Multi-channel piecewise selective averaging of cognitive evoked potentials with variable latency

Arthur Flexer
The Austrian Research Institute for Artificial Intelligence
Schottengasse 3, A-1010 Vienna, Austria
arthur@ai.univie.ac.at
Herbert Bauer
Institute of Psychology, University of Vienna
Liebiggasse 5, A-1010 Vienna, Austria

Abstract

This work is about the development of an alternative way of averaging evoked potentials (EP) of cognitive activities. Since the main assumption of invariant waveforms time locked to the eliciting events does not hold for cognitive EPs, averaging results in distorted estimates. Our alternative selective averaging finds similar subsequences of fixed length with variable latency which are common to all multi-channel EPs by transforming the multivariate time series to discrete sequences via vector quantization and applying a sequence alignment algorithm. The method yields a significant improvement over common averaging in terms of noise attenuation and is shown to be valid by comparison with results for random data. Results for EP data obtained during a spatial imagination task are reported.

1 Introduction

An evoked potential (EP) is the electro cortical potential measurable in the human electro encephalogram (EEG) before, during and after sensoric, motoric or cognitive events. This work is about the development of a new method for the analysis of EPs of cognitive activities for which already existing methods for the simpler sensoric or motoric EPs are not suited.

An EP is defined as the combination of the brain electric activity that occurs in association with the eliciting event and 'noise', which is brain activity not related to the event together with inference from non-neural sources. Since the noise contained in EPs is significantly stronger than the signal, the common approach is to compute an average across several EPs recorded under equal conditions to improve the signal-to-noise ratio. For common averaging of a set of N EPs, the assumed model for the ith EP (given here for only one channel of EP) is the following [6]:

$$x_i(t) = s(t) + n_i(t); \quad i = 1, 2, \ldots N; \quad 0 \leq t < T \tag{1}$$

where $x_i(t)$ is the ith recorded EP, $s(t)$ the underlying signal, $n_i(t)$ the noise associated with the ith EP, and T the duration over which each EP is recorded. The average $\hat{s}(t)$ over the sample of N EPs is used to estimate the underlying signal $s(t)$:

$$\hat{s}(t) = \frac{1}{N}\sum_{i=1}^{N} x_i(t) = s(t) + \frac{1}{N}\sum_{i=1}^{N} n_i(t) \tag{2}$$

Averaging will attenuate the noise $n_i(t)$ and not the signal $s(t)$ given that signal and noise linearly sum together to produce the recorded EPs $x_i(t)$ and the evoked signal $s(t)$ is the same for each recorded EP $x_i(t)$ and the noise contributions $n_i(t)$ can be considered to constitute statistically independent samples of a random process. Averaging is standard for the analysis of exogenous components of motoric and sensoric events like the $N100$ waveform which lie within $100ms$ after the eliciting event. If the above assumptions do not hold, the averaging will result in a biased, distorted estimate of the signal.

Cognitive evoked potentials are endogenous components of the EPs and start about $100ms$ after the onset of the recording and can last several seconds. They do not elicit one specific EP waveform time locked to the onset of the recording, and their analysis is a largely unsolved problem in psychophysiology. Classical methods like [15, 9, 8, 13] and [14] are designed for univariate time series of simpler motoric or sensoric EPs only. They usually assume that the recorded univariate signal is the same for all EPs during the whole duration of the recording but allow variable latencies of the common waveform. Therefore they cannot really cope with the harder problem of analysing cognitive EPs. Other more modern approaches like e.g. Independent Component Analysis [7] use EP data of short duration ($\leq 1sec$) after averaging. Another approach is to average despite improper assumptions thereby loosing all information about faster components and to analyse only positive and negative shifts of very slow potentials [2].

Our hypothesis concerning cognitive EPs is that only pieces (subsequences) of the whole EPs with variable latencies can be expected to be due to the cognitive task. Our approach to discover such subsequences which are common to all EPs across all recording channels is to replace the sequence of the original multi dimensional vectors (our data was measured via 22 electrodes) by a sequence of codebook vectors obtained via vector quantization. The trajectories across codebook vectors are univariate discrete time series to which we can apply a multiple sequence alignment procedure [1] that has originally been designed for molecular biology.

The data was recorded with 22 electrodes with an equidistant matrix montage during a spatial imagination task (Three-dimensional Cube Test 3DC [5]) and consists of 319 EP trials from 10 good female spatializers plus 167 EP trials from 8 poor female spatializers. After appropriate preprocessing (digital low pass filtering to frequencies below $8Hz$ and eliminating the DC-like trend by subtracting a linear fit), each EP trial lasts $8.5sec$ consisting of 2125 samples, each being a 22 dimensional real valued vector.

2 Computation of trajectories across codebook vectors

All EP time series are vector quantized together by using all EP vectors at all sample points as input vectors to a clustering algorithm disregarding their ordering in time. Then the sequence of original vectors x is replaced by the sequence of codebook vectors $\hat{x}$. K-means clustering (see e.g. [3, p.201]) is used for vector quantization using the sum of squared differences $d(x, \hat{x}) = \sum_{i=0}^{k-1} \mid x_i - \hat{x}_i \mid^2$ as measure of distance, where both x and $\hat{x}$ are of dimension k. Since observation of the sum of distances $d(x, \hat{x})$ with growing size of codebooks did not indicate an optimal codebook size, we pragmatically decided to use 64 codebook vectors based on the following consideration: We vector quantized the EP data with increasing numbers of codebook vectors. We then took the original EPs and substituted at each sample point the original real valued vector x_l with the appropriate codebook vector $\hat{x}_i$ (i.e. where $d(x_l, \hat{x}_i) < d(x_l, \hat{x}_s)$ for all s). We then visually inspected both the original EP time series and the coarser codebook time series as series of topographical patterns (spherical spline interpolations of the 22 values at a single point in time) and checked, whether the important features (positive and negative peaks and their development in time) of the topographies still were existent in the coarser codebook approximation. The high number of different discrete symbols (64 code book vectors) did not allow for a more principled information theoretic approach to obtain an optimal codebook size. For the 64 codebook vectors, we calculated a 64×64 distance matrix D_C.

3 Sequence alignment and selective averaging

We chose a so-called *fixed length subsequence* approach for comparison of the sequences made of 64 discrete symbols (corresponding to 64 codebook vectors $\hat{x}$). Given two sequences E and F of length m, all possible overlapping subsequences having a particular window length W from E are compared to all subsequences from F. For each pair of elements the score taken from the distance matrix D_C is recorded and summed up for the comparison of subsequences. The distance between two subsequences of length W from two sequences E and F is therefore:

$$D_{align}(b_e, b_f, W) = \sum_{i=0}^{W-1} d(E_{b_e+i}, F_{b_f+i}) \qquad (3)$$

The indices b_e and b_f are the beginning points of the subsequences in the sequences E and F and E_{b_e+i} and F_{b_f+i} are the corresponding codebook vectors. Successive application of this pairwise methods allows for the alignment of more than two sequences. Such a *fixed subsequence* approach that is explicitly designed for *multiple sequence alignment* is given in [1]. It computes a multiple alignment by iteratively comparing sequences to the multiple alignment obtained so far, keeping always just the L best subsequences as an intermediate result. The succession of sequences is chosen at random. When a subsequence is compared to an intermediate "more"-way (let us say p-way) subsequence, the resulting score is computed as the sum of the p pairwise comparisons of the subsequences in the intermediate solution with the new subsequence that is to be aligned. The number of all such crosswise comparisons within the final overall alignment is given by $P = \sum_{i=1}^{p-1} i$. The number of all element-wise comparisons within the final overall alignment is given by WP, and its average per element, the average element-wise within alignment distance, by:

$$\bar{D}_{align} = \frac{1}{WP} \sum_{i=1}^{p-1} \sum_{j=i+1}^{p-1} D_{align}(b_i, b_j, W) \qquad (4)$$

Desired is a set of beginning points b_i^{min} for which $\bar{D}_{align}$ is minimal. The b_i^{min} are the same for all $d = 22$ channels of the corresponding ith EP. For each channel of EP we can compute an alternative selective average $\hat{s}'(t)$ where the duration T is equal to the length of the subsequences, W, and the beginning points of the averaging are the parameters b_i^{min}:

$$\hat{s}'(t) = \frac{1}{N} \sum_{i=1}^{N} x_i(b_i^{min} + t); \quad 0 \leq t < W \qquad (5)$$

This approach guarantees that the obtained multiple alignments contain subsequences that are part of all the original sequences. The number of single element-wise comparisons is $LW(m+1-W)P$. For a given L and m, this function is proportional to p^2, in contrast with m^p comparisons in "brute force" searching where not just the L best but all possible alignments are considered. As experiments with L equal 100, 1000 and 10000 showed, it is sufficient to keep 100 intermediate results to avoid the omission of good alignments that are weak in the first few sequences but strong in the later ones. Experiments varying the window length W from 31 to 62, 125 and 187 showed that $W = 125$ (corresponding to $500ms$ of EP) is short enough to yield alignments of satisfactory quality which are still long enough to be significant in terms of their psychophysiological interpretation. For more detail on tuning of the parameters L and W see [4].

4 Results

In a related study [4] working on a subset of our EP data we have shown the statistical significance of our approach. We verified that our procedure yields better results for real human EPs than for unstructured random input in terms of average element-wise within alignment distance $\bar{D}_{align}$ (see Equ. 4). We compared results obtained from 21 EPs of one test subject with time-shuffled EPs and artificial EPs. The latter consisted of random Gaussian sequences whose power spectrum was changed appropriately to resemble the characteristics of real EPs. A one-way analysis of variance plus additional Duncan t-Tests allowed us to rank the result for real EP as being significantly better than the result for time-shuffled EP, which is again significantly better than the result for random Gaussian EP.

To compare the gain in noise attenuation of the common average and of our selective average, the respective estimated standard deviations of the background noise, $\hat{\sigma}(t)$ and $\hat{\sigma}'(t)$, are being compared.

$$\hat{\sigma}(t) = \left[\frac{\sum_{i=1}^{N} [x_i(t) - \hat{s}(t)]^2}{N-1} \right]^{\frac{1}{2}} \qquad (6)$$

$$\hat{\sigma}'(t) = \left[\frac{\sum_{i=1}^{N}\left[x_i(b_i^{min}+t) - \hat{s}'(t)\right]^2}{N-1}\right]^{\frac{1}{2}} \qquad (7)$$

Since the $\hat{\sigma}(t)$ and $\hat{\sigma}'(t)$ are given for each of the $d = 22$ channels and for the duration of $t = m$ or $t = W$ respectively, the following average estimates of the standard deviations of the background noise are being computed:

$$\hat{S} = \frac{1}{dm}\sum_{j=1}^{d}\sum_{t=0}^{m-1}\hat{\sigma}_j(t) \qquad (8)$$

$$\hat{S}' = \frac{1}{dW}\sum_{j=1}^{d}\sum_{t=0}^{W-1}\hat{\sigma}'_j(t) \qquad (9)$$

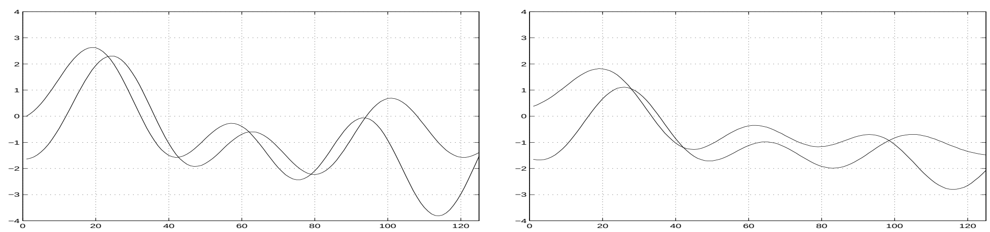
$\hat{S}$ is the estimate for the common averaging and $\hat{S}'$ for the selective averaging. An $\hat{\sigma}_j(t)$ is the $\hat{\sigma}(t)$ for channel j given by Equ. 6. An $\hat{\sigma}'_j(t)$ is the $\hat{\sigma}'(t)$ for channel j given by Equ. 7. For all EPs of good and poor spatializers the common average $\hat{s}$ as well as five selective averages $\hat{s}'$ have been computed. Results for the good spatializers were $\hat{S} = 7.68$ vs. mean $\hat{s}' = 4.35 \pm .068$ and for the poor spatializers $\hat{S} = 7.84$ vs. mean $\hat{s}' = 4.37 \pm .048$. Computing Z-values shows the differences in noise attenuation to be significant: $Z_{good} = |(4.35 - 7.68)/(.068/\sqrt{5})| = |-109.5| > Z_{99} = 2.58$; $Z_{poor} = |(4.37 - 7.84)/(.048/\sqrt{5})| = |-161.6| > Z_{99} = 2.58$. The estimated expected magnitude of the noise residual is now only ≈ 0.56 times that of the respective common averages. This is a gain in noise attenuation of more than 40%.

The repeated runs of the fixed subsequence algorithm showed a problem with the reproducibility of the selective averages. Although a majority of the obtained beginning points b_i^{min} are identical, there is a difference in the selective averages due to the different succession in which the sequences are being compared. We are able to overcome this problem by combining five selective averages to one overall selective average: we obtain overall beginning points b_i^{min5} at the points in time where the sums of the respective five $\bar{D}_{align}$ (cf. Equ. 4) of the single selective averages are minimal. This "averaging" across five solutions diminishes their variability. We computed two such overall selective averages by combining 2×5 selective averages to two sets of beginning points b_i^{min5}. The resulting pairs of overall selective averages are now very similar which is corroborated by visual inspection (Fig. 1).

The results of computing selective averages for both good and poor spatializers via beginning points b_i^{min5} are given in Fig. 2 as sequences of topographical patterns. Each topography is a spherical spline interpolation of the 22 values at a single point in time of the selective averaging window. Given are topographies at $40, 80, \cdots, 440, 480 msec$ of the window for poor spatializers (top two rows) and good spatializers (lower two rows). We can see that for both good and poor spatializers there is one specific dominant topographical pattern visible, albeit at changing levels of amplitude. It is a pattern of more positive amplitudes at frontal to central regions relative to more negative amplitudes at occipital to parietal regions. This common topographical pattern is generally more negative for poor spatializers.

5 Discussion

Our selective averaging makes possible the analysis of cognitive EPs by finding common pieces (subsequences) of a set of EPs which have fixed length but variable latencies and are sufficiently similar across all EP channels. We were able to validate our approach both on a statistical basis and in terms of the psychophysiological content of the obtained results. The comparison with artificial random EPs shows the statistical significance of our algorithm. The selective averaging almost doubles the noise attenuation compared to common averaging. The more negative amplitudes at occipital to parietal regions visible in Fig. 2 are well in line with both clinical findings and related EP studies about spatial processing. A study [12] using the exact same experimental design as in this work, but with male subjects and applying common averaging to analyse positive and negative shifts of very slow potentials, confirmed the importance of occipito-parietal regions by showing significant DC-negativation during spatial tasks. In agreement with our results, the authors also observed a significantly higher investment of cortical effort with poorer spatializers investing more activity (i.e. showing more pronounced negative topographical patterns). We additionally observed that the temporal structure of spatial information processing is a sequence of activations and in-activations characterized by sequences of negative- and positive-going



(a) (b)

Figure 1: Comparison of selective averages computed via beginning points b_i^{min5} obtained by minimizing the sum of $\bar{D}_{align}$ from five repeated runs each. (a) is for electrode F1, (b) for P2, both for poor spatializers.

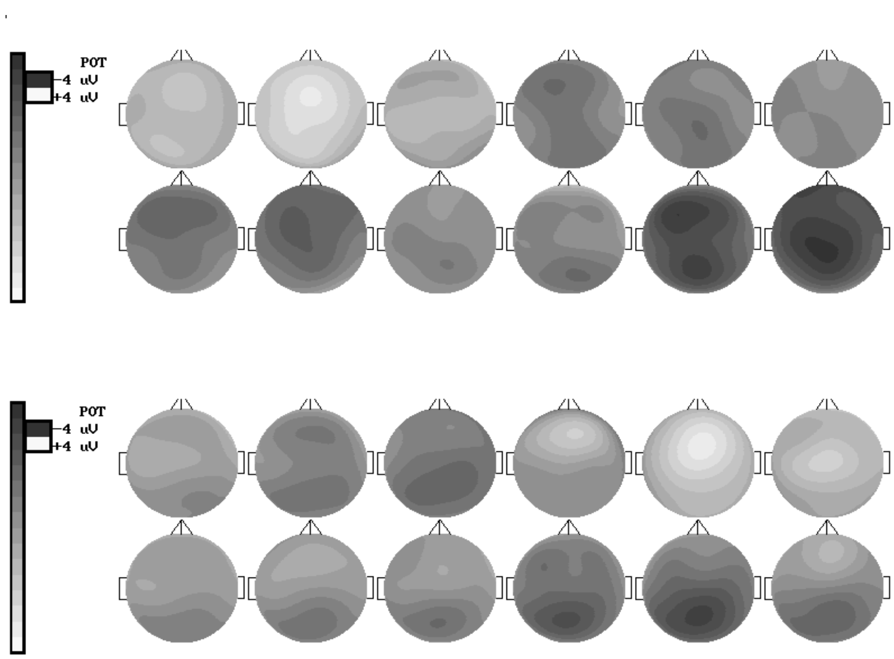


Figure 2: Sequences of topographies for poor spatializers (top two rows) and good spatializers (lower two rows). Scale is from -4 to $+4mV$.

potential changes of one dominant pattern. This shows that our method allows for a richer analysis of the temporal structure of the cognitive processes under investigation.

Finally we would like to comment on possible improvements concerning the techniques being used. Our decision to use batch k-means clustering for vector quantization was based on a thorough review of literature plus some comparative computer experiments which both showed batch k-means clustering to be a good choice. Our empirical decision for a codebook size of 64 could be replaced employing an objective measure obtained via a Bayesian approach to mixture modelling (see e.g. [11]). However, codebook centers obtained via clustering serve only as intermediate representations enabling efficient sequence comparison. Computation of selective averages is done using the original signals and not their codebook representations. Errors made during clustering could harm final results only in a very indirect way. Our sequence comparison algorithm finds subsequences which are the same across different subjects both in their topographical *and* timely appearance for the whole length of the analysis window. If we want to allow subsequences to additionally show considerable variation on the time axis, i.e. compression and expansion, algorithms that are able to deal with gaps in the comparison of sequences are needed. Hidden Markov Models (see e.g. [10]) are a promising candidate to solve this problem.

Acknowledgements: The EEG recordings have been made by R. Gstättner, Dept. of Psychology, University of Vienna. Parts of this work were done within the BIOMED-2 BMH4-CT97-2040 project SIESTA, funded by the EC DG XII. The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry of Science and Transport. The author was supported by a doctoral grant of the Austrian Academy of Sciences.

References

[1] Bacon D.J., Anderson W.F.: Multiple Sequence Alignment, Journal of Molecular Biology, 191, 153-161, 1986.

[2] Bauer H.: Slow potential topography, Behavior Research Methods, Instruments and Computers, 30(1), 20-33, 1998.

[3] Duda R.O., Hart P.E.: Pattern Classification and Scene Analysis, John Wiley & Sons, N.Y., 1973.

[4] Flexer A., Bauer H.: Discovery of Common Subsequences in Cognitive Evoked Potentials, in Zytkow J.M. & Quafafou M.(eds.), Principles of Data Mining and Knowledge Discovery, Second European Symposium, PKDD '98, Lecture Notes in Artificial Intelligence 1510, p.309-317, Springer, 1998.

[5] Gittler G.: Dreidimensionaler Wuerfeltest (3DW). Ein Rasch-skalierter Test zur Messung des raeumlichen Vorstellungsvermoegens. Theoretische Grundlagen und Manual, Beltz Test, Weinheim, 1990.

[6] Glaser E.M., Ruchkin D.S.: Principles of Neurobiological Signal Processing, Academic Press, 1976.

[7] Makeig S., Jung T-P., Bell A.J., Ghahremani D., Sejnowski T.J.: Blind Separation of Auditory Event-related Brain Responses into Independent Components, Proc. Natl. Acad. Sci. USA, 94:10979-10984, 1997.

[8] McGillem C.D., Aunon J.I.: Measurements of signal components in single visually evoked brain potentials, IEEE Transactions on Biomedical Engineering, 24, 232-241, 1977.

[9] Pfurtscheller G., Cooper R.: Selective averaging of the intracerebral click evoked responses in man: an improved method of measuring latencies and amplitudes, Electroencephalography and Clinical Neurophysiology, 38: 187-190, 1975.

[10] Rabiner L.R., Juang B.H.: An Introduction To Hidden Markov Models, IEEE ASSP Magazine, 3(1):4-16, 1986.

[11] Roberts S., Husmeier D., Rezek I., Penny W.: Bayesian Approaches to Gaussian Mixture Modelling, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 11, 1133-1142, 1998.

[12] Vitouch O., Bauer H., Gittler G., Leodolter M., Leodolter U.: Cortical activity of good and poor spatial test performers during spatial and verbal processing studied with Slow Potential Topography, International Journal of Psychophysiology, Volume 27, Issue 3, p.183-199, 1997.

[13] Weerd J.P.C.de, Kap J.I.: A Posteriori Time-Varying Filtering of Averaged Evoked Potentials, Biological Cybernetics, 41, 223-234, 1981.

[14] Westerkamp J., Aunon J.I.: Optimum multi-electrode a posteriori estimates of single-response evoked potentials, IEEE Transactions on Biomedical Engineering, 34, 13-22, 1987.

[15] Woody C.D.: Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals, Medical and Biological Engineering, 5, 539-553, 1967.