NeuroImage **9,** 477–489 (1999)
Article ID nimg.1999.0429, available online at http://www.idealibrary.com on IDEAL®

Dynamical Cluster Analysis of Cortical fMRI Activation

Axel Baune,*,† Friedrich T. Sommer,† Michael Erb,* Dirk Wildgruber,* Bernd Kardatzki,*
Günther Palm,† and Wolfgang Grodd*

*Section Experimental MR of the CNS, Department of Neuroradiology, University of Tübingen, D-72076 Tübingen, Germany; and †Department of Neural Information Processing, University of Ulm, D-89069 Ulm, Germany

Received August 19, 1998

Localized changes in cortical blood oxygenation during voluntary movements were examined with functional magnetic resonance imaging (fMRI) and evaluated with a new dynamical cluster analysis (DCA) method. fMRI was performed during finger movements with eight subjects on a 1.5-T scanner using single-slice echo planar imaging with a 107-ms repetition time. Clustering based on similarity of the detailed signal time courses requires besides the used distance measure no assumptions about spatial location and extension of activation sites or the shape of the expected activation time course. We discuss the basic requirements on a clustering algorithm for fMRI data. It is shown that with respect to easy adjustment of the quantization error and reproducibility of the results DCA outperforms the standard *k*-means algorithm. In contrast to currently used clustering methods for fMRI, like *k*-means or fuzzy *k*-means, DCA extracts the appropriate number and initial shapes of representative signal time courses from data properties during run time. With DCA we simultaneously calculate a two-dimensional projection of cluster centers (MDS) and data points for online visualization of the results. We describe the new DCA method and show for the well-studied motor task that it detects cortical activation loci and provides additional information by discriminating different shapes and phases of hemodynamic responses. Robustness of activity detection is demonstrated with respect to repeated DCA runs and effects of different data preprocessing are shown. As an example of how DCA enables further analysis we examined activation onset times. In areas SMA, M1, and S1 simultaneous and sequential activation (in the given order) was found. © 1999 Academic Press

Key Words: **functional MRI; cluster analysis; evaluation methods; motor system.**

INTRODUCTION

Due to the advent of functional magnetic resonance imaging (fMRI) regional hemodynamic changes following neuronal activation can now be monitored in the brain with high spatial resolution (Ogawa *et al.*, 1990). But the various methods used for fMRI signal extraction in conjunction with the applied mode of cerebral activation are still subject of an intense and contradictory discussion (Bandettini *et al.*, 1995a; Kim *et al.*, 1997; Ogawa *et al.*, 1990). There are single-voxel-based methods of activation detection like *z*-mapping (Le Bihan *et al.*, 1993; Cohen and Bookheimer, 1994), parametric tests (i.e., *t* test) (Xiong *et al.*, 1995a) and nonparametric tests (i.e., Kolmogorov–Smirnov statistics) (Crawley *et al.*, 1995; Wu and Lewin, 1994; Xiong *et al.*, 1995a), or cross-correlation analysis (Bandettini *et al.*, 1993; Xiong *et al.*, 1995a).

This paper proposes a new dynamical clustering analysis (DCA) algorithm for an explorative examination of the detailed time courses of fMRI signals. Besides the choice of an appropriate distance measure this analysis requires no prior assumptions about the spatial location and extension of activation or the shape of the expected time course. DCA with the Euclidean distance measure was applied to data from echo planar imaging (EPI) with fast, repetitive measurements of a single slice during voluntary finger movement. The appropriateness of the Euclidean distance measure can be checked by comparison to a number of other results existing for this well-studied experimental paradigm (Gerloff *et al.*, 1996; Neafsey *et al.*, 1978; Okano and Tanji, 1987; Richter *et al.*, 1997; Thaler *et al.*, 1988; Wildgruber *et al.*, 1997). Other clustering methods that have been applied to fMRI data were *k*-means analysis (Ding *et al.*, 1994) and fuzzy *k*-means analysis (Baumgartner *et al.*, 1997, 1998; Jarmasz and Somorjai, 1998; Moser *et al.*, 1997). We have chosen a hard clustering algorithm (like *k*-means) since there are fewer free parameters to be estimated from the data and because the membership functions of a fuzzy result were usually thresholded and also interpreted as a hard assignment. There are other explorative analysis methods of time courses with few prior assumptions like PCA and ICA that will be briefly addressed in the discussion.

What is the goal of clustering analysis and what are

1053-8119/99 $30.00

the basic requirements on a particular algorithm? Ideally clustering should reveal structures in the fMRI data only based on similarities defined by the chosen distance measure. The hard clustering result consists of a disjunct partitioning of the data (the clusters) and a representing cluster center for each partition (cluster) (Duda and Hart, 1973). Each cluster contains voxels with similar signal courses (Ding *et al.*, 1994; Toft *et al.*, 1997). The set of cluster centers should be representive for the structure in the data that is caused by what is considered as the signals in the fMRI data. As signals we consider here not only changes in the blood oxygenation level (BOLD effect) but also other physiological components. This goal implies assumptions on the nature of the data which should influence the analysis result. A clustering result can be characterized by the quantization errors, i.e., the distances between representants and data points, and the reduction degree, i.e., the number of data points divided by the number of cluster centers k. If all data properties are considered as signal (noiseless data) one would simply like a vanishing mean quantization error with the highest possible reduction degree. For noisy data this would mainly reveal noise properties and therefore one will adjust to a finite quantization error depending on the assumed signal-to-noise ratio. For fMRI data at 1.5 T the signal-to-noise ratio is less than 5%. Essential for a clustering algorithm is how the result is influenced by parameter settings. This relationship should fulfill the following requirements: (i) The initial parameter setting should allow an easy adjustment of a desired quantization error. (ii) The result should be insensitive with respect to random parameters such as seeds and the particular random sequence in the update process (reproducibility of results).

We used these criteria to compare the new DCA algorithm with the standard k-means method. In the particular case of fast fMRI data we examined the relation between clustering and a much simpler and faster outlier detection (OD). As an additional feature DCA simultaneously calculates a two-dimensional visualization of the cluster centers to reveal the main data structure. The structuring by DCA allows a fast and very efficient further analysis of the data. On the set of cluster centers, their corresponding spatial distribution and the low-dimensional visualization, properties become salient that are very unlikely to be detected by a voxel-by-voxel-based analysis. Furthermore, we evaluated activity maps with cross-correlation analysis on single voxels and cluster centers to test and compare the clustering results. Note that with finite quantization error the activity maps obtained after clustering will not exactly coincide with single-voxel detection.

In the following we explain the new DCA algorithm in general and discuss the results of an application to fMRI data from experiments examining voluntary finger movements. The DCA activation maps resulting from different data preprocessing procedures are compared with each other and with standard cross-correlation results. Finally the DCA was used to examine sequential activation of supplementary motor area (SMA), primary motor cortex (M1), and sensory cortex (S1) in the voluntary motor task.

METHODS

fMRI was performed on a 1.5-T scanner (Siemens Vision) using a single-slice EPI sequence (TE = 43 ms, $\alpha = 40°$, FOV = 192 mm, 64 × 64 matrix, slice thickness 4 mm) with a repetition time (TR) of 107 ms. In each experiment 1024 subsequent images were acquired for a time period of 110 s. Eight right-handed healthy volunteers were instructed to press a button repeatedly as fast as possible with the right index finger during presence of a light signal. Four times the light signal was switched on for 5 s after a 20-s pause interval. For each volunteer five trials of these measurements were performed. Prior to the fMRI experiment high-resolution T1-weighted anatomical imaging was performed to determine an appropriate slice position that covers SMA, M1, and S1 (Rademacher *et al.*, 1992). To minimize movement artifacts the head was fixed by foam rubber within the head coil and a strap was placed over the forehead. Absence of disturbing head motion was ensured by viewing the series of images in cine mode. Four regions of interest (ROIs) were determined within the acquired axial slice corresponding to the SMA, M1, and S1 and superior sagittal sinus (SIN). The position of cerebral sulci on the anatomical images was used as reference for defining the motor regions (Rademacher *et al.*, 1992).

Before DCA the fMRI measurements acquired across all trials for each subject were either averaged (AVG) or concatenated (CAT): Averaging was performed across all trials for each subject to reduce noise and interference by physiological processes like breathing and heart rate. Concatenation was performed after removing the initial phase of the measurement. Noise reduction and suppression of interference by physiological processes were carried out by smoothing the data with a low-pass filter (successive 5× and 3× TR boxcar function). For type AVG and type CAT data voxels with a mean signal intensity under a heuristically determined threshold of 75 (belonging to voxels outside the brain) were excluded from further processing. Furthermore, the first 75 images (≡8 s) in the time courses were cut off to disregard onset effects of the fMRI signal until reaching a steady state of the longitudinal magnetization. Each signal time course was normalized to a mean signal intensity of zero in order to classify different time courses independent of intensity offsets caused by variation of EPI signal intensity of different

brain tissues (cerebrospinal fluid (CSF) and gray and white matter).

To test how different data preprocessing strategies influence the results of DCA, type AVG and CAT data were filtered in four alternative ways leading to five distinct data sets (A–E) for each subject: The data set (A) contained the unchanged type AVG or type CAT data. The data set (B) contained only time courses with an amplitude above a heuristically chosen intensity threshold (averaged data, five intensity steps; concatenated data, eight intensity steps). This filter excluded time courses with very small activity changes. Data set (C) contained only time courses of voxels in the previously specified ROIs focusing the analysis to brain regions that are expected to be involved in movement preparation and execution. The data set (D) contained only time courses satisfying the conditions of B and C. Data set (E) consisted of the same time courses as data set D, but the time courses were normalized to an equal amplitude. Signal normalization was introduced to study the partial volume effect: In voxels containing parts of white matter and CSF the induced signal changes might be reduced due to the smaller volume of gray matter and not only due to a lower hemodynamic activation level. To limit the distortion caused by amplification of low-amplitude noise the normalization was applied to the data set D.

On the preprocessed fMRI data sets of each subject an unsupervised DCA classifying the detailed time courses with an Euclidean distance measure was applied (Duda and Hart, 1973; Schwenker *et al.*, 1996). The basic DCA algorithm is shown in Fig. 1. This algorithm calculates simultaneously representative prototypes of time courses (Somorjai *et al.*, 1997; Toft *et al.*, 1997) and a corresponding two-dimensional (2D) visualization for online visualization of the results. Different from standard cluster analysis (k-means CA) or fuzzy k-means CA the DCA algorithm needs no prior initial specification on the number k and the shape of prototypes (seeds). DCA extracts these informations from the data employing generation and fusion processes of clusters at run time. The dynamical process of generation and fusion is controlled by two threshold parameters (see Fig. 2a). For each time course the Euclidean distance to all cluster centers is calculated and as usual in cluster analysis the time course is assigned to the closest center. However, if in DCA the lowest distance is greater than the generation threshold, the time course is introduced as a new cluster center. After each assignment or generation process the cluster centers

choose factor f and estimate thresholds θ_{new}, θ_{merge}
set $k = 0$ (no prototypes)
  choose a data point: $x \in X$
  calculate distances: $d_j = \mathrm{d}(x, c^j),\ j = 0, \ldots, k$
  detect winner: $j^* = \mathrm{argmin}_j(d_j)$
  if $(d_{j^*} > \theta_{new})$ or $k = 0$
    $c^k := x$ and $\begin{cases} p^k := p^{j^*} \pm \mathrm{random}(d_{j^*}) & : \ k > 0 \\ p^k := 0 & : \ k = 0 \end{cases}$
    $k := k + 1$
  else
    adapt c^{j^*} by (1) and p^{j^*} by (2)
    calculate distances: $D_l = \mathrm{d}(c^l, c^{j^*}),\ l = 0, \ldots, k$
    detect winner: $l^* := \mathrm{argmin}_{l \neq j^*}(D_l)$
    if $(D_{l^*} \leq \theta_{merge})$
      merge(c^{l^*}, c^{j^*})
      $k := k - 1$
    adapt thresholds: $\theta_{merge} := \theta_{merge} * f$
      $\theta_{new} := \theta_{new} * 1/f$
  Goto: choose data point

$$\Delta c^j = \frac{1}{|C^j| + 1}(x - c^j) \tag{1}$$

$$\Delta p^j = \eta * \alpha \sum_{i \neq j}^{k} \Phi'[d^2(p^i, p^j)] \left(\Phi[d^2(c^i, c^j)] - \Phi[d^2(p^i, p^j)]\right) \left(p^i - p^j\right) \tag{2}$$

FIG. 1. The basic dynamical cluster analysis algorithm (DCA) and 2D visualization of calculated cluster centers. $\mathrm{d}(\cdot, \cdot)$ denotes an appropiate distance measure, $\mathrm{argmin}_j(x_j)$ calculates the index j of the minimal component of vector x, random (x) is a random value function for real values in the range $0 \ldots x$, and merge $(\cdot, \cdot)$ denotes a procedure for merging two clusters.

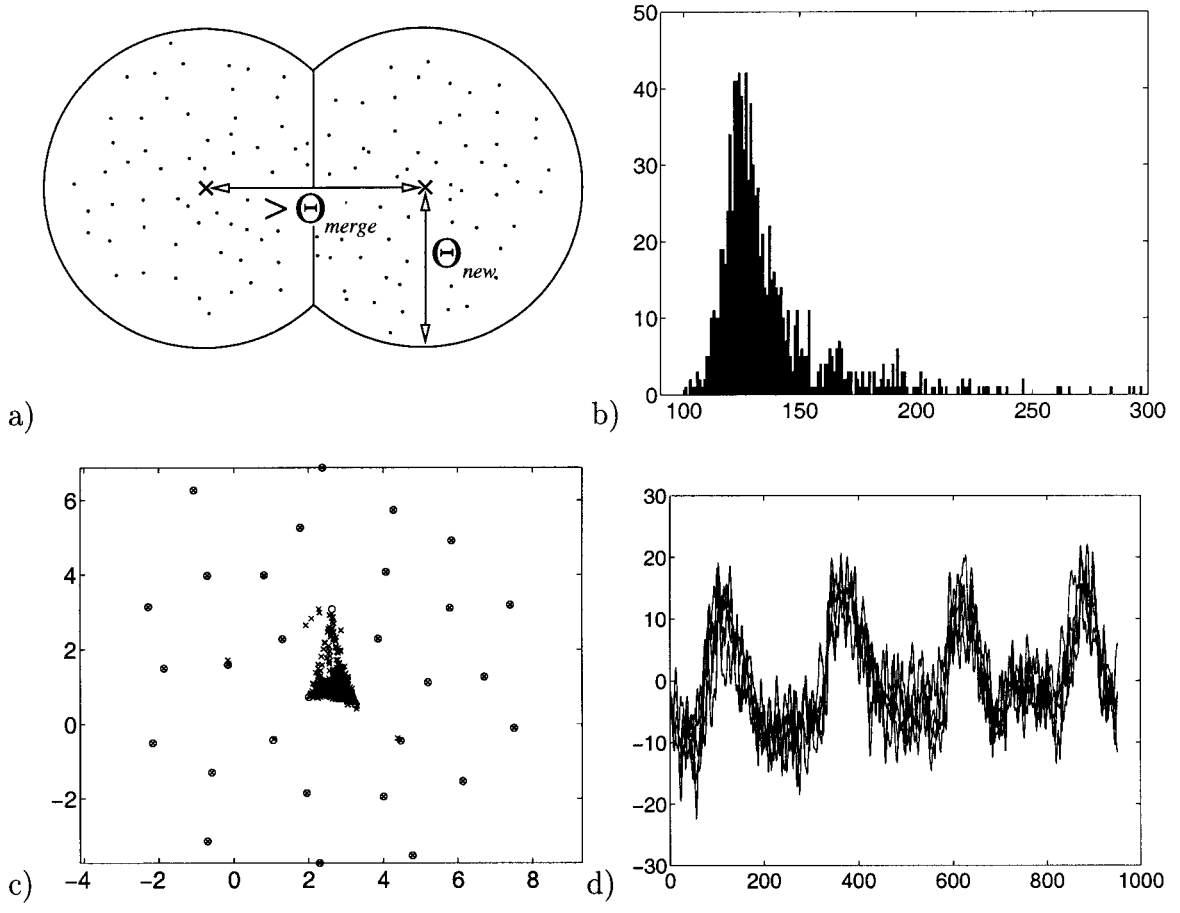

FIG. 2. (a) Schematic drawing of dynamical generation and fusion of cluster centers in DCA. Each cluster center (×) represents all data points (dots) with the smallest distance to the center. If the distances between a data point and all cluster centers are greater than a specified threshold θ_{new} (displayed by the circular borders), then the data point is used for generation of a new cluster center. If the distance between two cluster centers is smaller than a specified threshold θ_{merge} the two clusters are merged to one cluster. (b) Initial values of the thresholds are estimated by calculating a histogram of distances between the individual time courses and the mean signal time course over all voxels. For the generation threshold a value at 90% of the accumulation peak area is taken, in this example 175. The fusion threshold is chosen at 80% of the accumulation peak, here 160. (c) 2D visualization of cluster centers (circles) and data points (crosses). (d) Example of single-voxel time courses within a cluster center.

are adapted to always represent the mean of all time courses with the smallest distance to the center. If the distance between two cluster centers drops below the fusion threshold, the two clusters are merged into one cluster.

For the initial threshold setting the mean signal time course over all voxels in each data set was calculated. In the distance histogram between individual time courses and the general mean time course the generation threshold was set to the value of 90% of the accumulation peak area (see Fig. 2b). The fusion threshold was set to 80% of the accumulation peak area. After an initial phase the appropriate number of cluster centers becomes more and more stable, but sometimes generation and fusion processes do not stop. Therefore, the two thresholds are successively modified after each learning epoch, i.e., one complete run through the data set. The generation threshold is adapted to higher values, and the fusion threshold to lower values ($f = 0.99$, see Fig. 1) in a manner that keeps the total number of centers unchanged. Successively, this adjustment leads to a decreasing number of generation and fusion processes and to a transition to the standard *k*-means cluster algorithm, where convergence has been proven (Duda and Hart, 1973).

The cluster centers are projected into a 2D space by multidimensional scaling (Jain and Dubes, 1988; Schnell, 1994) to provide a survey about the distance structure in the high-dimensional space of the time courses (see Fig. 2c). A gradient descent algorithm on the stress function is used (Palm and Schwenker, 1996; Schwenker *et al.*, 1996), so that the 2D distance structure of center projections and high-dimensional distance structure of corresponding cluster centers become as similar as possible. 2D projection of the data points is done by a radial basis function network

(Haykins, 1994) with Gaussian functions, where the means are the high-dimensional cluster centers and the standard deviation is the ninth of the generation threshold in DCA. The output vector sum of the radial basis functions is normalized to 1 and used as weights for a linear combination of the 2D cluster center projections.

DCA yielded a set of cluster centers for each fMRI data set of each subject. A cluster center is the representative signal time course for a subset of voxels with similar time courses (see Fig. 2d). The members of a cluster can be displayed as a spatial pattern in the corresponding anatomical MRI image. We defined a cluster to be active if the cross-correlation peak between cluster center and reference function exceeded a heuristically determined threshold of 0.5 (see Bandettini *et al.*, 1995a). The reference function was defined, after visual inspection of the DCA cluster centers of all subjects, as four-step functions with 3 s off and on phase directly before and after the external trigger (start of light trigger). Values of the reference function outside the off/on time intervals were undefined and not used for cross-correlation. The chosen off/on length corresponds to the minimum expected hemodynamic response length of about 3 s and was confirmed by the DCA results. A DCA activation map was defined as the superimposed spatial patterns of all active clusters. The response delay of a voxel was defined as the time step where the cross-correlation function between the corresponding time course and the reference function reached its maximum. Further, the response delay of a ROI was defined as the average response delay of all active voxels in the ROI. The pairwise differences of the ROI response delays were defined as time shifts of activation between the regions.

For comparison of the DCA with other methods the whole data set (AVG A) (see above) was additionally analyzed with *k*-means CA (Duda and Hart, 1973) for different *k* values. The initial cluster center seeds in *k*-means CA were determinated randomly from the analyzed data set. Activation detection was performed analogously to DCA. For each parameter setting we performed $l = 10$ runs. To quantify reproducibility of the results of different runs of a method we considered the mean Euclidean distance between corresponding clusters. Corresponding clusters for different runs were defined by the membership of the data points to the clusters (i.e., for each data point the representing cluster centers in the different runs were defined as corresponding clusters). The difference between the results of two runs was measured by the Euclidean distance between corresponding cluster centers averaged over all data points. The variability in the results of the different runs was measured by the mean value of the pairwise differences of the runs, i.e., by averaging the $l(l-1)/2$ pairwise distances.

To compare the DCA activation maps with a standard fMRI evaluation method the data set AVG A was also analyzed with standard cross-correlation analysis (Bandettini *et al.*, 1993; Xiong *et al.*, 1995a). In this case we defined a voxel as active if the cross-correlation between its time course and action function exceeded a threshold of 0.3. In order to reduce false positives the locations of activation were spatially filtered with a kernel of 1 voxel (3 mm) radius, i.e., only voxels with surrounding neighbors over the specified correlation threshold were defined as active.

RESULTS

Reproducibility of the results has been defined in the introduction as one of our crucial requirements for a clustering method. Depending on the random choice of seeds and random selection in the update order of data points in a particular method, results of different runs with the same parameters can vary. Therefore, we first tested the variability of DCA results and compared it with the *k*-means CA with different *k* values on the data of one subject (type AVG A data) using the Euclidean distance. While the repeated DCA runs led to almost the same number (±2%) of cluster centers, comparison of the variabilities in the results of *k*-means CA and DCA in Table 1 shows significantly lower values for the DCA reflecting smaller deviations of the cluster partitioning and cluster center shapes in the different runs of DCA. Thus, cluster generation and fusion in the adaptive phase of DCA improves the reproducibility of the results.

DCA was applied to type AVG A data of the eight subjects with the threshold setting as described under Methods. We kept the DCA thresholds fixed since there is no good argument why clustering results of different data sets should have identical *k* to be compared. The

TABLE 1

Variability of Cluster Center Sets Obtained in Repeated Runs of *k*-Means CA and DCA

Centers	Mean variability
k-means	
10	17.59 (±31.31)
20	18.46 (±28.88)
31	22.16 (±27.93)
40	25.02 (±28.5)
50	26.56 (±32.33)
60	26.36 (±32.97)
DCA	
31	3.12 (±13.03)

Note. Shown are the mean Euclidean distances (and standard deviation in parentheses) between corresponding cluster centers of 10 independent analysis runs for each CA method and each number of cluster centers.

DCA results comprised on average 33 cluster centers (standard deviation, 9.2). An example for the similarity of shapes of time courses that were collated in a single cluster by DCA with the Euclidean distance can be seen in Fig. 2d. Figure 2c displays for one subject the two-dimensional representation of the (high-dimensional) cluster centers and single-voxel time courses, respectively. Note first the very uneven distribution of voxel counts over the set of cluster centers. We found a dichotomy between a very few clusters (1–3) representing 90–95% of the data with high quantization error and cluster centers with small membership numbers representing their data points with low quantization error. This finding was typical for all subjects and qualitatively also observed with the *k*-means method. Variation of the thresholds from the highest distance between the data points to lower values showed that the bipartition is already observable with 2–3 cluster centers and is preserved for lower values. With lowering the fusion threshold the total number of cluster centers is increased. We compared the clustering results with a much simpler computation of such a bipartition by OD. As outliers we defined data points outside an Euclidean vicinity sphere around the general mean signal time course (see Fig. 3c). Note, second, that the low dimensionally projected cluster centers in Fig. 2c are almost equidistant. We also checked the distances between the high-dimensional cluster centers and found for all subjects a quite regular high-dimensional hypertetrahedron.

In the remainder of this section we describe one particular aspect of the clustering analysis: detecting activated clusters. Table 2 shows properties of activity maps obtained by applying the same correlation criterion as described under Methods after DCA and *k*-means CA analysis, respectively. The mean sizes of activation spots obtained with *k*-means CA were smaller compared to DCA, even for cases with a higher number of cluster centers than in DCA. As expected also the activity maps after *k*-means clustering were less stable in repeated runs (deviation of activated voxels $>33\%$) than those obtained after DCA clustering. For many runs the DCA activation maps were identical or deviated only in a small percentage of voxels (deviation $<7\%$).

Exemplary activation maps for one subject obtained by different methods are displayed in Figs. 3a–3c. Figure 3a gives the anatomical image with the standard activation map obtained by single-voxel cross-correlation detection overlayed in white. Figure 3b shows the activity map after the DCA of data set AVG A. Inactive clusters are displayed in gray tones, and active clusters with colors. A comparison of Fig. 3b with the anatomy in Fig. 3a reveals that most activation detected by DCA was within brain regions that were expected to be active in the performed motor task. The DCA-activated regions formed by all colored clusters agree (with the exception of some voxels in the area of the superior sagittal sinus) very well with the cross-correlation analysis map. However, DCA provides more detailed information about different activation time shapes by the additional discrimination between different activated clusters. Figure 3c displays the result of OD applied to the data set AVG A. For the radius of the vicinity sphere we chose a value equal to the generation threshold in DCA. Comparison of Figs. 3b and 3c shows that the active voxels obtained with DCA were mostly contained in the OD filtered pattern.

Characteristic signal time courses of cluster centers obtained with the DCA are shown in Figs. 3d and 3e. Figure 3d displays active cluster centers that clearly reveal the four periods of the performed motor task, but in addition have differences in shape and onset times between the cluster centers, for example, the onset time difference between the second and the third from the bottom. The corresponding spatial pattern of the third cluster is distributed over M1 and S1 (see Fig. 3b). Comparable patterns of active clusters extending into multiple ROIs were found in all subjects. Time courses of inactive clusters are given in Fig. 3e. The upper cluster is localized in the region of the superior sagittal sinus and clearly reveals blood pulsations due to heart beat and respiration. The two remaining signal courses in Fig. 3e belong to voxels located outside the defined ROIs. Again, comparable types of signal courses were found for all subjects.

Figure 4 shows effects on the activity maps detected after DCA caused by the various filtering procedures described under Methods (AVG A-AVG E) and the OD (colors of the active clusters correspond to the mean onset time of the corresponding voxel time courses). Activity maps corresponding to different filtering differ in detail but are distributed in the same brain areas. In general, data reduction by filtering diminished the area of the activation, as can be seen by comparing Figs. 4b and 4d with Fig. 4a. Combined application of two filters further reduced the activated voxels (compare Fig. 4e with Figs. 4b and 4d). With amplitude normalization after twofold filtering (see Fig. 4e) the resulting activation area increased and became spatially more continuous (see Fig. 4f). Cluster centers for data with amplitude normalization are displayed in Fig. 5. Note that the first three time courses of Fig. 3d now form a single cluster. Figure 4c shows activity maps obtained after DCA on data that were reduced by OD filtering (vicinity sphere radius is equal to generation threshold in DCA). A comparison with the activity map of the data set AVG A (see Fig. 4a) exemplifies our general observation that cluster analysis (*k*-means and DCA) of the fMRI data singled out the same voxels that can be found by OD. Figures 4g and 4h show activation maps after DCA on concatenated data sets (CAT) where voxels with small amplitudes have been removed (CAT B, Fig. 4g), and

with additional restriction to the ROIs (CAT D, Fig. 4h). Averaging or concatenating (with low-pass filtering) led to quite similar activation maps: despite an increase of the data dimension by concatenating DCA clustering seems to be similar.

Table 3 summarizes the averaged time shifts over all subjects in seconds between the examined ROIs obtained by the DCA method on data set AVG A. These results indicate sequential activation of SMA, M1, and S1 in the given order, which has also been found in other fMRI studies (Richter *et al.*, 1997; Wildgruber *et al.*, 1997). Moreover, the delay time between SMA and M1 is in the same range of magnitude as reported by electrophysiological experiments (Gerloff *et al.*, 1996; Neafsey *et al.*, 1978; Okano and Tanji, 1987; Thaler *et al.*, 1988).

TABLE 2

Comparison between Activation Maps Obtained with *k*-Means CA and DCA

Centers	Mean active	Max. active	Active hull	*P* (active)
		k-means		
10	32.4 (±7.75)	50	51	0.635 (±0.15)
20	36.8 (±13.68)	69	69	0.533 (±0.15)
31	31.7 (±12.29)	65	67	0.473 (±0.16)
40	39.1 (±17.11)	74	91	0.43 (±0.13)
50	35.4 (±10.7)	56	75	0.47 (±0.13)
60	35.6 (±12.4)	59	70	0.5 (±0.12)
		DCA		
31	71.1 (±4.73)	78	80	0.88 (±0.06)

Note. Each line of the table expresses results after 10 repeated runs with the same parameters. The columns of the table display the number of cluster centers; the mean number of activated voxels; the maximal number of activated voxels in a single run; the size of the hull of activated voxels in all runs; i.e., the superset of activation maps; and the mean probability that a single run classifies a voxels in the hull as active—i.e., a low probability means a high variability in the activation detection. Standard deviations of the mean values are given in parentheses.

DISCUSSION

We proposed a new DCA method using the Euclidean distance between detailed time courses of voxels to reveal hemodynamic processes in the brain. This approach raises two questions:

(a) Is the chosen distance measure appropriate? To examine this question we chose a well-known experimental paradigm where the activity maps calculated on the clustering results could be compared with data from the literature obtained with a variety of brain mapping techniques.

(b) What are the advantages of the new clustering algorithm? To address this question we checked the requirements on a clustering algorithm for fMRI analysis formulated in the introduction. The new clustering method has a much improved reproducibility: Table 1 shows that for repeated runs on the same data sets the variability of the proposed new cluster algorithm DCA is significantly below that of *k*-means. The adjustment parameters in DCA are the initial values for the thresholds governing cluster fusion and generation. These parameters relate more directly to the quantization error than the parameter *k* in *k*-means analysis (Duda and Hart, 1973) and fuzzy *k*-means analysis that directly sets the reduction degree. Our initial threshold setting where the integral over the distance histogram in the data set assumed 90/80% corresonds to an analysis characteristic far away from small overall quantization error (which would be provided by thresholds at low values of the integral). The 90/80% setting heuristic could be uniformly applied to all examined data sets. Since the tradeoff between reduction degree and quantization error varies from data set to data set the *k* value in *k*-means analysis must often be optimized iteratively by a result driven search for each data set (cluster validity problem).

The dynamical process of cluster generation and fusion does not necessarily converge for fixed thresholds. In a phase where the total number of cluster has settled—what we always observed—the thresholds in DCA are adapted so that a successive relaxation to the *k*-means algorithm is performed. The termination of DCA is therefore ensured by the provable convergence of the *k*-means algorithm (Duda and Hart, 1973); however, convergence time might vary for different initial threshold settings. Generally, CA is a computationally expensive analysis method for fMRI data: *k*-means CA takes about 1.5–2 times and DCA takes about 5–7 times the CPU time of standard cross-correlation analysis. Of course, this comparison does not take into account that CA provides richer information about the data than cross-correlation analysis.

There are a variety of ways for the further use of the DCA results. The prototypes (cluster centers) allow a

FIG. 3. Results in one subject determined by DCA vs standard correlation and vicinity filtering. (a) The anatomical image with defined ROIs and overlay of activation obtained with standard cross-correlation (white). (b) Distributions of voxels corresponding to the various clusters obtained with the DCA. Gray tones display inactive and colors active clusters. (c) Spatial distribution of voxels remaining after masking a vicinity sphere around the general mean signal time course of all voxels. (d) Representative signal time course of selected active cluster centers. The time courses are marked with membership counts of voxels and corresponding colors in b. Small bars below the time courses show the application phases of the external trigger. (e) The center of a cluster located in the sagittalis sinus (top) and two other centers of cluster located outside the defined ROIs (two bottom rows).

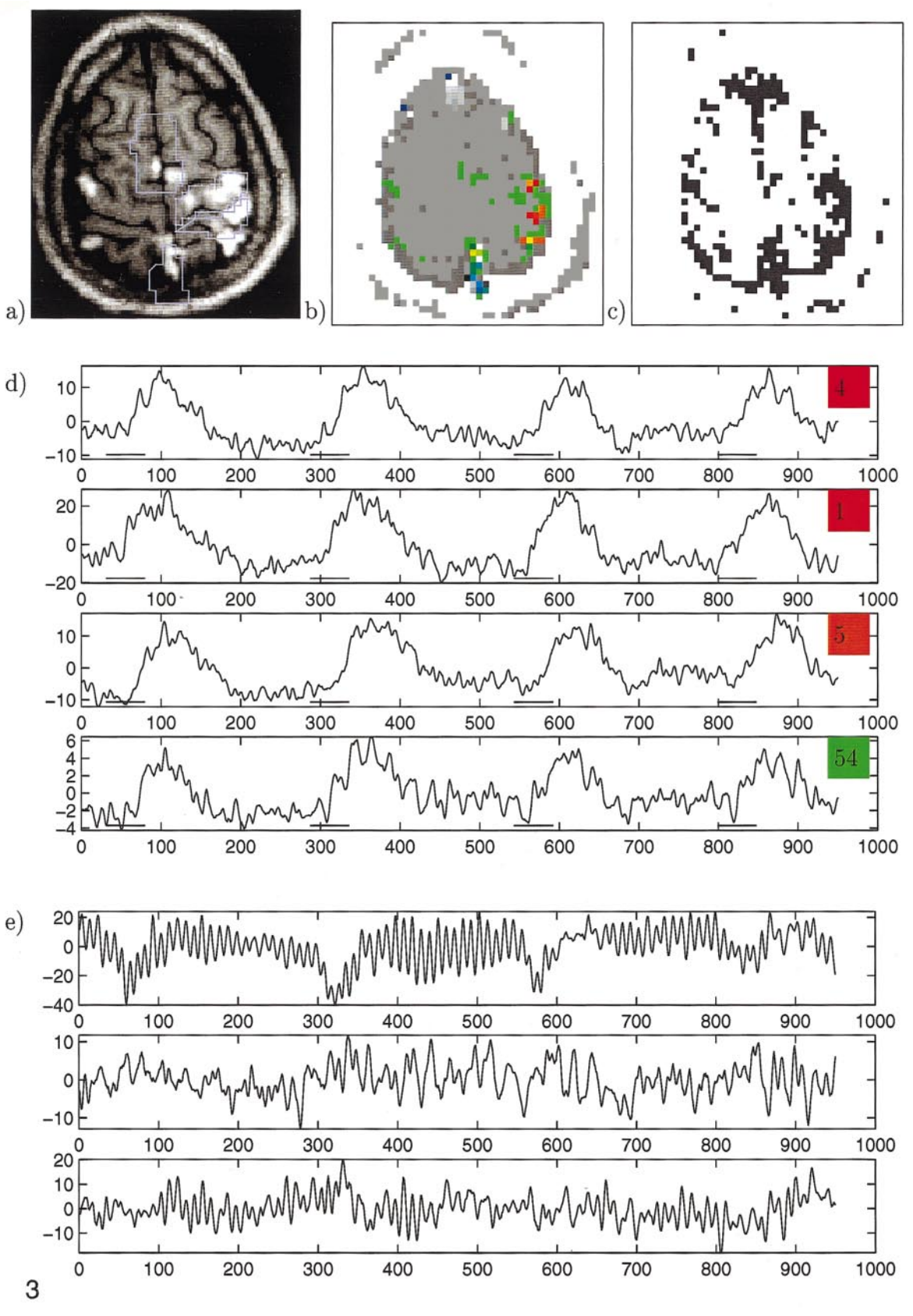

FIGURE 3

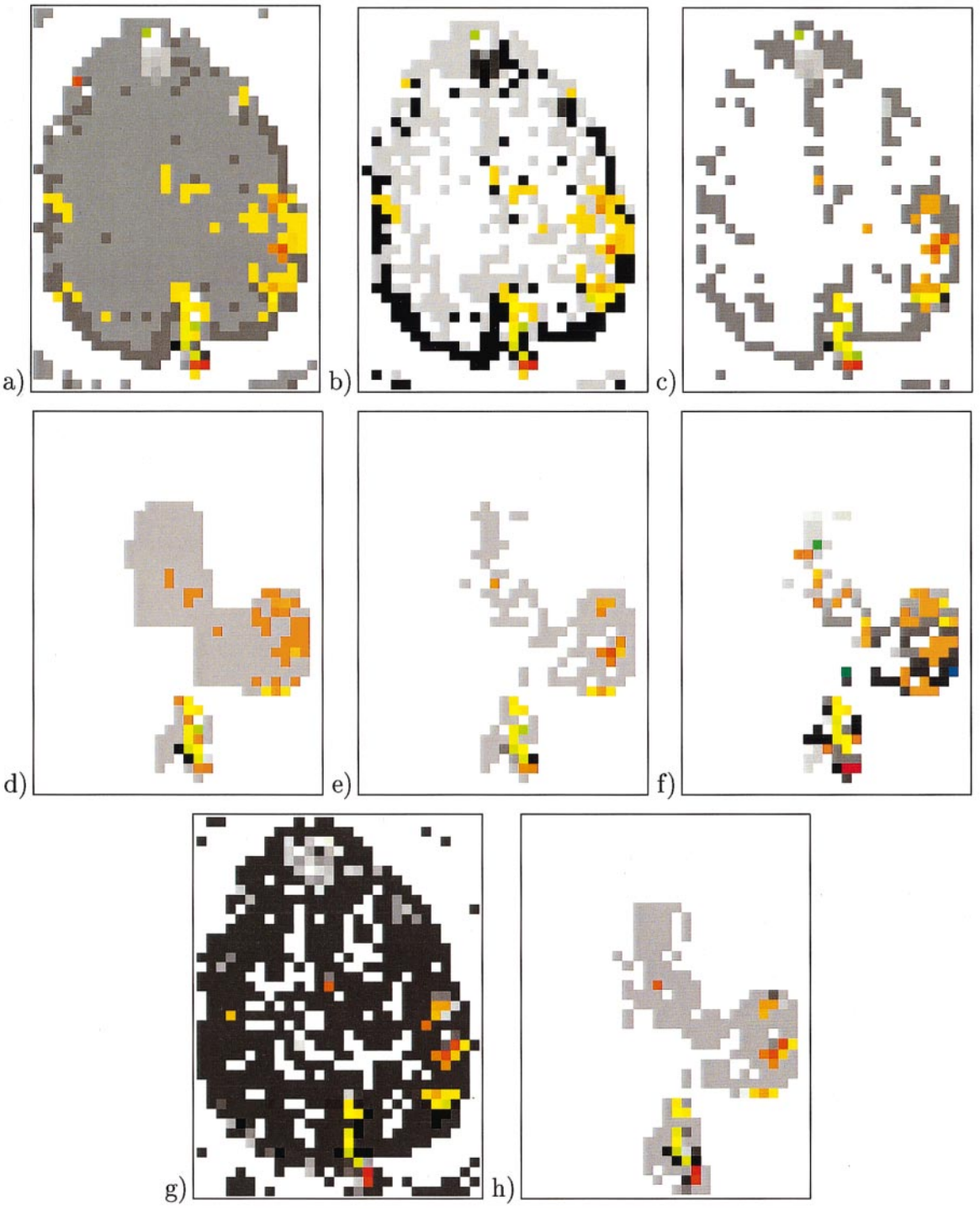

FIG. 4. Comparison of DCA results using the different sets of filtered data of one subject. The colors code different onset times of the hemodynamic response in active clusters. Colors from red, orange, yellow, and green to blue correspond to response times from early to late. (a,b,d,e) DCA activation maps for the data sets AVG A–AVG D. (c) Result of the DCA after OD filtering. (f) Result of the DCA on the equal amplitude normalized data set AVG E. (g,h) The DCA results on the concatenated data sets CAT B+CAT D.

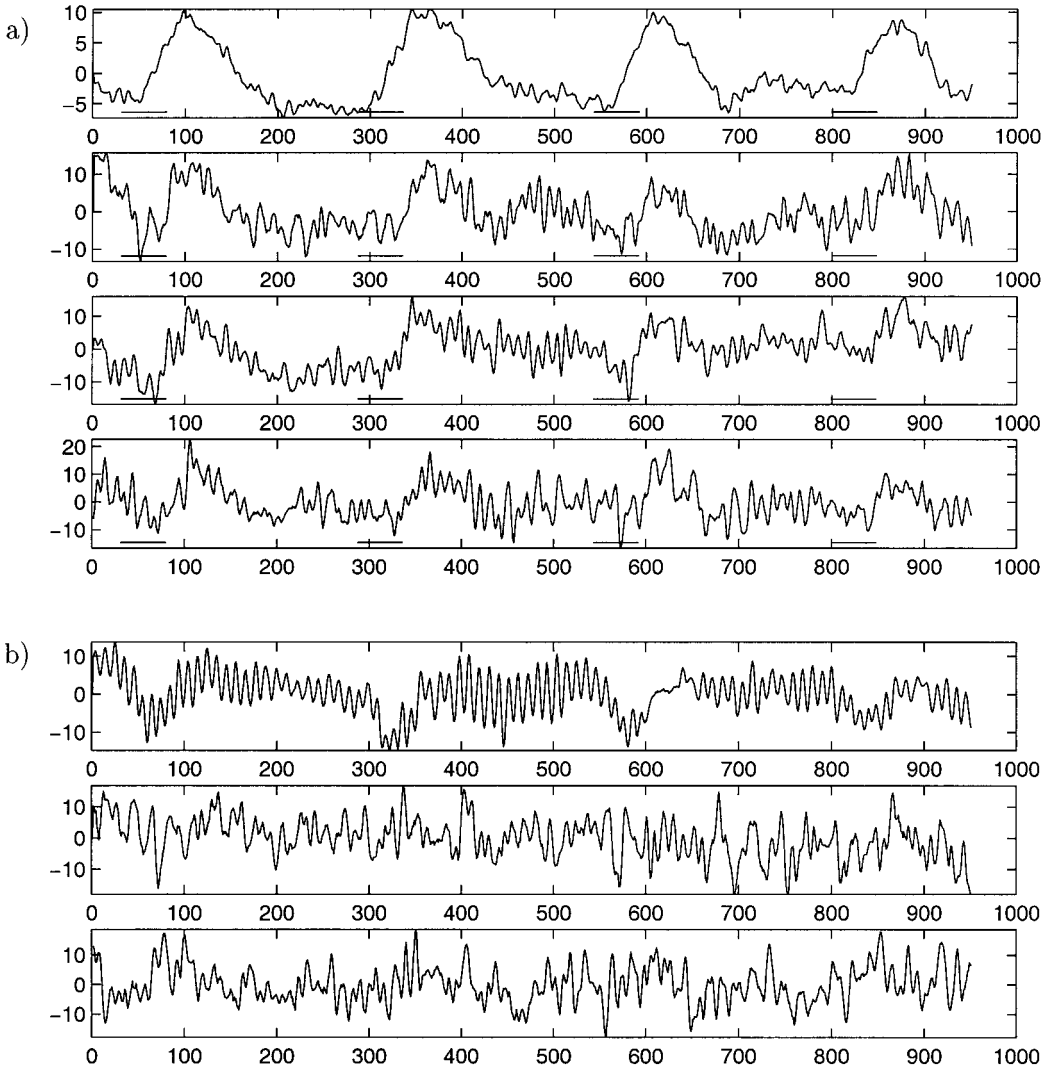

FIG. 5. Typical time courses of clusters obtained with amplitude normalization. (a) Time courses of active cluster centers. (b) The center of a cluster located in the sagittalis sinus (top) and two other centers of cluster located outside the defined ROIs (two bottom rows).

valuable survey about the predominating shape characteristics in the data set. Clusters with dominating components of changes due to the BOLD effect, physiological components, and sometimes also measurement artifacts are easily recognizable (for instance, we discovered a signal component presumably caused by the AC heating of the VHF transmitter tube in the Siemens Vision). The spatial distribution of clusters can be displayed as an overlay of the corresponding maps onto the anatomical image. Cluster centers can be selected with respect to interesting features salient in time shape and spatial distribution and can be further analyzed on a voxel-by-voxel basis. It would be highly unlikely to find these voxel sets with single-voxel tests.

We applied our approach to analyze signal courses by DCA with the Euclidean distance measure on the concrete example of movement-induced hemodynamic changes. For the well-studied voluntary movement task, correlation analysis on the cluster centers revealed the expected activation maps (Figs. 3a and 3b) and active clusters were already visually salient in the provided data survey (Figs. 3d and 3e). Beyond correlation analysis the clustering methods are capable of

TABLE 3

Mean Time Shifts in Seconds, and Standard Deviation in Parentheses, between the Examined Brain Regions in All Subjects

$\Delta_{SMA\text{-}M1}$	$\Delta_{M1\text{-}S1}$	$\Delta_{S1\text{-}SIN}$	$\Delta_{SMA\text{-}S1}$	$\Delta_{SMA\text{-}M1/S1}$
0.35 (±0.96)	0.12 (±0.66)	0.73 (±0.54)	0.47 (±0.36)	0.41 (±0.65)

Note. $\Delta_{SMA\text{-}M1}$, $\Delta_{M1\text{-}S1}$, $\Delta_{S1\text{-}SIN}$, and $\Delta_{SMA\text{-}S1}$ denote the time shifts between SMA and M1, M1 and S1, S1 and SIN, and SMA and S1, respectively. $\Delta_{SMA\text{-}M1/S1}$ is the time shift between activated voxels contributing to SMA and active voxels in M1 and S1 together.

discriminating between different levels of activation (Baumgartner *et al.*, 1997, 1998; Jarmasz and Somorjai, 1998; Moser *et al.*, 1997), even if the temporal pattern is indistinguishable. However, the interpretation of such different levels of activation is not straightforward because of the partial volume effect. We investigated this question by clustering after amplitude normalization. This increased the similarity of the time curves (Fig. 5) and led to a lower number of activated cluster centers found by DCA. Comparison of twofold filtering (Fig. 4e) and twofold filtering plus amplitude normalization (Fig. 4f) showed additional activated voxels in the neighborhood of previously activated voxels. Comparison of Fig. 4f with Figs. 3b and 3d shows that the additional activated voxels after amplitude normalization were mostly voxels with a lower amplitude of activation (green voxels in Fig. 3b). Because of the distribution of activated voxels with a low amplitude (green voxels in Fig. 3b) in the neighborhood of voxels with higher activation amplitudes (red and orange voxels in Fig. 3b), we interpret the variation of amplitudes as an effect of the partial volumes. However, the partial volume effect cannot be appropriately corrected by amplitude normalization because of the low signal-to-noise ratio in time courses with small BOLD changes. Amplitude normalization yields only reasonable activity maps after previous data filtering (see Methods). In a forthcoming paper we will consider how the partial volume effect can be reduced if the anatomical structure and geometrical neighborship is used for a location-dependent amplitude renormalization. Hard clustering with DCA with such a preprocessing must then be compared with methods separating intrinsically several superimposed signal components in single voxels like true fuzzy clustering techniques, PCA, ICA, or the recently proposed blind separation analysis (McKeown *et al.*, 1998). The full exploitation of the membership function as well as blind source separation also require additional assumptions about the spatial distribution of activation.

The 2D visualization showed a remarkably even distribution of the centers reflecting an almost regular hypertetrahedron configuration of the centers (of course, there is no exact distance preserving projection of this configuration in fewer dimensions). The hypertetrahedron configuration of the centers might be a consequence of the high dimension of our data (950 for data set AVG; 4750 for data set CAT) that is of the same order or higher than the number of data points. It must be emphasized that in this domain even most of the common clustering algorithms are not well tested; however, our results indicate that DCA can be applied. The 2D visualization further showed that always a high fraction of data is represented by a few clusters. These findings suggest that for the fMRI data interesting data points like activated voxels are in small clusters of outliers in the data. This assumption was confirmed by our results with the simple OD filter.

Time courses of single voxels belonging to activated clusters have been used to determine the mean response times of the ROIs considered in the motor task. An advantage of cluster methods is that they naturally reveal temporal and spatial activation processes independent from ROIs (Ding *et al.*, 1994). We observed that often a single activated cluster extented into different ROIs—signifying similar time courses at different loci. This could be an indication of functional connectivity (Friston *et al.*, 1993; Strother *et al.*, 1995) between ROIs that are responsible for movement preparation and execution. Single cell recordings performed in monkeys (Okano and Tanji, 1987) and cats (Neafsey *et al.*, 1978; Okano and Tanji, 1987) during voluntary movements reported in each of the regions SMA and M1 cell populations with early and late neural response times. These results imply that there is no well-defined onset time of the whole ROI, which might also explain the observed high standard deviation of onset times within the ROIs. Furthermore, the obtained time shifts might not necessarily reflect the time delays of neuronal activation, in particular, if hemodynamic response characteristics vary in the different ROIs (Binder *et al.*, 1993; Buckner *et al.*, 1996; Lee *et al.*, 1995); however, a recent paper of Menon *et al.* (1998) reports that the relative timing between the onset of the fMRI response in different brain areas appears to be preserved.

In summary, we have examined cluster analysis on the time courses of (fast) fMRI data using the Euclidean distance. DCA has been proposed as a new adaptive clustering method. We have defined important requirements of a clustering algorithm for noisy data and have shown that DCA meets them much better than the *k*-means method. We have checked the DCA results for different data preprocessing by comparing activated clusters which are detected by the usual correlation criterion and determined response delays of different ROIs involved in the motor task. However, onset times within clusters show a high standard deviation and therefore alternative distance measures and preprocessing might be more appropriate for this feature. These are matters of our further investigation.

ACKNOWLEDGMENTS

This study was financially supported by Fortüne No. 282, the research program of the University Hospital Tübingen. Intellectual support in stimulating discussions with F. Schwenker (University of Ulm, Ulm, Germany) is gratefully acknowledged. We also further acknowledge Uwe Klose (University of Tübingen, Tübingen, Germany) especially for programming of the fast EPI sequence.

REFERENCES

Bandettini, P. A., Jesmanowicz, A., Wong, E. C., and Hyde, J. S. 1993. Processing strategies for time-course data sets in functional MRI of the human brain. *Magn. Reson. Med.* **30:**161–173.

Bandettini, P. A., Wong, E. C., Binder, J. R., Rao, S. M., Jesmanowicz, A., Aaron, E. A., Lowry, T. F., Forster, H. V., Hinks, R. S., and Hyde, J. S. 1995a. Functional MR imaging using the BOLD approach: Dynamic characteristics and data analysis methods. In *Diffusion and Perfusion Magnetic Resonance Imaging* (D. Le Bihan, Ed.), pp. 335–349. Raven Press, New York.

Bandettini, P. A., Binder, J. R., DeYoe, E. A., Rao, S. M., Jesmanowicz, A., Hammeke, T. A., Haughton, V. M., Wong, E. C., and Hyde, J. S. 1995b. Functional MR imaging using the BOLD approach: Applications. In *Diffusion and Perfusion Magnetic Resonance Imaging* (D. Le Bihan, Ed.), pp. 335–349. Raven Press, New York.

Baumgartner, R., Scarth, G., Teichtmeister, C., Somorjai, R., and Moser, E. 1997. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. I. Reproducibility. *J. Magn. Reson. Imag.* **7:**1094–1101.

Baumgartner, R., Windischberger, C., and Moser, E. 1998. Quantification in functional magnetic resonance imaging: Fuzzy clustering vs. correlation analysis. *Magn. Reson. Imag.* **16**(2):115–125.

Le Bihan, D., Jezzard, P., Turner, R., Cuenod, C. A., Pannier, L., and Prinster, A. 1993. Problems and limitations in using Z-maps for processing of brain function MR images. *SMR, Book of Abstracts,* 11.

Binder, J. R., Jesmanowicz, A., Rao, S. M., Bandettini, P. A., Hammeke, T. A., and Hyde, J. S. 1993. Analysis of phase differences in periodic functtional MRI activation data. In *Proceeding of the Society of Magnetic Resonance in Medicine, 12th Annual Scientific Meeting, New York,* Vol. 3, p. 1383.

Buckner, R. L., Bandettini, P. A., O'Craven, K. M., Savoy, R. L., Petersen, S. E., Raichle, M. E., and Rosen, B. R. 1996. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA* **93:**14878–14883.

Cohen, M. S., and Bookheimer, S. Y. 1994. Localisation of brain function using magnetic resonance imaging. *TINS* **17**(7):268–277.

Crawley, A. P., Wood, M. L., and Mikulis, D. J. 1995. Evaluation of data processing methods for fMRI. *SMR/ESMREMB, Book of Abstracts,* 819.

Ding, X., Tkach, J., Ruggieri, P., and Masaryk, T. 1994. Analysis of time-course functional MRI data with clustering method without use of reference signal. In *Proceedings of the International Society for Magnetic Resonance, 2nd Meeting, San Francisco, California,* Vol. 2, p. 630.

Duda, R., and Hart, P. 1973. *Pattern Classification and Scene Analysis.* Wiley, New York.

Forman, S. D., Cohen, J. D., Fitzgerald, M., and Eddy, W. F. 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of acluster-size threshold. *Magn. Reson. Med.* **33:**636–647.

Friston, K. J., Firth, 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.

Gerloff, C., Grodd, W., Altenmüller, E., Kolb, R., Naegele, T., Klose, U., Voigt, K., and Dichgans, J. 1996. Coregistration of EEG and fMRI in a simple motor task. *Hum. Brain Map.* **4:**199–209.

Haykins, S. 1994. *Neural Networks—A Comprehensive Foundation,* Macmillan Co., New York.

Jain, A., and Dubes, R. 1988. *Algorithms for Clustering Data.* Prentice Hall, Englewood Cliffs, NJ.

Jarmasz, M., and Somorjai, R. L. 1998. Time to join! Cluster-merging in unsupervised fuzzy clustering of functional MRI data. In *Proceedings of the International Society for Magnetic Resonance in Medicine, 6th Scientific Meeting and Exhibition, Sydney,* p. 2068.

Kim, S.-G., Richter, W., and Uğurbil, K. 1997. Limitations of temporal resolution in functional MRI. *Magn. Reson. Med.* **37:**631–636.

Lee, A. T., Glover, G. H., and Meyer, C. H. 1995. Discrimination of large venous vessels in time-course spiral blood-oxygen-level-dependent magnetic-resonance functional neuroimaging. *Magn. Reson. Med.* **33:**745–754.

McKeown, M. J., Makeig, S., Brown, G. G., Jung, T.-P., Kindermann, S. S., Bell, A. J., and Sejnowski, T. J. 1998. Analysis of fMRI data by blind separation into independant spatial components. *Hum. Brain Map.* **6:**160–188.

Menon, R. S., Luknowsky, D. C., and Gati, J. S. 1998. Mental chronometry using latency-resolved functional MRI. *Proc. Natl. Acad. Sci. USA* **95:**10902–10907.

Moser, E., Diemling, M., and Baumgartner, R. 1997. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. II. Quantification. *J. Magn. Reson. Imag.* **7:**1102–1108.

Neafsey, E. J., Hull, C. D., and Buchwald, N. A. 1978. Preparation for movement in the cat. I. Unit activity in the cerebral cortex. *Electroencephalogr. Clin. Neuropysiol.* **44:**706–713.

Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. USA* **87:**9868–9872.

Okano, K., and Tanji, J. 1987. Neuronal activities in the primate motor fields of the agranular frontal cortex preceding visually triggered and self-paced movement. *Exp. Brain Res.* **66:**155–166.

Palm, G., and Schwenker, F. 1996. Adaptive analysis and visualisation in high dimensional data spaces. *Cybernet. Syst. '96, Vienna, Austria,* **2:**1009–1013.

Rademacher, J., Galaburda, A. M., Kennedy, D. N., Filipek, P. A., and Caviness, V. S. 1992. Human cerebral cortex: Localization, parcellation, and morphometry with magnetic resonance imaging. *J. Cogn. Neurosci.* **4:**352–374.

Richter, W., Andersen, P. M., Georgopoulos, A. P., and Kim, S.-G. 1997. Sequential activity in human motor areas during a delayed cued finger movement task studied by time-resolved fMRI. *NeuroReport* **8:**1257–1261.

Schnell, R. 1994. *Graphisch gestützte Datenanalyse.* R. Oldenbourg Verlag, München/Wien.

Schwenker, F., He, J., Kestler, H., Littmann, E., Schieszl, S., and Palm, G. 1996. Anwendungen neuronaler Netze. In *Finanzmarktanalyse und -prognose mit innovativen Verfahren.* (G. Bol, G. Nakhaeizadeh, and K. H. Vollmer, Eds.), pp. 35–69. Physica-Verlag, Heidelberg.

Somorjai, R. L., Dolenko, B., Scarth, G., and Summers, R. 1997. How different are the clusters in fMRI?—A multivariate statistical assessmentt using profile analysis. In *Proceedings of the International Society for Magnetic Resonance in Medicine, 5th Scientific Meeting and Exhibition, Vancouver, B.C.,* Vol. 1, p. 354.

Strother, S. C., Kanno, I., and Rottenberg, D. A. 1995. Principal component analysis, variance partitioning, and "functional connectivity." *J. Cereb. Blood Flow Metab.* **15:**353–360.

Thaler, D. E., Rolls, E. T., and Passingham, R. E. 1988. Neuronal activity of the supplementary motor area (SMA) during internally and externaly triggered wrist movement. *Neurosci. Lett.* **93:**264–269.

Toft, P., Hansen, L. K., Nielsen, F. Å., Strother, S. C., Lange, N., Mørch, N., Svarer, C., Paulson, O. B., Savoy, R., Rosen, B., Rostrup, E., and Born, P. 1997. On clustering of FMRI time series. In *Third*

International Conference on Functional Mapping of the Human Brain, Copenhagen, NeuroImage **5:**456.

Wildgruber, D., Erb, M., Klose, U., and Grodd, W. 1997. Sequential activation of supplementary motor area and primary motor cortex during self-paced finger movement in human evaluated by functional MRI. *Neurosci. Lett.* **227:**1–4.

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.

Wu, D., and Lewin, J. S. 1994. Evaluation of non-parametric statistical measurements and data clustering for functional MR data analysis. *SMR, Book of Abstracts,* 629.

Xiong, J., Gao, J.-H., Lancaster, J. L., and Fox, P. T. 1995a. Comparison of common statistics used for functional MRI data analysis, *SMR/ESMREMB, Book of Abstracts,* 827.

Xiong, J., Gao, J.-H., Lancaster, J. L., and Fox, P. T. 1995b. Clustered pixel analysis for functional MRI activation studies of the human brain. *Hum. Brain Map.* **3:**287–301.