ON TIME DELAY ESTIMATION AND ITS APPLICATION TO THE STUDY OF COMMUNICATION BETWEEN DIFFERENT CEREBRAL AREAS

C. Fortgens

Department of Neuropsychiatry, Psychiatric Centre Rosenburg, The Netherlands

Summary

Five methods for time delay estimation are discussed: the Basic Cross Correlation method, the Generalized Cross Correlation method, the Average Amount of Mutual Information, the Maximum Entropy of the Difference of 2 time series and the Directed Transinformation. In case of *a priori* knowledge, expressions for the maximum likelihood time delay estimators are derived. The application of these different methods in physiological applications will be discussed. Simulations are provided to illustrate some of the characteristics of the proposed methods.

Introduction

Brain research has revealed that different cerebral areas work in concert to produce a 'single' mental activity. Cerebral systems have been identified that support memory functions, while other systems mediate language or consciousness, etc. However, usually we do not know how the different subsystems cooperate. A relative simple parameter in the interaction process between neural areas might prove to be the time delay involved in the communication between these subsystems. Knowledge about the time relationships may give some insight in the hierarchical organization of major brain processes.

In our clinic, we are particularly interested in the communication process between both cerebral hemispheres. It is assumed that this cerebral communication is somehow reflected in the electrical brain signals (EEG).

Firstly, current models and algorithms will be discussed and secondly, a new application of an information theoretic measure is proposed which can be used in both stationary processes as well as in transient-like processes, such as the event related brain potentials.

Cross Correlation Methods

In many studies on time delay estimation, the model used can be presented like in Fig. 1, where $s(i)$ and $s(i-D)$, corrupted with independent noise are known:

$$x(i)=s(i)+n_1(i)$$
$$y(i)=s(i-D)+n_2(i) \qquad \text{F1}$$

s, n_1 and n_2 are independent, while n_1 and n_2 are Gaussian processes. Perhaps the most commonly used method is the Basic Cross Correlation method (BCC),

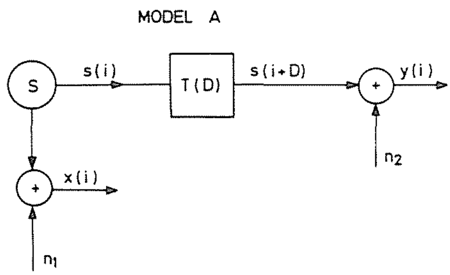

Fig. 1. Model A is a simple delay system, s is the signal source and $T(D)$ is a linear transform that introduces a delay of D samples. n_1 and n_2 are zero-mean normal processes

or its frequency equivalent, the Coherence Phase Shift method. These methods proceed as follows. The estimator $\hat{D}$ is that value of D that maximizes the expression:

$$\mathrm{BCC}(D)=\frac{1}{N-D}\sum_{i=D+1}^{N} x_m(i)\cdot y_m(i-D)\,, \qquad \text{F2}$$

where x_m and y_m are the actual (measured) data.

One may notice that the BCC only depends on knowledge of the cross variance $R_{xy}(t)$ and therefore is an a posteriori estimator.

An improvement on the BCC can be made by filtering the signal before the cross correlator, like shown in Fig. 2.

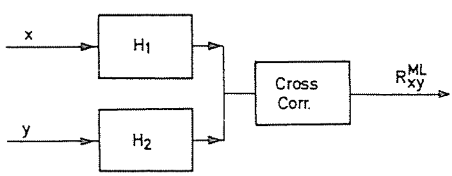

Fig. 2. Generalized Cross Correlation method

This is the so-called Generalized Cross Correlation method (GCC). Knapp and Carter [3] derived expressions for H1 and H2 that yield a 'heuristic' Maximum Likelihood time delay estimator. These expressions are:

$$R_{xy}^{\mathrm{ML}}(\tau)=\sum_{f}\frac{C_{xy}(f)}{(1-C_{xy}(f)}\cdot\frac{G_{xy}(f)}{|G_{xy}(f)|}\cdot e^{j2\pi f\tau} \qquad \text{F3}$$

where:

$$C_{xy}(f)=\frac{(G_{xy}(f))^2}{G_{xx}(f)\cdot G_{yy}(f)} \qquad \text{F4}$$

is the Magnitude Squared Coherence (MSC) and G is the Fourier transform of the correlation function R.

Actually, Knapp and Carter [3] did not derive a ML-time-delay estimator, because they used estimated spectra as their *a priori* knowledge, estimated from the same data set as in which the time delay had to be estimated. This was also recognized by Scarbrough et al. [7]. Their "optimal" ML processor used the *a priori* auto-spectra, and the estimated cross-spectrum. Again, also this "optimal" ML-processor is not a ML-estimator. In fact, the algorithm is not stable under all circumstances. For instance, if the estimated cross spectrum is equal to the square root of the product of the auto-spectra then the Magnitude Squared Coherence is equal to 1 and therefore F3 is not defined.

In appendix A expressions are derived for the ML-time-delay estimator in case of *a priori* known noise processes, and in case of a priori known noise processes plus signal process. For the second case which is also the one considered by Scarbrough et al. [7], the ML-estimator uses all 2nd order moments, and not as they suggested, only the cross covariances. In appendix B it is shown, that BCC is a ML-estimator if the noise processes are known to be zero-mean, independent white Gaussian processes.

The Average Amount of Mutual Information (AAMI) and Maximum Entropy of the Difference (MED)

A method using an information theoretic measure in time delay estimation has been described by Mars and Van Arragon [5]. The nonlinear model they used is shown in Fig. 3.

Mars and Van Arragon used the Average Amount of Mutual Information (AAMI) to estimate the delay D from the measurements x and y. The properties of the mutual amount of information have been studied extensively by Gel'fand and Yaglom [1]. Mars and Van Arragon took as $\hat{D}$ that value of D which maximizes the expression:

$$\mathrm{AAMI}_{XY}(x, y|D) = \iint_{XY} \hat{p}_{XY}(x, y|D) \log \frac{\hat{p}_{XY}(x, y|D)}{\hat{p}_X(x)\, \hat{p}_Y(y)}\, dxdy\,. \qquad \text{F5}$$

Because of the nonlinearity, normality of x and y is not assumed. The power density functions were estimated using a sophisticated kernel estimator. Only stationarity is assumed for signal and noise sources. So the AAMI-estimator is

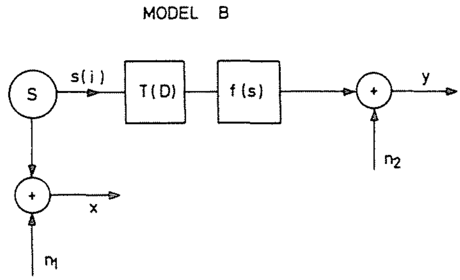

Fig. 3. Model B is equal to Model A, except for the non-linearity $f(s)$ in the delay line

an *a posteriori* estimator. The division by $\hat{p}_X(x)$ and $\hat{p}_Y(y)$ in F5 is merely a matter of normalization. In appendix C it will be derived that in case if x and y are normal processes, the AAMI method proves to be equal to the BCC.

Now we have seen 2 *a posteriori* estimators, BCC and AAMI, that actually depend on 2 variables: xy or $\hat{p}(x, y)$. However, as we might guess from appendix A, an improvement can possibly be made by using also xx and yy or $\hat{p}(x, y^D)$, where x and y^D are vectors of random variables presenting the process and D denotes a shifted version of the process. A suitable measure is the Maximum Entropy of the Difference (MED) between series x and series y. If $z^D = x - y^D$ then:

$$\mathrm{MED}(D) = -\int_{-\infty}^{\infty} \hat{p}(z^D) \log \hat{p}(z^D)\,. \qquad \text{F6}$$

If $n1$ and $n2$ are independent Gaussian processes then:

$$\hat{p}(z^D) = \frac{\exp\left[-\frac{1}{2}(z^D)^T \hat{R}_D^{-1} (z^D)\right]}{(2\Pi)^{n/2} |\hat{R}_D|^{\frac{1}{2}}} \qquad \text{F7}$$

where n is the dimension of z^D and $\hat{R}_D = R_{x-y^D}$.

Substituting F7 in F6:

$$\mathrm{MED}(D) \approx -\log(|\hat{R}_D|)\,. \qquad \text{F8}$$

Now take as $\hat{D}$ that value of D, that maximizes (det $(\hat{R}_D))^{-1}$. Rewriting det $(\hat{R}_D)$:

$$|\hat{R}_D| = |\hat{R}_{xx} + \hat{R}_{yy} - \hat{R}_{xy^D} - \hat{R}_{y^D x}|\,. \qquad \text{F9}$$

This expression shows that all elements of the estimated covariance matrices are being used. It should be noted that we did not require s to be a Gaussian process. Furthermore, MED(D) averages over all possible realisations of the processes x and y. If a priori information is provided, a p. d. f.-estimator must be invoked that takes this knowledge into account [8]. If the delayed signal is somehow attenuated, then z^D must be changed into $z^{D*} = A \cdot x - y^D$, while the maximization problem has become one with 2 variables D and A.

The physiological model

Basically we have introduced 2 models, firstly a delay system without signal distortion, secondly, a delay system with signal distortion. In case of the first type of model, ML estimators have been derived, though they are usually only of theoretical interest (because usually we do not have relevant a priori information).

In the sequel, I will disregard the frequency domain approaches. Although they may offer calculational advantages, they are hampered by the fact that delays are always restricted to +180 and −180 degrees phase shifts and they offer no conceptual advantages in the case of time delay estimation.

Now, before considering the presented estimators in more details, let us examine the model which we want to use in solving the delay estimation problem in neural tissues. I think that we do not have sufficient physiological support for either model A or model B. Reality deviates from these models in a number of aspects.

1. Different nerves conduct the electrical nerve pulses at different velocities. Because of these different velocities, the signal will be distorted at micro level and therefore perhaps also at the macro level. Here, micro level refers to measurements in single cells and nerves, where macro level refers to measurement of relative large populations of cells.
2. Often communication channels between different brain areas are bidirectional. Feedback loops are always possible.
3. Gaussian distributions are usually postulated but rarely justified. Deviations from normality in the measured process should be taken into account. The

algorithms that are based on models that implicate normality, must be 'sufficiently' insensitive to the actual deviations from these assumptions.

4. It is possible that coding and decoding devices are part of the communication system. Little is known about the characteristics of these devices. Nonlinearities, as well as memory are likely to be present.
5. In many physiological recordings, the instrument noise is small compared to unwantedly recorded physiological background signals. There is no ground for the assumption that these background signals, recorded at different electrode locations, are independent, neither will they be independent of the signal we are actually concerned with.
6. During the generation of, for instance evoked potentials, special neural pathways may be involved. Therefore, the channel properties may depend on the delay of time since the stimulus onset.

It will be clear from this summary of necessary model properties that both models that have been discussed might give rise to poor results. Therefore, we present a model for which information theoretic measures have been proposed by Marko (1973). We believe that this model of bidirectional communication is well suited for the problem measuring exchange processes between cerebral areas.

Directed transinformation

The bidirectional communication model presented by Marko (1973), is shown in Fig. 4.

Marko derived expressions for the Directed Trans-Information (DTI) from A1 to A2 and vice versa. The DTI measure designates the reduction of the entropy in for instance x because of knowledge of past events in y. The sum of these measures in both directions is equal to the mutual information. Harashima has worked out the ideas of Marko interpreting x and y as time

MODEL C

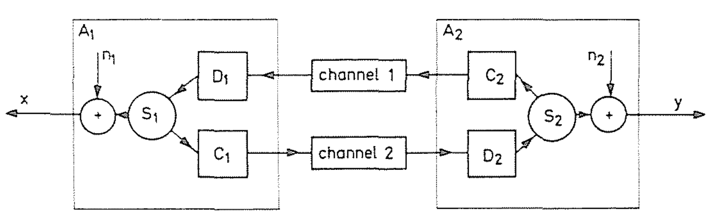

Fig. 4. Bidirectional communication between A_1 and A_2. S_1 and S_2 are signal sources; n_1 and n_2 are noise processes, independent of S_1 and S_2. C_1, C_2, D_1 and D_2 are coding resp. decoding devices

series, resulting in an expression for conditional mutual information (6, 9). This conditional mutual information expresses the statistical influence of variable x at time k on variable y at time $k+m$. In case x and y are multivariate random vectors, this statistical influence can be written as (6):

$$\mathrm{DTI}(x_k \rightarrow y_{k+m} | X^N Y^N y_k) = \frac{1}{2} \log \frac{|R(X^N Y^N y_k y_{k+m})| \, |R(X^N Y^N x_k y_k)|}{|R(X^N Y^N y_k)| \, |R(X^N Y^N x_k y_k y_{k+m})|} \quad \text{F10}$$

where

$$x = (x_0, \ldots, x_k, \ldots, x_{k+m}, \ldots) = X^N x_k, \ldots, x_{k+m}, \ldots$$
$$y = (y_0, \ldots, y_k, \ldots, y_{k+m}, \ldots) = Y^N y_k, \ldots, y_{k+m}, \ldots$$

and $R(.)$ is the covariance matrix.

This measure of uncertainty decrease of $yk+m$ because of xk has been applied in biological signals (Inouye et al., 1981, Saito, 1981). For calculational conveniency (numerical stability?), this DTI measure was expressed in ARMA-model coefficients. In this study we like to stay with F10, because this formule offers some practical and conceptual advantages. First of all, the 'heuristic' identification of the order of the ARMA-model is avoided. Secondly, the covariance matrices can also be estimated for evoked potentials. In this case x and y are vectors which elements have a fixed relationship to the stimulus onset. The covariance matrix R_{xy} can be estimated from N realizations $\mathbf{x}^j$ and $\mathbf{y}^j$ $(j=1, N)$:

$$\mathbf{x}^j = \{x_1^j, \ldots, x_s^j, \ldots, x_n^j\}$$
$$\mathbf{y}^j = \{y_1^j, \ldots, y_s^j, \ldots, y_n^j\} \quad \text{F11}$$

Sample s corresponds to the stimulus onset.

$$\hat{R}_{\mathbf{xx}}(i,j) = \frac{1}{N}\left(\sum_{k=1}^{N} x_{mi}^k \, x_{mj}^k\right)$$
$$\hat{R}_{\mathbf{yy}}(i,j) = \frac{1}{N}\left(\sum_{k=1}^{N} y_{mi}^k y_{mj}^k\right) \quad \text{F12}$$
$$\hat{R}_{\mathbf{xy}}(i,j) = \frac{1}{N}\left(\sum_{k=1}^{N} x_{mi}^k y_{mj}^k\right)$$

How well does the Marko-model reflects the biological reality? The model does not assume the presence of a single delay, multiple delays are included, though strictly, the model does not assume anything on delays. Because of possible coding and decoding devices, memory and nonstationary transfer functions, delays and impulse responses may not at all describe reality sufficiently. Though DTI measures the influence of process x at time k on process y at time $k+m$, m is not necessarily a physical delay. The mutual information measures do not indicate the anatomy of the substrate. However, because of additional knowledge about connections between certain cerebral areas etc., we are allowed to see m as a delay factor. We should like to stress that this interpretation is by no means 'logical'.

Simulations

Simulations are done to illustrate the time delay estimation methods that have been discussed. ML-estimators are not included because a priori knowledge is usually not available in physiological experiments. The BCC-algorithm, the TDI-algorithm and the MED-algorithm were fed with data generated by the model that is presented in Fig. 5. Moving Average filters with white Gaussian input are used to generate the time series. Results were

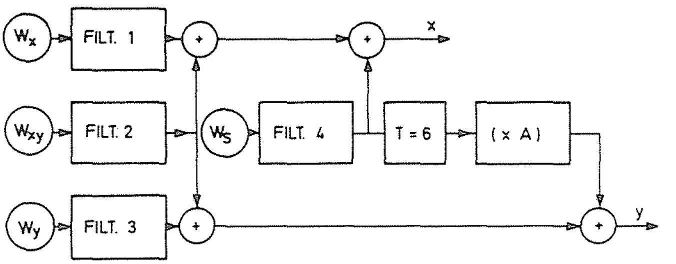

Fig. 5. Simulation model for generating delayed time series corrupted with (in)dependent noise

obtained using different signal to noise ratios. Firstly, FILTER 2 in the model was set to zero and the attenuation factor A was set to 1.0.
The actual model-parameters were:

0 db: FILTER 1 = (0.8, 0.6)
FILTER 2 = (0.0)
FILTER 3 = (0.8, 0.6)
FILTER 4 = (0.7, 0.5, 0.4, 0.3)
A = 1.0

−6 dB: FILTER 1 = (1.6, 1.2)
FILTER 2 = (0.0)
FILTER 3 = (1.6, 1.2)
FILTER 4 = (0.7, 0.5, 0.4, 0.3)
A = 1.0

Figure 6 shows the resulting curves of the different algorithms. One sees that BCC depends on the correlation function of s (broad peaks). DTI shows sharper peaks and MED yields very sharp peaks. MED does not depend on the correlation function of s, while the entropy drops very rapidly if delay D is not equal to the real delay.

In the second set of simulations, the situation is less ideal. FILTER 2 is not equal to 0.0 and $A = 0.5$. In this case we do not have uncorrelated noise while the delayed signal is attenuated. Algorithm MED has been adjusted to estimate the attenuation factor from the data. It actually maximizes the entropy of $\hat{R}_{\hat{A}x-y^D}$ (though this adjustment had little influence on the performance of the estimator).
The actual sets of model-parameters were:

0 dB: FILTER 1 = (0.57, 0.42)
FILTER 2 = (0.57, 0.42)
FILTER 3 = (0.57, 0.42)
FILTER 4 = (0.7, 0.5, 0.4, 0.3)
$A = 0.5$

−6 dB: FILTER 1 = (1.6, 1.2)
FILTER 2 = (0.0)
FILTER 3 = (1.6, 1.2)
FILTER 4 = (0.7, 0.5, 0.4, 0.3)
$A = 0.5$

Figure 7 shows the results of the different algorithms. Especially BBC and MED are sensitive to the correlated noise environment. DTI is not much affected by the presence of this noise.

Conclusion

The best time delay estimators are only best for the model to which they apply. If reality deviates from the chosen model, then the performance of the time delay estimator can deteriorate very rapidly. We have seen that especially uncorrelated noise in a case where correlated noise was expected, can be very

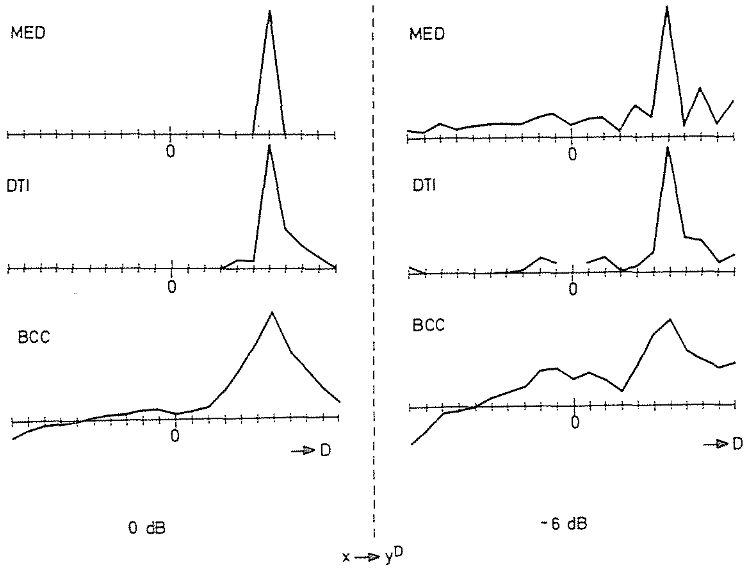

Fig. 6. Simulation results of MED, DTI and BCC. The noise processes are uncorrelated. The attenuation factor is $A = 1.0$

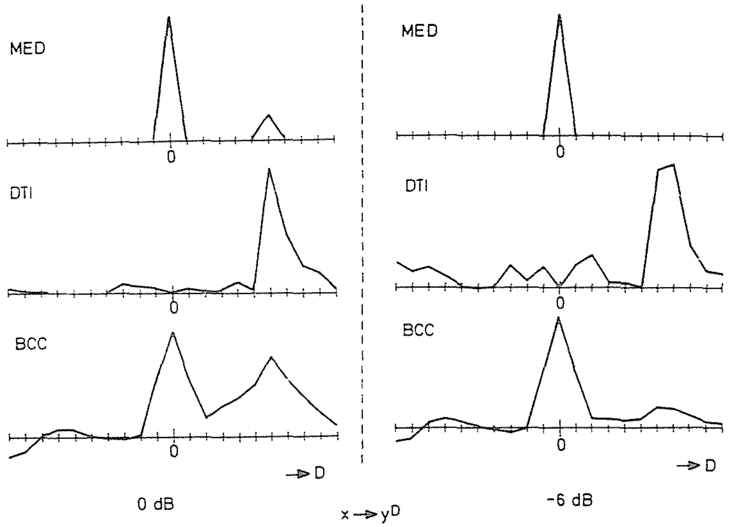

Fig. 7. Simulation results of MED, DTI and BCC. The noise processes are correlated and the attenuation factor $A = 0.5$

hazardous. In the case of a priori known spectra, ML-estimators are probably the best. But one should have very good reasons to believe that the a priori spectra indeed apply to his particular data set. If one does not have any prior information except the knowledge that noise processes are uncorrelated, MED seems to be a good choice. BCC performs sligthly worse than MED but consumes less computer time. In the case one does not know if the noise processes are uncorrelated, DTI offers a good solution. DTI does not seem to depend much on this type of noise.

The value of the DTI method for physiological experiments remains to be established.

Appendix A

In this appendix, 2 maximum likelihood estimators for time delay D are derived, firstly in case of *a priori* known noise processes n_1 and n_2, and secondly, in case of *a priori* known noise processes, as well as the signal process. Using the maximum likelihood philosophy, one takes as estimator $\hat{D}$ that value which maximizes the probability of measurements $\mathbf{x}$ and $\mathbf{y}$ that actually occurred, taking into account all *a priori* known statistical properties of noise processes and signal processes.

First model: $x(i)=s(i)+n_1(i)$

$$y(i)=s(i-D)+n_2(i)$$

$$\mathbf{x}=\{x_1, \ldots, x_n\}$$
$$\mathbf{y}=\{y_1, \ldots, y_n\}$$

s, n_1 and n_2 are independent processes. n_1 and n_2 are zero-mean Gaussian random processes with covariance matrices $R_{n_1 n_1}$ resp. $R_{n_2 n_2}$.

To find the ML-estimator $\hat{D}$ for time delay D, we need to maximize the probability density function for given data, with respect to D.

$$p(\mathbf{x}, \mathbf{y} \mid R_{n_1 n_1}, R_{n_2 n_2}, D)=\frac{\exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{y}^D)^T R^{-1}(\mathbf{x}-\mathbf{y}^D)\right]}{(2\pi)^n \, |R|^{\frac{1}{2}}}, \qquad \text{A1}$$

where $\mathbf{y}^D$ is a shifted version of $\mathbf{y}$ and $R=R_{n_1 n_1}+R_{n_2 n_2}$. Of course, in case of $R=\emptyset$, no ML-estimator exists.

Maximizing A1 is the same as minimizing $ML1$ (D) where:

$$ML1\,(\mathrm{D})=(\mathbf{x}-\mathbf{y}^D)^T R^{-1}(\mathbf{x}-\mathbf{y}^D) \qquad \text{A2}$$

$ML1$ (D) is the weighted summation of all elements of the estimated covariance matrix R_{xy^D}, where the weight factors are functions of the *a priori* known elements of $R_{n_1n_1}$ and $R_{n_2n_2}$.

Second model:

The second model is equal to the first model, except that s is known to be Gaussian and R_{ss} is given. This is the model Knapp and Carter (1976) and Scarbrough et al. (1981) considered. We can proceed as we did with the first model:

$$p(\mathbf{x}, \mathbf{y} \mid R_{ss}, R_{n_1n_1}, R_{n_2n_2}, D) = \frac{\exp\left[-\frac{1}{2}(\mathbf{x}, \mathbf{y}^D)^T R^{-1} (\mathbf{x}, \mathbf{y}^D)\right]}{(2\Pi)^n \, |R|^{\frac{1}{2}}} \qquad \text{A3}$$

where:

$$R = \begin{pmatrix} R_{ss} + R_{n_1n_1} & R_{ss} \\ R_{ss} & R_{ss} + R_{n_2n_2} \end{pmatrix}$$

Again, the ML estimator does not exist when R is singular. Maximizing A3 is the same as minimizing $ML2\,(D)$ where:

$$ML2\,(D) = (\mathbf{x}, \mathbf{y}^D)^T R^{-1} (\mathbf{x}, \mathbf{y}^D)$$

$\hat{D}$ is the value of D, that minimizes $ML2\,(D)$. $ML2\,(D)$ is a weighted sum of all elements of the estimated covariance matrices R_{xx}, R_{xy^D} and R_{yy}, where the weight factors are functions of the *a priori* known elements of $R_{n_1n_1}$, $R_{n_2n_2}$ and R_{ss}.

Appendix B

If *a priori* is known, that n_1 and n_2 are white zero-mean Gaussian processes, the Basic Cross Correlation method and the ML-estimator yield the same result. The maximum likelihood estimator can be expressed as (see appendix A):

$$ML1\,(D) = (\mathbf{x} - \mathbf{y}^D)^T R^{-1} (\mathbf{x} - \mathbf{y}^D) \qquad \text{B1}$$

Because n_1 and n_2 are white Gaussian processes we can write:

$$R^{-1} = \frac{1}{R_{n_1n_1}(0,0) + R_{n_2n_2}(0,0)} \cdot E \qquad \text{B2}$$

where E is the unity matrix.

Substituting B2 in B1:

$$ML1\,(D) = \mathbf{xx} + \mathbf{y}^D\mathbf{y}^D - 2\mathbf{xy}^D$$

$$\min\{ML1\,(D)\} = \max\,(\mathbf{xy}^D) \qquad \text{B3}$$

B3 is by definition equal to the BCC method.

Likewise, minimizing $ML2\,(D)$ (appendix A) in the case n_1, n_2 and s are white Gaussian processes, is equal to the BCC-method.

Appendix C

When x and y have a bivariate Gaussian distribution, the AAMI method is equivalent to the BCC method.

$$\max\{\mathrm{AAMI}_{XY}(x, y^D)\} = \max\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{p}_{XY}(x, y^D)\log\frac{\hat{p}_{XY}(x, y^D)}{\hat{p}_X(x)\,\hat{p}_Y(y)}dxdy\right\} \qquad \text{C1}$$

where:

$$\hat{p}_{XY}(x, y^D) = \frac{\exp\left[-\frac{1}{2}(x, y^D)^T R^{-1}(x, y)^D\right]}{2\Pi\,|R|^{\frac{1}{2}}}$$

while:

$$R = \begin{pmatrix} xx & xy^D \\ y^D x & yy \end{pmatrix} \qquad \text{C2}$$

Substituting C2 in C1:

$$\max\{\mathrm{AAMI}_{XY}(x, y^D)\} = \max\left\{\frac{xx \cdot yy}{xx \cdot yy - (x, y^D)^2}\right\} \qquad \text{C3}$$

In the case considered, C3 yields maximizing xy^D. This procedure is, therefore, by definition equal to the BCC method.

References

1. GEL'FAND, I. M.—YAGLOM, A. M.: Calculation of the amount of information about a random function contained in another such function. American Mathematical Society Translations, *12*, 199—246. (1959).
2. INOUYE, T.—YAGASAKI, A.—TAKAHASHI, H.—SHINOSAKI, K.: The dominant direction of interhemispheric EEG changes in the linguistic process. Electroencephalography and clinical Neurophysiology, *51*, 265—275. (1981).

3. KNAPP, C. H.—CARTER, G. C.: The generalized correlation method for estimation of time delay. IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP—24, 320—327 (1976).
4. MARKO, H.: The bidirectional communication theory, a generalization of information theory. IEEE Transactions on Communications, COM—21, 1345—1351. (1973).
5. MARS, N. J. I.—ARRAGON VAN, G. W.: Time delay estimation in nonlinear systems. IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP—29, 619—621 (1981).
6. SAITO, Y.—HARASHIMA, H.: Tracking of information within multichannel EEG record — causal analysis in EEG —. In: Yamaguchi, N. and Fujisawa, K. (eds), Advances in EEG and EMG Data Processing. Elsevier/North-Holland, Biomedical Press (1981).
7. SCARBROUGH, K.—AHMED, N.—CARTER, G. C.: On the simulation of a class of time delay estimation algorithms. IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP—29, 534—540 (1981).
8. SHORE, J. E.: Minimum cross-entropy spectral analysis. IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP—29, 230—237 (1981).
9. WYNER, A. D.: A definition of conditional mutual information for arbitrary ensembles. Information and Control, *38*, 51—59 (1978).

Ir. Cor FORTGENS, Department of Neuropsychiatry, Psychiatric Centre Rosenburg, Oude Haagweg 377, 2552 GB, The Hague, The Netherlands