain why HBA had surprisingly larger RMS-S than SPM and MICHIGAN on the gray matter border although they are all nonlinear transformations. HBA presented a still larger RMS-S along the prefrontal cortex, probably because HBA has a smaller number of landmarks on the prefrontal cortex than on the parieto-occipital cortex. SPM and MICHIGAN had a relatively large RMS-S over the medial occipital cortex extending to the posterior cingulate cortex, suggesting that these regions may be weak points for SPM and MICHIGAN.

Smoothing reduced the RMS-S across the brain, but relatively large RMS-S along the gray matter border was still observed under intense smoothing because the images were smoothed uniformly across the brain disregarding the structure and CBF distribution. This suggests that uniform smoothing is a way of reducing RMS-S but is not the best way, nor does it correct for the inappropriate standardization.

It is the gray matter that is activated if any region is activated, and standardization is considered to be incomplete if there is a mismatch in gray matter distribution. Therefore, when anatomical standardization methods are compared, one with smaller RMS-S on the gray matter border is always more appropriate for the analysis of PET activation data. In this sense, nonlinear methods are better than linear methods, and SPM and MICHIGAN, which use PET image values, are more appropriate than HBA, which uses only morphological information. However, smaller RMS-S and absence of high RMS-S on the gray matter border is not a sufficient condition for a complete standardization. This is because a match in gray matter distribution does not guarantee a match in anatomically corresponding cortical areas among individuals, much less of functionally corresponding cortical areas. Therefore, it is desirable for the users of SPM or MICHIGAN to confirm the absence of irrational distortion by transforming each subject's MRI with the same transformation parameters as have been used on the subject's PET images. It should be noted here that the above-discussed points are valid for normal subjects and may not be applicable to patients with substantially abnormal radioactivity distribution.

It should also be noted that the detectability of activation foci is low in the area with large RMS-S due to low S/N in intersubject averaging analysis. Investigators should be careful about it when they note a failure to detect activation in those areas. Therefore, it is desirable to create and examine RMS-S images in every study that uses an intersubject averaging technique.

Subject-by-Task Interaction

The RMS for the subject-by-task interaction, representing individual variation in activation, had inconsistent high regions in Broca's area, left prefrontal, cingulate, and visual cortex, depending on the data set (Fig. 6). Except in the occipital cortex, the high RMS-ST regions are located within the activated areas identified in intersubject averaging analysis. This is either because of incomplete standardization of the activated area in each subject or because of individual variation in the amount of activation within the area or both. In the mean time, those high RMS-ST regions are the cause of the different appearance between the 2WI-MX z map and the other z maps. The high RMS-ST regions in the visual cortex may reflect an activation in some subjects and no activation in the others. In fact, some subjects in Akita answered that they used a strategy of "imagery" to perform the verb-generation task. If this is the case, this RMS-ST approach may be useful to explore a difference in the pattern of activation between subgroups of subjects.

The presence of high RMS-ST regions suggests that 2WI better describes the CBF variations than 2W. In fact the F images for 2W vs 2WI in Fig. 7, which represent the significance of $(\alpha\beta)_{ij}$ as an additional term in Eq. (7), appeared similar to RMS-ST and had high values in the above mentioned areas. Although the F value is difficult to interpret quantitatively because of multiple comparisons, this suggests that 2WI is valid compared to 2W and that the subject-by-task interaction not only exists theoretically but also is observable in the present data.

Replication Effects

The F images for 2WI vs 3W in Fig. 7, which represents the significance of additional terms regarding replication effects, i.e., $\gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$ in Eq. (9), did not show high values in or around the activation foci. This suggests that the CBF variation due to those effects is less significant than the experimental error and that the present data failed to support the validity of 3W compared to 2WI. The small degree of freedom in 3W is also responsible. This finding is probably related to the failure of 3W providing higher peak z values than 2WI in the z image for the task main effect (Tables 1 and 2).

The presence of habituation is generally believed in psychology and is demonstrated by CBF measurement (Warach *et al.*, 1992; Friston *et al.*, 1993). Therefore, absence of significant replication effects does not preclude the use of 3W even if the task main effect is to be investigated. Some investigators may also argue that, from the standpoint of experimental design, the statistical model is to be built before the data acquisition and is not to be selected after the data are analyzed. However, the F approach shown in Fig. 7 and the variation components analysis proposed in the present study allows investigators to check the significance of each effect and the validity of the ANOVA design *a posteriori* compared to the experimental error and the precision provided by the data themselves.

Reproducibility between Centers

In the present study, four sets of data were analyzed independently. Although figures were presented on Fukui data and partly on Tokyo data, Table 1 indicates that the peak in the z map is similar among the four sets of data. This indicates that the observed activation regions are highly reproducible despite a rather small number of subjects (=6) and the difference in the PET camera. The apparent difference in the peak z value may be attributed to the S/N of the original data in each center, which depends on the dose, count rate, acquisition time, and various correction methods. Poline *et al.* (1996a) carried out a similar multicenter study in Europe and found highly reproducible results between centers. Detailed comparison between centers as well as between this work and previous European and American investigations on the verb generation in terms of race and language difference will be reported in another paper.

CONCLUSION

We have compared four methods of anatomical standardization on identical data sets. A nonlinear method is better than a linear method, and the method that seeks to match the value of the PET image to that of the standard PET image (SPM and MICHIGAN) works better than the method that seeks to match the landmarks determined by PET-registered MRI (HBA). However, investigators are recommended to inspect the morphological result of the transformation for each subject if MRI is not used for anatomical standardization. Under the specific experimental condition of the present study, SPM worked best of all the methods studied. However, this conclusion cannot be generalized to every PET activation study. The best choice may depend on the activation paradigm, scanning protocol, and number of subjects as well as the population.

We have also compared various ANOVA designs on identical data sets. The two-way design (2W) sufficiently demonstrated the activation foci in the present data. However, the subject-by-task interaction (2WI) was also significant, which supported the two-way design with interaction (2WI). The three-way design (3W) incorporating the replication effects was not strongly supported by the present data. The mixed-effects design (2WI-MX) did not work well due to large intersubject variation. However, this conclusion may

also depend on the activation paradigm, scanning protocol, and number of subjects.

We have proposed variance component images including the root mean square for the subject effect and that for the subject-by-task interaction. The RMS-S is useful for evaluation of the anatomical standardization, and the RMS-ST is useful for evaluation of individual variation in activation and for assessment of 2WI design compared to 2W.

Acknowledgments

We thank Takao Fushimi and Masayuki Tanaka at the TMIG for their assistance in designing and preparing the task. We also thank Hinako Toyama at the TMIG for arranging the computer network and providing software to facilitate the data analysis. We thank all the staff of the participating four PET centers for their technical assistance in data acquisition and preprocessing.

REFERENCES

Alpert, N. M., Chestler, D. A., Correia, J. A., Ackerman, R. H., Chang, J. Y., Finklestein, S., Davis, S. M., Brownell, G. L., and Taveras, J. M. 1982. Estimation of the local statistical noise in emission computed tomography. *IEEE Trans. Med. Imaging* **MI-1:** 142–146.

Alpert, N. M., Barker, W. C., Gelman, A., Weise, S., Senda, M., and Correia, J. A. 1991. The precision of positron emission tomography: Theory and measurement. *J. Cereb. Blood Flow Metab.* **11:** A26–A30.

Arndt, S., Cizadlo, T., Andreasen, N. C., Zeien, G., Harris, G., O'Leary, D. S., Watkins, G. L., Ponto, L. L., and Hichwa, R. D. 1995. A comparison of approaches to the statistical analysis of [^{15}O]H2O PET cognitive activation studies. *J. Neuropsychiat. Clin. Neurosci.* **7:** 155–168.

Arndt, S., Cizadlo, T., O'Leary, D., Gold, S., and Andreasen, N. C. 1996. Normalizing counts and cerebral blood flow intensity in functional imaging studies of the human brain. *NeuroImage* **3:** 175–184.

Ardekani, B. A., Braun, M., Hutton, B. F., Kanno, I., and Iida, H. 1995. A fully automated multimodality image registration algorithm. *J. Comput. Assisted Tomogr.* **19:** 615–623.

Carson, R. E., Yan, Y., Daube-Witherspoon, M. E., Freedman, N., Bacharach, S. L., and Herscovitch, P. 1993. *IEEE Trans. Med. Imaging* **MI-12:** 240–250.

DeGrado, T. R., Turkington, T. G., Williams, J. J., Stearns, C. W., and Hoffman, J. M. 1994. Performance characteristics of a whole-body PET scanner. *J. Nucl. Med.* **35:** 1398–1406.

Fox, P. T., Perlmutter, J. S., and Raichle, M. E. 1985. A stereotactic method of anatomical localization for positron emission tomography. *J. Comput. Assisted Tomogr.* **9:** 141–153.

Fox, P. T., and Lancaster, J. L. 1994. Neuroscience on the net. *Science* **226:** 994–996.

Friston, K. J., Passingham, R. E., Nutt, J. G., Heather, J. D., Sawle, G. V., and Fackowiak, R. S. J. 1989. Localisation in PET images: Direct fitting of the intercommissural (AC–PC) line. *J. Cereb. Blood Flow Metab.* **9:** 690–695.

Friston, K. J., Frith, C. D., Liddle, P. F., Dolan, R. J., Lammertsma, A. A., and Frackowiak, R. S. J. 1990. The relationship between global and local changes in PET scans. *J. Cereb. Blood Flow Metab.* **10:** 458–466.

Friston, K. J., Frith, C. D., Liddle, P. F., and Frackowiak, R. S. J. 1991. Comparing functional (PET) images: The assessment of significant change. *J. Cereb. Blood Flow Metab.* **11:** 690–699.

Friston, K. J., Frith, C. D., Liddle, P. F., and Frackowiak, R. S. J. 1993. Functional connectivity: The principal-component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.* **13:** 5–14.

Friston, K. J., Ashburner, J., Frith, C. D., Heather, J. D., and Frackowiak, R. S. J. 1995. Spatial registration and normalization of images. *Hum. Brain Mapping* **3:** 165–189.

Friston, K. J., Holmes, A., Poline, J-B., Price, C. J., and Frith, C. D. 1996. Detecting activations in PET and fMRI: Levels of inference and power. *NeuroImage* **40:** 223–235.

Grabowski, T. J., Frank, R. J., Brown, C. K., Damasio, H., Boles Ponto, L. L., Watkins, G. L., and Hichwa, R. D. 1996. Reliability of PET activation across statistical methods, subject groups, and sample sizes. *Hum. Brain Mapping* **4:** 23–46.

Herbster, A. N., Mintun, M. A., Nebes, R. D., and Becker, J. T. 1997. Regional cerebral blood flow during word and nonword reading. *Hum. Brain Mapping* **5:** 84–92.

Herholz, K., Thiel, A., Wienhard, K., Pietrzyk, U., von Stockhausen, H-M., Karbe, H., Kessler, J., Bruckbauer, T., Halber, M., and Heiss, W-D. 1996. Individual functional anatomy of verb generation. *NeuroImage* **3:** 185–194.

Herscovitch, P., Markham, J., and Raichle, M. E. 1983. Brain blood flow measured with intravenous $H_2{}^{15}O$. I. Theory and error analysis. *J. Nucl. Med.* **24:** 782–789.

Holmes, A. P., Blair, R. C., Watson, J. D. G., and Ford, I. 1996. Nonparametric analysis of statistic images from functional mapping experiments. *J. Cereb. Blood Flow Metab.* **16:** 7–22.

Iida, H., Miura, S., Kanno, I., Murakami, M., Yamamoto, S., and Amano, M. 1989. Design and evaluation of Headtome-IV, a whole-body positron emission tomograph. *IEEE Trans. Nucl. Sci.* **NS-37:** 1006–1010.

Iida, H., Miura, S., Kanno, I., Ogawa, T., and Uemura, K. 1996. A new PET camera for noninvasive quantification of physiological functional parametric images. In *Quantification of Brain Function using PET* (R. Myers, V. Cunningham, D. Bailey, and T. Jones, Eds.), pp. 57–61. Academic Press, London.

Kanno, I., Iida, H., Miura, S., Murakami, M., Takahashi, K., Hiroshi Sasaki, Inugami, A., Shishido, F., and Uemura, K. 1987. A system for cerebral blood flow measurement using an $H_2{}^{15}O$ autoradiographic method and positron emission tomography. *J. Cereb. Blood Flow Metab.* **7:** 143–153.

Kleinbaum, D. G., Kupper, L. L., and Muller, K. E. 1988. Applied Regression Analysis and Other Multivariable Methods. PWS-Kent Publ., Boston.

Lewellen, T. K., Kohlmeyer, S. G., Miyaoka, R. S., and Kaplan, M. S. 1996. Investigation of the performance of the General Electric Advance positron emission tomograph in 3D mode. *IEEE Trans. Nucl. Sci.* **43:** 2199–2206.

McIntosh, A. R., Grady, C. L., Haxby, J. V., Maisog, J. M., Horwitz, B., and Clark, C. M. 1996. Within-subject transformation of PET regional cerebral blood flow data: ANCOVA, ratio, and z-score adjustments on empirical data. *Hum. Brain Mapping* **4:** 93–102.

Minoshima, S., Koeppe, R. A., Frey, K. A., and Kuhl, D. E. 1994a. Anatomic standardization: Linear scaling and nonlinear warping of functional brain images. *J. Nucl. Med.* **35:** 1528–1537.

Minoshima, S., Koeppe, R. A., Frey, K. A., Ishihara, M., and Kuhl, D. E. 1994b. Stereotactic PET atlas of the human brain: Aid for visual interpretation of functional brain images. *J. Nucl. Med.* **35:** 949–954.

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

Palmer, M. R., Bergstrom, M., Pate, B. D., and Beddoes, M. P. 1986. Noise distribution due to emission and transmission statistics in positron emission tomography. *IEEE Trans. Nucl. Sci.* **33:** 439–442.

Poline, J. B., Vandenberghe, R., Holmes, A. P., Friston, K. J., and Frackowiak, R. S. J. 1996a. Reproducibility of PET activation studies: Lessons from a multi-center European experiment EU concerted action on functional imaging. *NeuroImage* **4:** 34–54.

Poline, J. B., and Mazoyer, B. M. 1996b. Analysis of individual brain activation maps using hierarchical description and multiscale detection. *IEEE Trans. Med. Imaging* **13:**702–710.

Poline, J.B., and Mazoyer, B.M. 1996c. Enhanced detection in brain activation maps using a multifiltering approach. *J. Cereb. Blood Flow Metab.* **14:**639–642.

Poline, J-B., Worsley, K. J., Evans, A. C., and Friston, K. J. 1997. Combining spatial extent and peak intensity to test for activations in functional imaging. *NeuroImage* **5:**83–96.

Roland, P. E., Graufelds, C. J., Wahlin, J., Ingelman, L., Andersson, M., Ledberg, A., Pedersen, J., Akerman, S., Dabringhaus, A., and Zilles, K. 1994. Human brain atlas: For high resolution functional and anatomical mapping. *Hum. Brain Mapping* **1:** 173–184.

Roland, P. E., and Zilles, K. 1994. Brain atlases—A new research tool. *Trends Neurosci.* **11:**458–467.

Rumsey, J. M., Horwitz, B., Donohue, B. C., Nace, K., Maisog, J. M., and Andreason, P. 1997. Phonological and orthographic components of word recognition, a PET–rCBF study. *Brain* **120:** 739–759.

Ruttimann, U., Rio, D., Rawlings, R., Andreason, P., and Hommer, D. 1997. PET analysis using a variance stabilizing transform. *NeuroImage* **5:**B24.

Senda, M., Kanno, I., Yonekura, Y., Fujita, H., Ishii, K., Lyshkow, H., Miura, S., Oda, K., Sadato, N., and Toyama, H. 1993. Comparison of three anatomical standardization methods regarding foci localization and it's between subject variation in the sensorimotor activation. In *Quantification of Brain Function* (K. Uemura, N. A. Lassen, T. Jones, and I. Kanno, Eds.), pp. 439–445. Elsevier, Amsterdam.

Senda, M., Kanno, I., Yonekura, Y., Fujita, H., Ishii, K., Lyshkow, H., Miura, S., Oda, K., Sadato, N., and Toyama, H. 1994. Comparison of anatomical standardization methods regarding the sensorimotor foci localization and between-subject variation in $H_2{}^{15}O$ PET activation, a three-center collaboration study. *Ann. Nucl. Med.* **8:** 201–207.

Snodgrass, J. G., and Vanderwart, M. 1980. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learning Memory* **6:** 174–215.

Talairach, J., and Szikla, G. 1967. *Atlas of Stereotaxic Anatomy of the Telencephalon.* Masson, Paris.

Talairach, J., and Tournoux, P. 1988. *Co-planar Stereotaxic Atlas of the Human Brain.* Thieme, New York.

Warach, S., Gur, R. C., Gur, R. E., Skolnick, B. E., Obrist, W. D., and Reivich, M. 1992. Decreases in frontal and parietal lobe regional cerebral blood flow related to habituation. *J. Cereb. Blood Flow Metab.* **12:**546–553.

Warburton, E., Wise, R. J. S., Price, C. J., Weiller, C., Hadar, U., Ramsay, S., and Frackowiak, R. S. J. 1996. Noun and verb retrieval by normal subjects studies with PET. *Brain* **119:** 159–179.

Wessel, K., Zeffiro, T., Toro, C., and Hallett, M. 1997. Self-paced versus metronome-paced finger movements, a positron emission tomography study. *J. Neuroimaging* **7:** 145–151.

Winer, B. J. 1971. *Statistical Principles in Experimental Design,* 2nd ed, pp. 309–347. McGraw-Hill Kogakusha, Tokyo.

Wise, R., Chollet, F., Hadar, U., Friston, K., Hoffner, E., and Frackowiak, R. 1991. Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain* **114:** 1803–1817.

Woods, R. P. 1996. Modeling for intergroup comparisons of imaging data. *NeuroImage* **4:**S84–94.

Woods, R. P., Iacoboni, M., Grafton, S. T., and Mazziotta, J. C. 1996. Improved analysis of functional activation studies involving within-subject replications using a three-way ANOVA model. In *Quantification of Brain Function Using PET* (R. Myers, V. Cunningham, D. Bailey, and T. Jones, Eds.), pp. 353–358. Academic Press, San Diego.

Worsley, K. J., Evans, A. C., Marrett, S., and Neelin, P. 1992. A three-dimensional statistical analysis for CBF activation studies in human brain. *J. Cereb. Blood Flow Metab.* **12:** 900–918.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., and Evans, A. C. 1996a. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum. Brain Mapping* **4:** 58–73.

Worsley, K. J., Marrett, S., Neelin, P., and Evans, A. C. 1996b. Searching scale space for activation in PET images. *Hum. Brain Mapping* **4:**74–90.