An Eigenfunction Approach to the Inverse Problem of EEG

Bo Hjorth
Research and Development Laboratory, Siemens-Elema Ab, Solna, Sweden

Ernst Rodin
Department of Neurology, Wayne State University and Epilepsy Center of Michigan, Detroit, Michigan, USA

Key words: Source localization; EEG eigenfunction analysis; EEG eigenvector; Factor analysis; Methodology; Source derivation; Average reference.

Summary: A mathematical method of extracting salient features from electroencephalographic data called eigenfunction analysis is presented. It allows the reduction of 21 channels of EEG data to a few components which can be separated into those which are likely to originate relatively close to the surface and others of deeper origin. It was demonstrated that the original tracings can be reconstituted from these few components. The eigenvectors give an indication of the location of sources and the degree to which the eigenfunction appears on source derivation and average reference recordings allows an estimation of relative depth. The method has been successfully applied to EEG tracings from 10 patients and is illustrated in the case of a young woman suffering from complex partial seizures associated with a deep left temporal lesion. The implications for marked data reduction and the development of objective assessment of clinical neurophysiologic data are discussed.

Introduction

In a previous publication (Hjorth and Rodin 1988) we have explored the possibility of separating EEG segments containing generalized spike-wave discharges into deep and superficial components. This was accomplished by subtracting source density values from average referenced electrode potential values. The resultant data were then subjected to eigenfunction analysis. It is the purpose of this presentation to delineate the process of eigenfunction analysis in greater detail and show its applicability to localization of clinical electroencephalographic data.

Method

The underlying assumption of this study was that EEG tracings consist of correlated as well as uncorrelated elements. Correlation between tracings was considered to be the result of field components generated by common sources. Field components from several such sources superimpose to form the composite field from which the potentials are derived at defined electrode positions and conventionally recorded as tracings. By definition, two or more sources generating correlated elements were considered one generating system with respect to these elements. In a more formalized approach, the instantaneous potential value at electrode i may be denoted p(i) and the value simultaneously generated by system j denoted e(j). The superposition, or linear combination, described above can then be expressed by the algebraic transformation

$$p(1) = t(1,1)\,e(1) + t(1,2)\,e(2) + .. + t(1,j)\,e(j) + ..$$
$$p(2) = t(2,1)\,e(1) + t(2,2)\,e(2) + .. + t(2,j)\,e(j) + ..$$
$$\vdots$$
$$p(i) = t(i,1)\,e(1) + t(i,2)\,e(2) + .. + t(i,j)\,e(j) + ..$$
$$\vdots$$

All coefficients t(i,j) are assumed to remain constant during the epoch under study, while all values e(j) and all resultant potentials p(i) are assumed to vary with time. The variation of each value e(j) during the epoch is named an eigenfunction (also "proper function"), since it is uncorrelated to the other eigenfunctions, that is to the e(j)'s with other values of j (j = 1, 2, 3, etc.), and together with these can be used to synthesize all electrode potentials p(i) as described by the transformation above.

Each column of coefficients, for example t(i,1), describes how the corresponding eigenfunction, in this example e(1), is distributed over the electrodes (i). This means that the coefficients in some way reflect the location of the system which generates the corresponding eigenfunction. We have assumed the extent to which an eigenfunction contributes to an electrode potential, as expressed by the corresponding coefficient, to be an appropriate quantity for describing the relative vicinity of this electrode to the generating system. A location is conventionally defined by means of three geometrically orthogonal directions, but in the EEG recording situation the non-orthogonal directions of the electrode positions are at least as useful, since the coefficients representative

Accepted for publication: November 8, 1988
Correspondence and reprint requests should be addressed to Bo Hjorth, Skaanevaegen 26, S-19143 Sollentuna, Sweden.

of the electrodes match exactly the array of input data required by any mapping system for topographic display of EEG activity.

Mathematical Derivation

The transformation described above can be equivalently expressed by means of the more concentrated notations of matrix algebra as

$$p = [T]\, e, \qquad (1)$$

where the array of coefficients t(i,j) in the explicitely written transformation have been extracted to form the corresponding array of elements in matrix [T]. The set of electrode potentials p and the set of eigenfunctions e are formally named vectors.

Determining the locations of the systems which generate the eigenfunctions thus requires that matrix [T] is known. Deriving the eigenfunctions furthermore requires that its inverse, [invT], can be computed, in which case the eigenfunctions are then derived from the recorded potentials as

$$e = [invT]\, p. \qquad (2)$$

Our method of computing [T] and [invT] is related to factor analysis (Comrey, 1973) and, in analogy with factor analysis, based on the correlation matrix for available data. The correlation matrix [C] for the electrode potentials p is derived from the covariance matrix [Cov(p)] as

$$[C] = [1/s]\, [Cov(p)]\, [1/s], \qquad (3)$$

where [1/s] is a diagonal matrix containing the reciprocals of the standard deviations s of the electrode potentials p as its elements (all off-diagonal elements being zeroes). The physical equivalent of variance is mean power and the equivalent of standard deviation is RMS-potential (Root Mean Square). In view of later computations, it is advantageous to consider the normalized electrode potentials p', i.e., each potential divided by its RMS-value for the epoch:

$$p' = [1/s]\, p. \qquad (4)$$

Inversely, the potentials can then be expressed as

$$p = [s]\, p', \qquad (5)$$

where [s] is the inverse of [1/s]. The normalization of p makes it possible to express the covariance matrix for p in terms of p';

$$[Cov(p)] = [s]\, [Cov(p')]\, [s].$$

Using this alternative expression for [Cov(p)] in (3) shows that

$$[C] = [Cov(p')],$$

i.e., that [C] is identical to the covariance matrix for the normalized electrode potentials p'. This matrix can, according to the rules of matrix algebra, be expanded into a product of three matrices,

$$[V]\, [E]\, [invV] = [Cov(p')], \qquad (6)$$

where [V] contains "eigenvectors" as columns, [E] "eigenvalues" along its main diagonal (all off-diagonal values being zero), and [invV] again containing the eigenvectors, but in this matrix as rows. The latter means that each of [V] and [invV] is the transpose of the other and, since they are orthogonal matrices, they are also mutually inverse so that no matrix inversion is required. This is true also when one (or more) eigenfunction is zero, i.e., the number of eigenfunctions is less than the number of electrode potentials, although the redundant eigenvectors are then nonsensical. A method for eigenvalue and eigenvector analysis according to (6), originally credited to the mathematician Jacobi, has been described and implemented as a computer program in the FORTRAN language by Carnahan et al. (1969).

The covariance matrix for p' is identical to the epoch average of the vector product of p' and its transpose;

$$[Cov(p')] = average(\, (p') \times transpose(p') \,).$$

Combining this expression with (6) and recalling that [invV] is the transpose of [V] yields

$$[E] = average(\, ([invV]\, p') \times transpose([invV]\, p') \,)$$

which means that [E] is the covariance matrix of ([invV] p'). This assumes that the average of each component of p' is zero, which is another way of saying that the time constant used for the EEG recording must be short compared to the epoch length. The fact that [E] is a diagonal matrix also means that the elements of vector ([invV] p') are uncorrelated and, hence, represent the eigenfunctions;

$$e = [invV]\, p'.$$

By substituting for p' by means of (4), this result can be expressed in terms of recorded potentials p as

$$e = [invV]\, [1/s]\, p \qquad (7)$$

and, inversely,

$$p = [s]\, [V]\, e. \qquad (8)$$

Finally, by combining (7) with (2) and (8) with (1), [T] and [invT] can be identified as

$$[T] = [s]\,[V]$$

and

$$[invT] = [invV]\,[1/s],$$

by means of which the eigenfunctions can be derived (2) and the corresponding locations t(i,j) determined.

Implementation of the Method

Although the method may seem logically straightforward, there are two points worth noticing; limitation in the number of eigenfunctions and degeneration of the correlation matrix in case of low-amplitude electrode potentials.

Eigenvalue and eigenvector analysis, as described by expression (6), works only for square matrices, i.e., when the number of eigenfunctions having their variances represented by the eigenvalues in [E] is formally identical to the number of electrodes. In practice, however, the number of eigenfunctions having a magnitude (eigenvalue) of any significance is always much smaller than the number of electrodes. For practical reasons, the eigenvalues are sorted into descending order and the eigenvectors rearranged correspondingly in [V] and [invV]. Usually a few eigenvalues stand out significantly from the general "noise level" of remaining eigenvalues and represent the choice of eigenfunctions to be studied. The clinical significance of each eigenfunction must be estimated on the basis of its features as a function of time and its location.

In actual recordings, one or more tracings may display very low or even zero amplitude, in which case correlation matrix [C] will collapse due to small or zero value of s in (3). This can be avoided by (falsely) assigning unity to any s below a chosen minimum and, at the same time, assigning zeroes to all off- diagonal elements in the row and in the column of [C] corresponding to the same electrode, and unity to the common diagonal element. As a result, this electrode potential will not be affected by transformations (7) and (8), since the analysis algorithm will then consider it uncorrelated to the other potentials and, hence, an eigenfunction. The "escape" of this electrode potential from analysis will not significantly affect the transformation of remaining data, since the basis for its exclusion is too small power. This manipulation of the correlation matrix, making some of its diagonal elements falsely represent uncorrelated signals, prevents degeneration of the correlation matrix and guarantees successful eigenvalue and eigenvector analysis of remaining data. The cost for this is that transformation (7) transfers electrode potentials having low amplitude into formal eigenfunctions that are not, in reality, uncorrelated to the other eigenfunctions. This is of little importance, since eigenfunctions associated with small eigenvalues (i.e., low power) will be disregarded in the final evaluation anyway.

The EEG data were obtained on an 18-channel electroencephalograph connected to a commercial computer system with data acquisition and field mapping capabilities. Analog signals were filtered between 1 and 70 Hz. Two hundred and fifty-six data points were available for a 2000 ms sample per channel allowing for a resolution of 7.8 ms intervals per data point. The International System of Electrode Placements was used but the three frontopolar values were extrapolated by the computer program. The data were initially obtained on a linked ears reference but subsequently transformed via computer to the average common reference. The data were then separated into "deep" and "superficial" components as described previously by Hjorth and Rodin (1988). The concept of superficial data is based on a paper by Perrin et al. (1987), showing that data recorded by the source derivation technique are generated in more superficial regions of the brain and, as a consequence, we have assumed the data remaining after extraction of superficial data from the original recording to represent "deep data". This aspect will be further discussed later. All data processing was done by means of an IBM compatible microcomputer.

Results

Eigenvalue and eigenvector analysis was successfully applied to the individual correlation matrices computed from 10 EEGs, i.e., the Jacobi iteration (Carnahan et al. 1969) showed fast convergence in each case. Eigenfunctions were derived by means of transformation (7) described in the preceding and, in order to reconstitute the original data, subjected to transformation (8). Reconstitution showed that no information was lost in our computer implementation of the method. In order to determine how various features in the EEG were related to different eigenfunctions, we also limited the number of eigenfunctions used in the reconstitution. Usually 3 to 5 eigenfunctions were sufficient to obtain a reconstitution that preserved all essential features of the original tracings.

Figure 1 shows the example of an EEG to which the analysis was applied after dividing the data into "superficial data" and "deep data". Figures 2 through 4 show the corresponding results which, according to our experience, are typical of the method. The EEG was from a young woman with a deep left mesial temporal lesion which was either a low grade neoplasm or a congenital malformation. She had not yet been surgically explored, because her seizures were quite minor and she was otherwise asymptomatic. Figure 2 shows the superficial data, the first 10 eigenfunctions resulting from analysis and a reconstitution based only on the first 4 of these eigenfunctions. The reconstitution demonstrates that

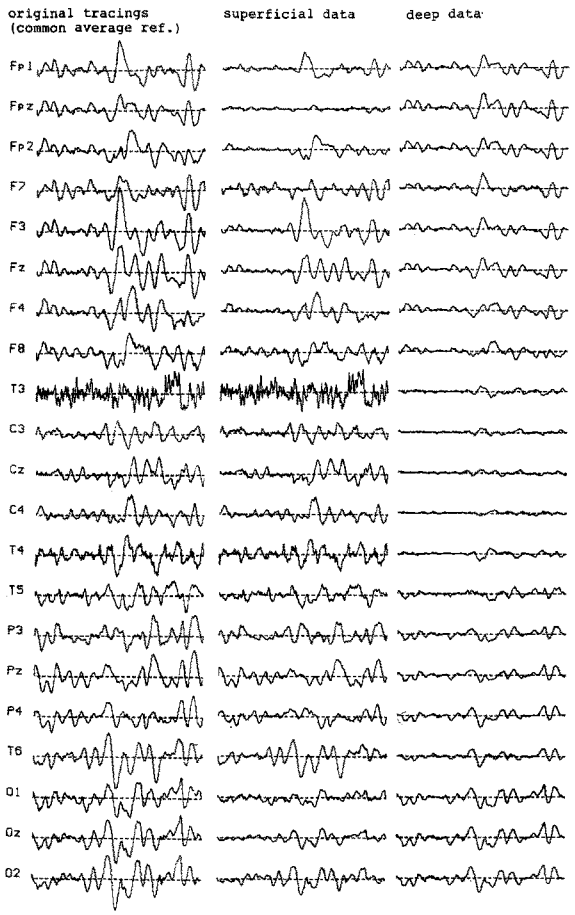

Figure 1: Source density data (center) derived from original EEG tracings (left) were considered to be representative of superficial sources, while data remaining after subtraction of superficial data from the original tracings were considered to be representative of deeper sources (right).

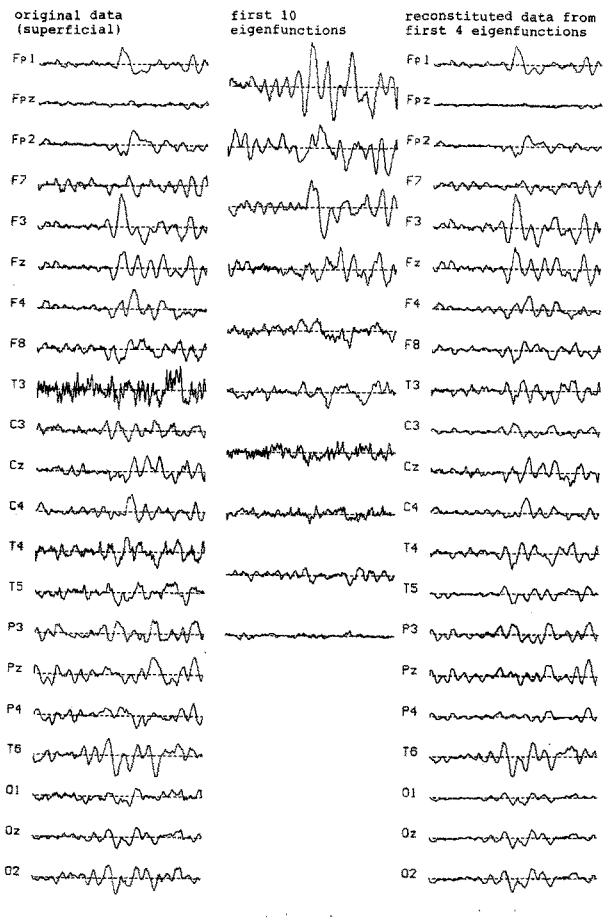

Figure 2: Analysis of the superficial data (left) yielded a number of eigenfunctions (center) which were sorted with respect to descending eigenvalue. Reconstitution (right) from the first 4 eigenfunctions showed that these contain the essential features of the original (superficial) data.

these 4 eigenfunctions account for practically all significant features visible in the original data. The fast frequencies at T3 representing temporalis muscle activity in the original tracings are not contained in the 4 selected eigenfunctions, and therefore do not appear in the reconstituted tracings.

Figure 3 shows the 4 superficial eigenfunctions used for reconstitution together with their location data, and also the (only) 2 eigenfunctions obtained from the deep data. The location data are derived as described in the Methods section, but multiplied by 100 in order to express the percentage to which the eigenfunction is present at each electrode. Negative location data indicate where the polarity is reversed in relation to the displayed eigenfunction or, differently expressed, where the eigenfunction is reversed. It should be noticed, however, that the presence defined by the location data cannot be referenced to any point external to the electrodes involved, since the elements of an eigenvector on which the location data are based, are always referenced to their own average. In spite of this relativity of the location data, vicinity to the origin of an eigenfunction is usually detectable as a peak in the spatial distribution of the location data, i.e., as a high value of either polarity at one position surrounded by lower values, while the values of opposing polarity are more evenly distributed in remaining areas. An example of such a location is F3 with respect to the first superficial eigenfunction, as can be seen in Figure 3.

The two first superficial eigenfunctions are both represented (superimposed) in the first deep eigenfunction. This may not be easily seen by visual inspection, but can be determined by computing the covariances. The first eigenfunction has equal amplitude in superficial and deep data and may, therefore, be estimated to originate at a depth of approximately 0.65 radius of the head, i.e., centered between 0.5 and 0.8 radius, according to our interpretation (Hjorth and Rodin 1988) of the results published by Perrin et al. (1987). The second superficial eigenfunction has a larger amplitude in its deep representation, thus indicating a depth closer to, but more than 0.5 radius, possibly 0.55 -0.6 radius. The third

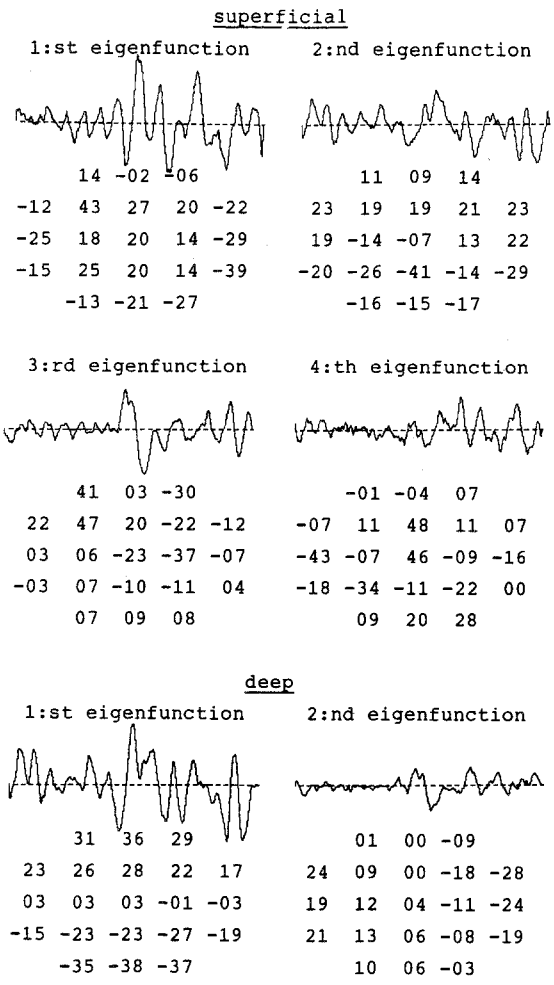

Figure 3: The first 4 superficial and the only 2 deep eigenfunctions shown together with their respective location data. These are expressed as the percentage to which the corresponding eigenfunction is present at each electrode (First line: F_{p1} F_{pz} F_{p2}, next line: F7 F3 Fz F4 F8, etc.).

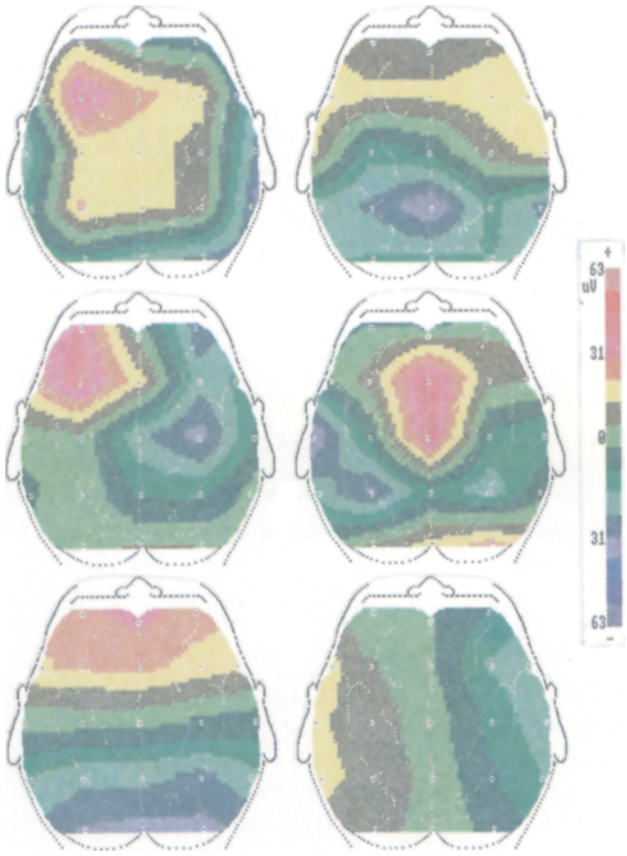

Figure 4: Topographic display of the four superficial and two deep eigenfunctions. Top row first and second superficial eigenfunction, middle row 3rd and 4th superficial, and bottom row 1st and 2nd deep eigenfunction. Positivity on this map does not mean positive electrode potentials at this location, but merely expresses the presence of the eigenfunction which consists of a negative positive sequence. The eigenfunction is also present in negative areas of the map, but with reversed polarity.

superficial eigenfunction has some of its elements represented in the second deep eigenfunction, but with a much reduced amplitude. This indicates that the origin is superficial, i.e., close to 0.8 radius. The fourth eigenfunction is purely superficial.

Figure 4 shows color maps of the numerical location data in Figure 3 of the first two superficial eigenfunctions in the top, the 3rd and 4th superficial eigenfunctions in the middle and the two deep ones in the bottom row. The presence of the "upright" eigenfunction is indicated by positivity and its reverse by negativity. The first and third eigenfunctions can be assigned a distinct common location at F3. The second eigenfunction, which is the deepest, shows superficially a lateral symmetry with maxima at F7 and F8, while the opposing polarity has a maximum at Pz. Since the reversed eigenfunction at Pz has the largest amplitude one may suspect that the origin is closer to this position. The fourth, purely superficial, eigenfunction displays a complicated location pattern which is not readily interpretable at this time.

The location data of a deep eigenfunction may be seen as representing the "far field", thus describing the main direction of a source as projected onto the global surface. The location data related to the first deep eigenfunction thus indicate the common direction of the global fields generated by the sources of the two first superficial eigenfunctions, since these fields superimpose to form one common eigenfunction in the deep data. The location data of the second deep eigenfunction indicate a global orientation perpendicular to that of the first eigenfunction. The appearance of the lesion on the MRI scan is shown in Figure 5.

Discussion

Although eigenfunctions may be very efficient descriptors for reducing the amount of data without loss

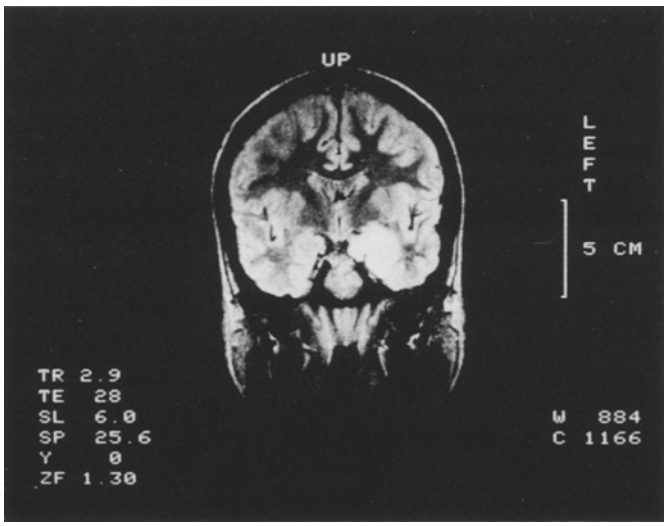

Figure 5: Appearance of the left mesial temporal lesion on MRI scan.

of information, there is no guarantee that they reveal any "deeper truth" concerning the generation of these data. If, however, the analyzed data are known to be generated by a physical system, the eigenfunctions are often, in analogy with factors derived by means of factor analysis, found to reflect fundamental physical parameters of this system. This means that the actual "truth" is often also the most efficient descriptive model for the data. With EEG potentials as input data, the physical model is relatively well defined; EEG data represent the potentials at defined positions in space, potentials are bound to have sources, and potential fields generated by sources superimpose to form the recorded potentials.

The possible interpretive errors resulting from eigenfunction analysis of EEG are related rather to the selection of time segments of input data than to the lack of a more precise interpretive model. The EEG used here as an example was selected because the origin of abnormal activity was stationary in space, ensuring this activity a well defined representation in the correlation matrix computed for the entire length of the EEG epoch. In certain types of epilepsy, with different wave elements being generated in different places, the epoch may have to be subdivided into short segments. Correlation based on short segments may, on the other hand, lead to incorrect results, since single elements may occasionally be similar although their origins are unrelated. Fortunately, the location data indicate this kind of situation, and may be used to "focus" on the origin of a certain activity by "tuning" the segment limits to produce a plausible location map.

In this application of eigenfunction analysis the eigenvalues are used primarily for sorting the eigenfunctions in an order corresponding to descending magnitude. While the basis for estimating the significance of a factor obtained from factor analysis is its eigenvalue, the significance of an eigenfunction may be determined by for example, features indicating normal or abnormal activity. An epileptic spike does not contribute much to the eigenvalue (mean power) over an epoch, but may be highly significant in the clinical evaluation. The 3rd eigenfunction in Figures 2 and 3 shows a distinct feature; the sudden onset of a process looking like a physical system response. This eigenfunction takes the rank of the 2nd eigenfunction in an analysis based on the latter half of the epoch, since its eigenvalue will then exceed that of the current 2nd function, but all features remain the same. A conclusion of this is that the eigenvalue provides little information compared to a visual inspection of the corresponding eigenfunction. One should, on the other hand, attach significance only to the most dominant eigenfunctions, since the others may represent "mathematical orthogonalization of residual noise". The 4th eigenfunction in Figure 3 may possibly be an example of the latter, since its location data present a too complicated pattern, unless there is some instationarity requiring the last quarter of the epoch to be analyzed as a separate segment.

It may also seem surprising that the muscle activity at T3 is not assigned an eigenfunction of its own, but this is related to the fact that this extremely local (superficial) activity of moderate amplitude has too small total field power to compete with the dominant components. The risk of obtaining non-unique solutions from the eigenvalue and eigenvector analysis increases as the power approaches that of the "eigenvalue noise level" (especially considering the manipulation of the correlation matrix described in the methods section). This, however, suggests a different approach to the identification of superficial components having low amplitude; by making a reconstitution from all eigenfunctions except those of larger amplitude, so that remaining superficial components can be studied without the masking effect of those already recognized. The method of "filtering out" components recognized as eigenfunctions with defined locations, by means of a reconstitution excluding these components, may be used also for suppressing eye movement artifacts, for example.

The set of location data for a certain eigenfunction, i.e., t(i,j) for a given j, acts as a "focussing device", by extracting potentials being generated in the particular region, and having the particular direction, of this eigenfunction. We have seen an example in which a sharp spike was generated in the same region as a more powerful eigenfunction (i.e., as part of this eigenfunction) without being readily visible in the original recording, due to the superposition of other components. The polarity of the location data may seem confusing to some clinical electroencephalographers, who are used to mapping the polarity of individual events in the recording, since the polarity of the location data merely tells whether the entire eigenfunction is "upright" or reversed at each par-

ticular electrode. The polarity of the eigenfunction at any selected instant of time at a certain electrode is the same as that of the instantaneous value of the eigenfunction in areas where the location data are positive, and reversed in areas where the location data are negative. As an example, the sharp negative transition (upward, according to EEG convention) in the 3rd eigenfunction in Figure 3 has its maximum (47% of the negative value of the eigenfunction) at F3 and the reversed maximum (37% of the same value with reversed polarity, i.e., transition positive) at C4. The same principle applies also when the instantaneous values of the eigenfunction are positive, e.g., at the positive peak some 150 ms later, when the contributions to the electrode potentials from the same eigenfunction will be 47% of this value at F3 and 37% of the same value, but negative, at C4. This is how the reconstituted data in Figure 2 are derived, not only for the negative and positive peaks but for each value during the epoch, with the contributions from the 4 first eigenfunctions superimposed at each electrode, thus illustrating a possible hypothesis for generation of the originally recorded EEG.

A disadvantage of the location data is that the abosolute location in the superficial data is lost, due to the fact that the elements of each eigenvector refer to their average. One way to restore the absolute location would be to map the covariance between the eigenfunction and each superficial tracing, something we have not yet done. The location of an eigenfunction can always be mapped with reference to an arbitrary electrode, as long as the electrode is one from which the analyzed tracings are derived. If, for simplicity of example, we assume that the location data for T3 and T4 represent the ears instead, the first superficial eigenfunction can be mapped as referenced to "linked ears" after subtraction of -27, the mean of -25 and -29, from all location data. The presence of this eigenfunction with reference to "linked ears" would thus be 43% - (-27%) = 70% at F3, 0% at 02 and -12% (reversed) at T6, for example. Since experience has proved that "linked ears reference" contributes additional information as compared to common average reference, this suggests that the processing format should be extended from 21 to 23 average referenced channels (unless Fpz and Oz are excluded), with the ear electrodes sensing the spatial field in parity with other electrodes. From a technical point of view, the idea of a common physical reference is an anomaly, since different sources with different locations and different orientations cannot have a common reference, neither near nor far. Since the necessary reference must then be purely arithmetic, the only natural choice is the common average. This is also supported by the observation that a number of (matrix) operations are possible only with common average as the reference for input and/or output data (the inverse to source density computation, eigenvector computation, adaptive derivation, see Hjorth 1982). After processing average referenced input data, the results can always be displayed with reference to any of the electrodes involved, or to any combination of these, as illustrated by the example above.

Eigenfunction analysis of the EEG in Figure 1 has revealed two dominant activities centered at F3; the lst and 3rd eigenfunctions derived from the superficial data. The partition of the original, average reference, tracings into "superficial" and "deep" data indicates that these origins are located at different depths. The nominal values of depth level, 0.65 and 0.8 radius for the two eigenfunctions, indicate the order in which origins are situated relative to each other, but should not be taken as true absolute values. One reason is that the idealized conductivity model, on which the idea of "superficial" and "deep" data is based, may not be sufficiently close to the physical reality of the individual case. Another reason is that a linear interpolation between 0.8 radius and 0.5 radius is not valid, since the decline of the source density (superficial data) from 100% at 0.8 radius to 0% at 0.5 radius is not linear as a function of source depth. An estimation of the proportion between the sizes of a certain superficial eigenfunction and its corresponding contribution (by superposition) to some of the deep eigenfunctions may seem somewhat arbitrary, but can easily be made exact by further development of the computer program.

From a clinical point of view, the methodology presented here offers considerable promise because it represents another step toward more precise as well as objective data analysis and interpretation. The fact that it is possible to extract the major components of a sample of EEG and reconstitute the original tracings without loss of informational content is important and so is the observation that within certain limits an estimate of depth of the various processes can be made on purely mathematical grounds. In the cases examined so far, the first eigenfunction corresponded in all instances to features of the record which a well trained clinical electroencephalographer would recognize in the original trace, but orientation of the vector and relative depth may be difficult if not impossible to determine. It is therefore quite likely that with further refinement of the method especially in regard to depth estimate, the computer-assisted clinical electroencephalographer can look forward to another powerful tool for accurately assessing the location of sources contained in electroencephalographic data.

References

Carnahan, B., Luther, H.A. and Wilkes, J.O. Applied Numerical Methods. John Wiley & Sons, Inc. 1969, SBN 471 13507 0: 250-260.

Comrey, A.L. A First Course in Factor Analysis, Academic Press, New York and London, 1973.

Hjorth, B. An adaptive EEG derivation technique. Electroenceph. clin. Neurophysiol., 1982, 54: 654-661.

Hjorth, B. and Rodin, E. Extraction of "deep" components from scalp EEG. Brain Topography, 1988, 1(1): 65-69.

Perrin, F., Bertrand, O. and Pernier, J. Scalp current density mapping: Value and estimation from potential data. IEEETransactions on Biomedical Engineering. BME-34, 1987, 4: 283-288.