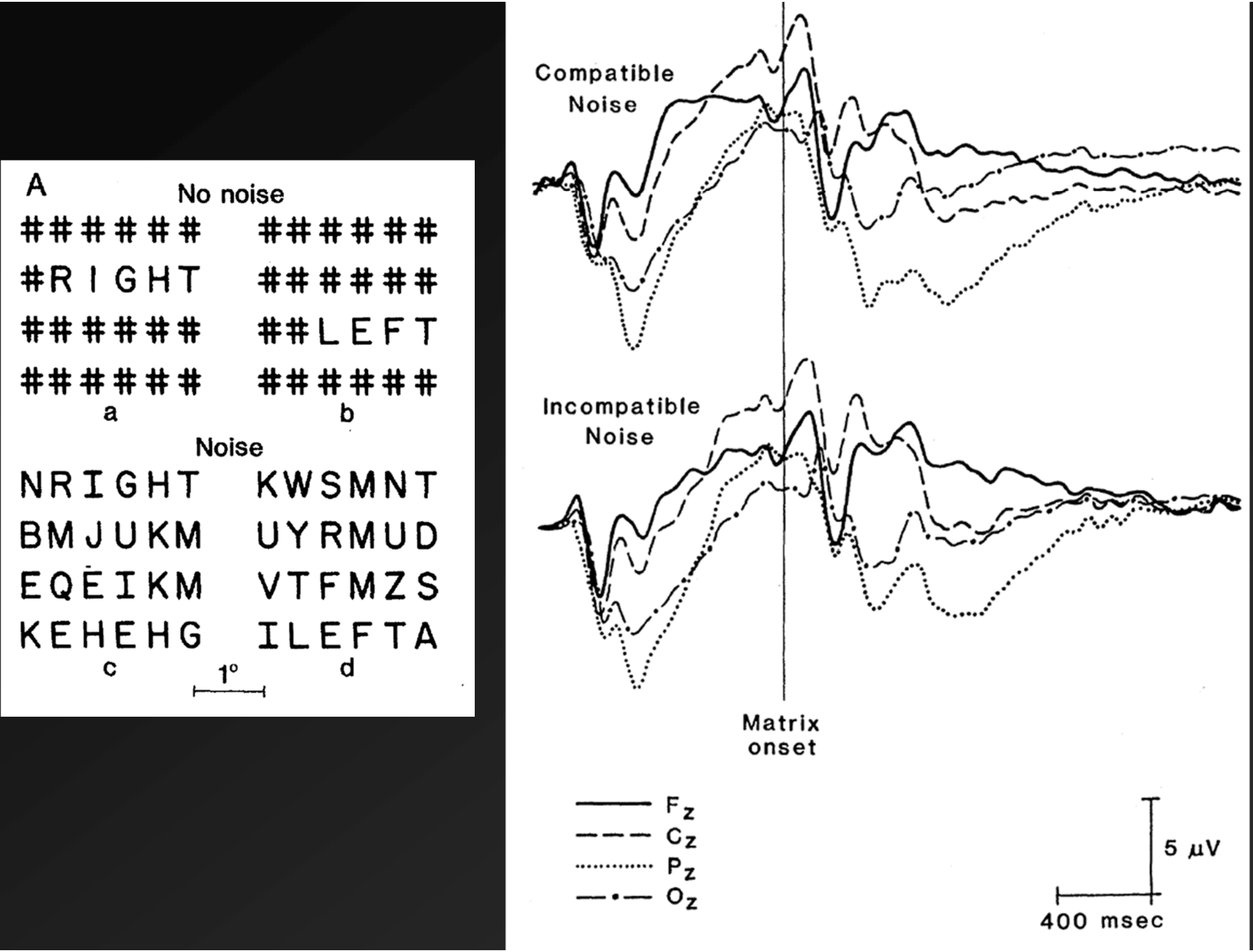
Advanced Signal Processing II
(aka Acronym Day)

Latency Jitter and Woody Filters (acronym free)
Inference Problems with Scalp Topography (acronym free)
PCA
ICA
BESA
Simultaneous ERP with ICA and fMRI!

The Problem of Latency Jitter

- The averaging assumption of invariance in signal is not always warranted
  - Especially for the later endogenous components
  - To the extent that the signal varies from trial to trial, the average will produce potentially misleading results
- Two common possibilities:
  - Smearing of components;
    - will underestimate amplitude of component (especially a problem if comparing groups, one group with more latency jitter)
  - Bimodal or multi-bumped components


A
No noise
######
#RIGHT
######
######
a
######
##LEFT
######
b
Noise
NRIGHT
BMJUKM
EQEIKM
KEHEHG
c
KWSMNT
UYRMUD
VTFMZS
ILEFTA
d
1°
Compatible
Noise
Incompatible
Noise
Matrix
onset
Fz
Cz
Pz
Oz
5 μV
400 msec

The Solution

- The Woody Adaptive Filter (Woody, 1967)
- Based on Cross-correlation
  - Assumptions less restrictive than averaging methods
    - Waveform (morphology) must be constant across trials
    - Latency need not be constant

Details

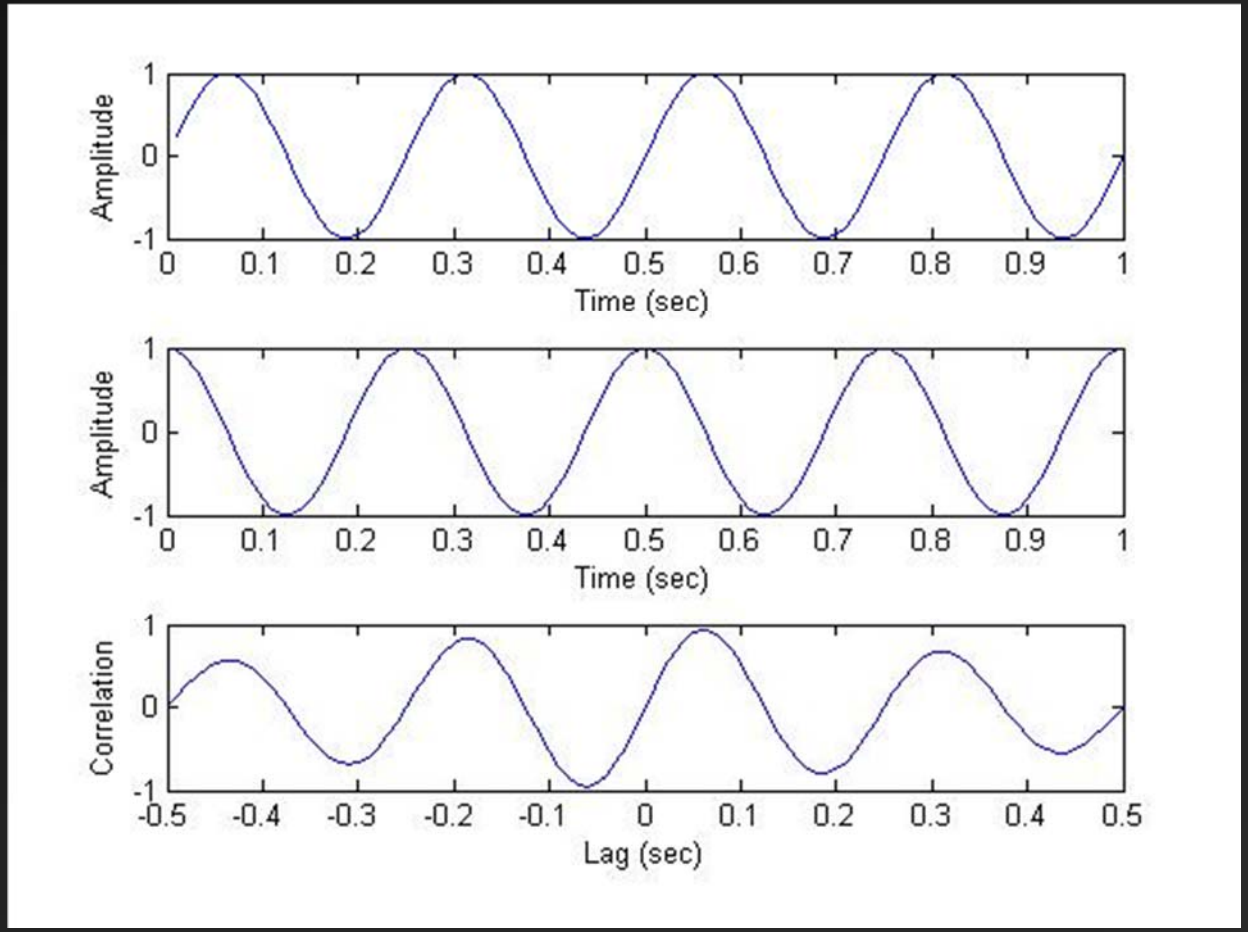
- Cross-correlational series
  - For two waveforms the correlation between each of them is computed
    - first with no lag in time
      a1, a2, ..., an
      b1, b2, ... bn
    - then with one lagged with respect to the other
      a1, a2, ..., an-1
      b2, b3, ... bn
  - A series of correlation values is obtained by progressively increasing the size of the lag

The Basic Idea

Sine

Cosine

Cross-Correlation



See … CrossCorr_Sin_Cos.m

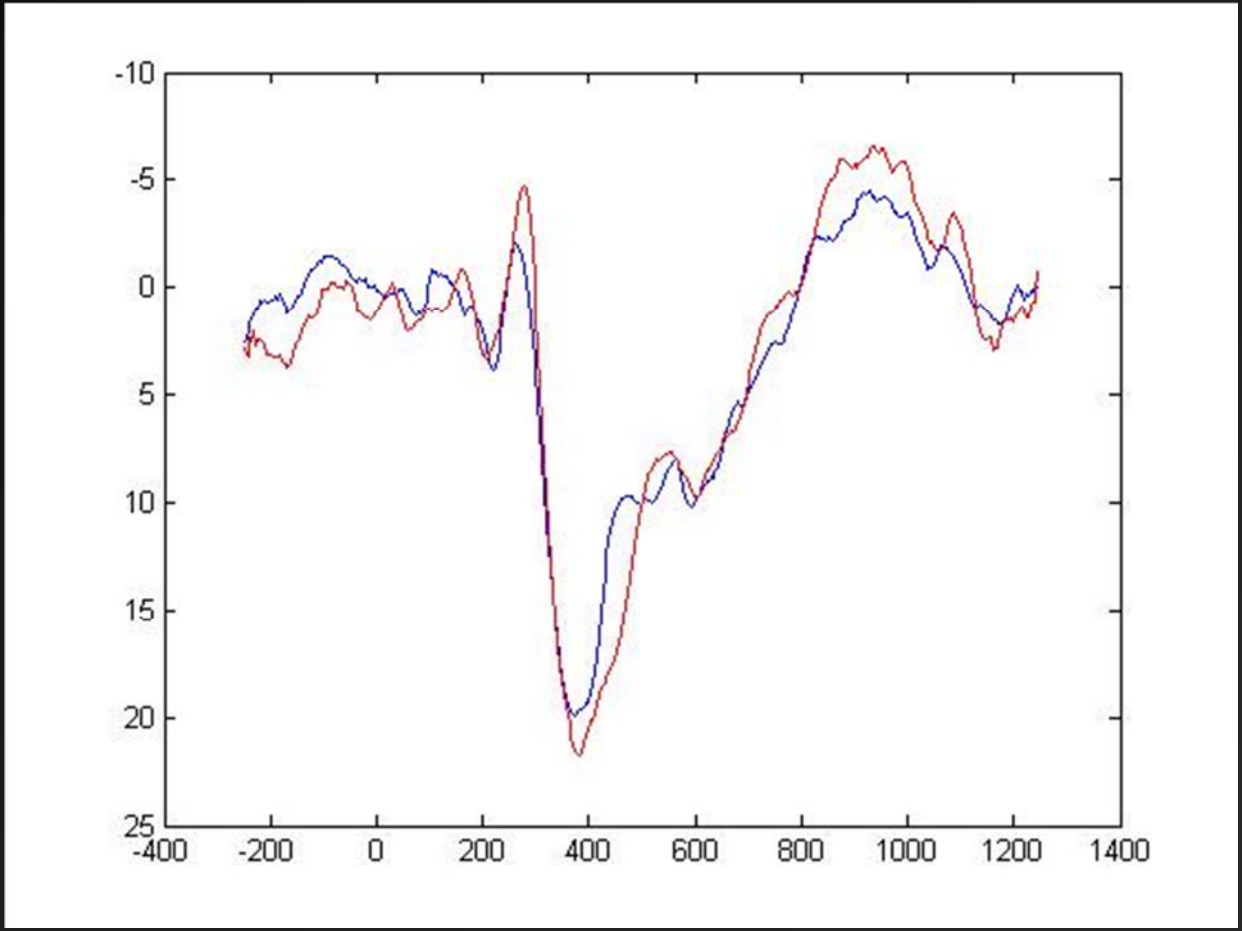
More Details

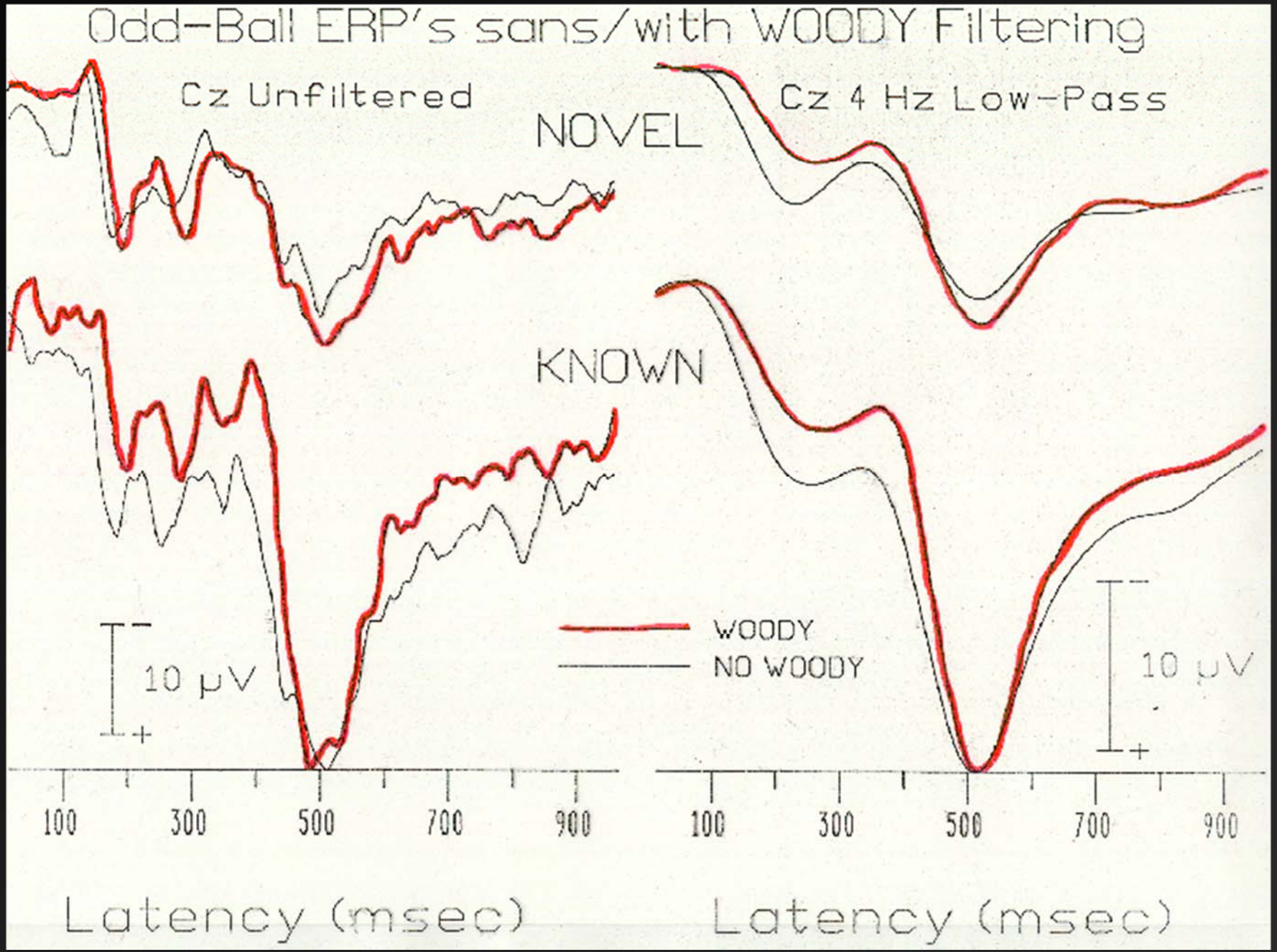
- Can be used as a "template matching" procedure
- Compare running average with raw EEG epochs
- This is a method of single-trial signal detection:
  - First create a template: either predetermined (e.g., sine wave) or empirically determined (e.g., average)
  - Then calculate cross-correlational series between each raw EEG epoch and the template
  - If some maximum correlation achieved, conclude signal is present
  - If correlation not achieved conclude absent
  - This can also be used as a method of determining the latency of a component (by examining the trial-by-trial shifts), or of determining the variability in response for a given individual (again by examining the trial-by-trail shifts)

Woody's Instantiation

- The Woody Adaptive Filter (Charles Woody, 1967) is a special case and application of cross correlational technique
- The term "adaptive" refers to the fact that the template is not established a priori, but generated and updated by an iterative procedure from the data themselves
- Procedure
  - Initial template is usually either a half cycle of a sine or triangle wave, or the unfiltered average of single trials
  - Cross-lagged correlations (or sometimes covariances) are then computed between each trial and this template typically over a limited range of samples (e.g., region of P300, not over "invariant" components)
  - Each trial is then shifted to align it with the template at the value which yields the maximum cross correlation (or covariance)
  - A new template is then generated by averaging together these time-shifted epochs
  - Procedure is repeated using this new average as the template
  - repeated until the maximal values of the cross correlation become stable
  - often, average cross-correlation value increment monitored; if $\underline{r}$ increases < .005 or .001, then stability achieved
- Some implementations, trials which do not reach a minimum criterion (e.g., .30-.50) are discarded from subsequent template construction and perhaps from subsequent analysis altogether

Woody Filtering Demo!



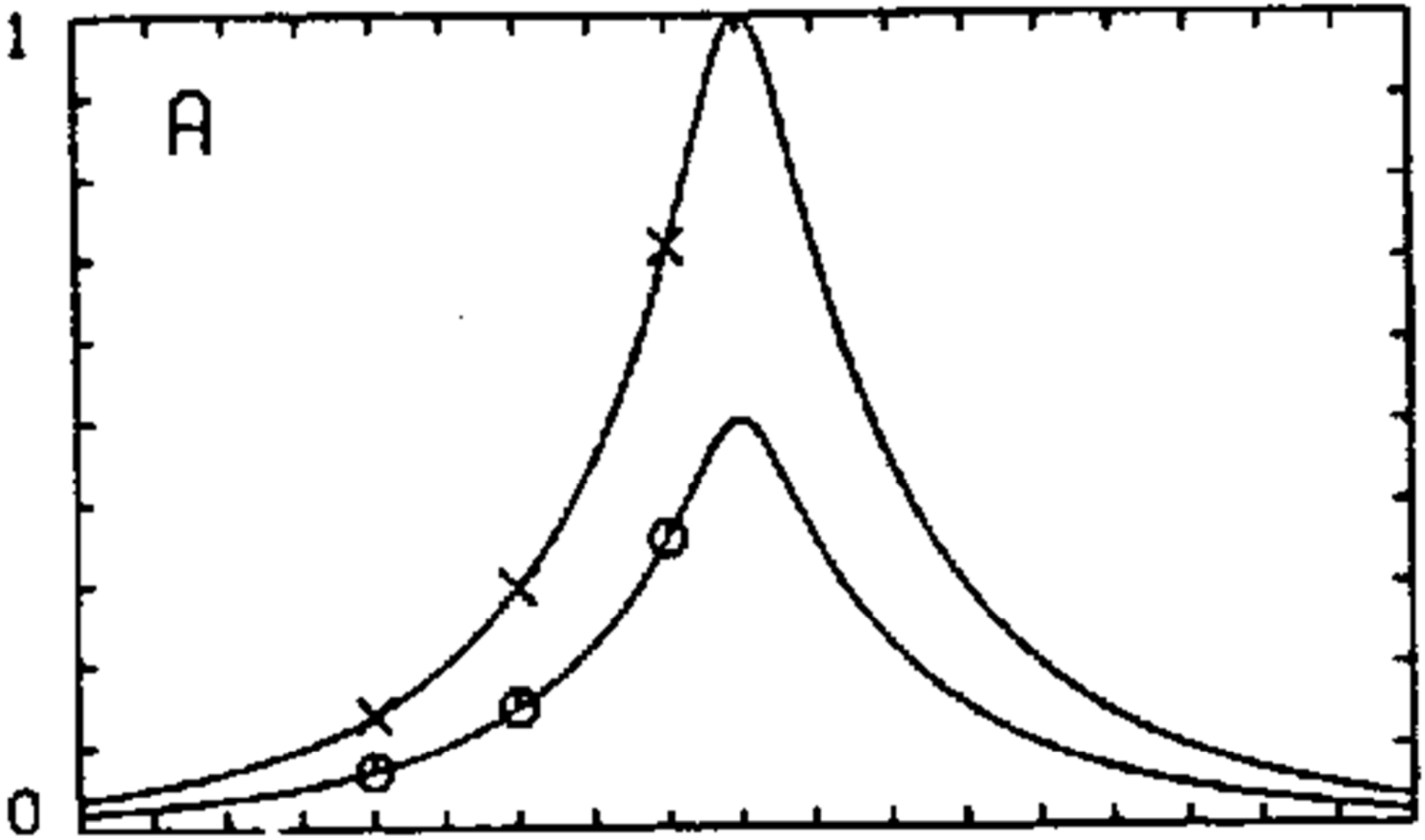

Odd-Ball ERP's sans/with WOODY Filtering
Cz Unfiltered
Cz 4 Hz Low-Pass
NOVEL
KNOWN
WOODY
NO WOODY
10 μV
10 μV
100
300
500
700
900
100
300
500
700
900
Latency (msec)
Latency (msec)

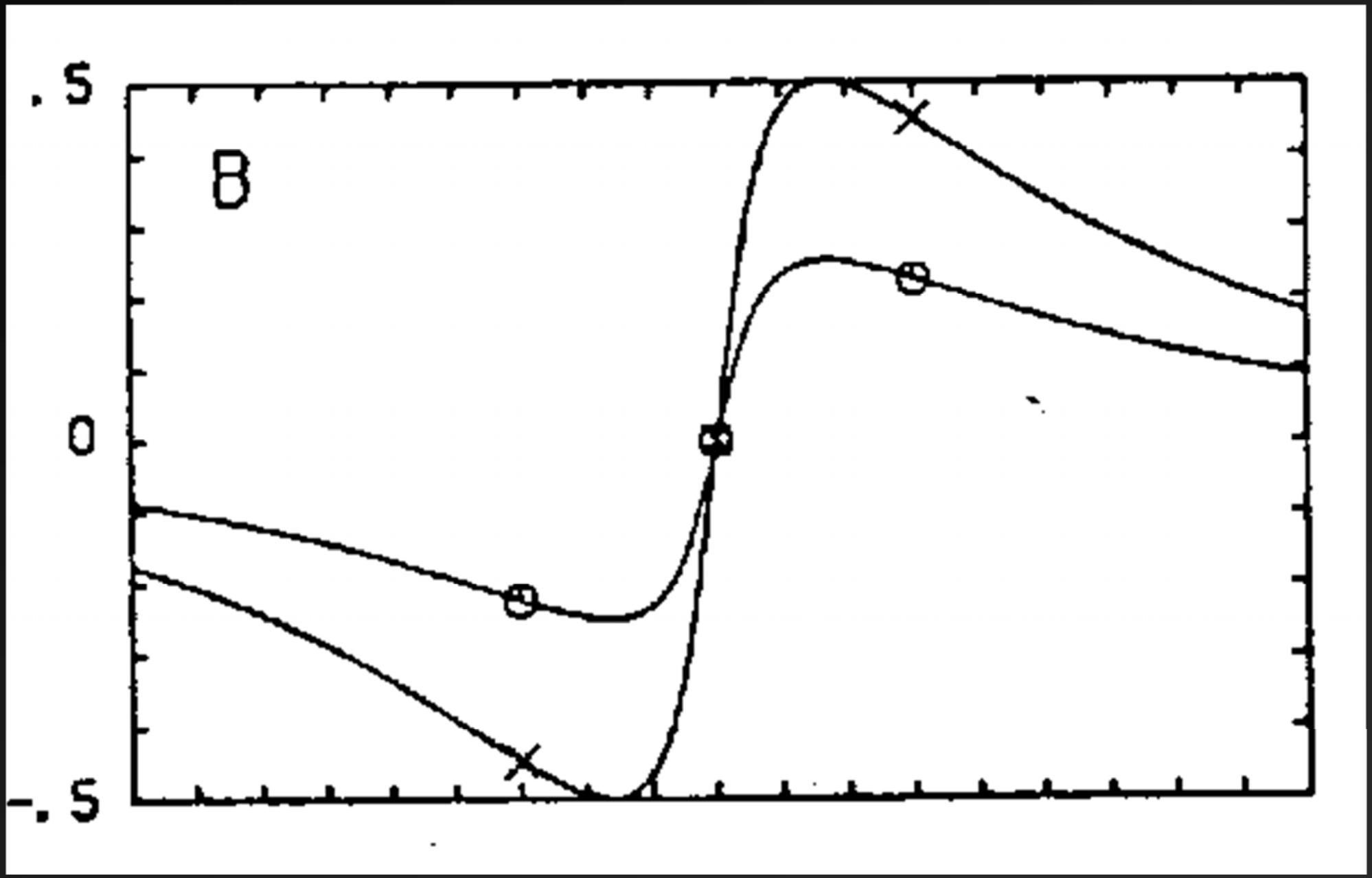
Validity

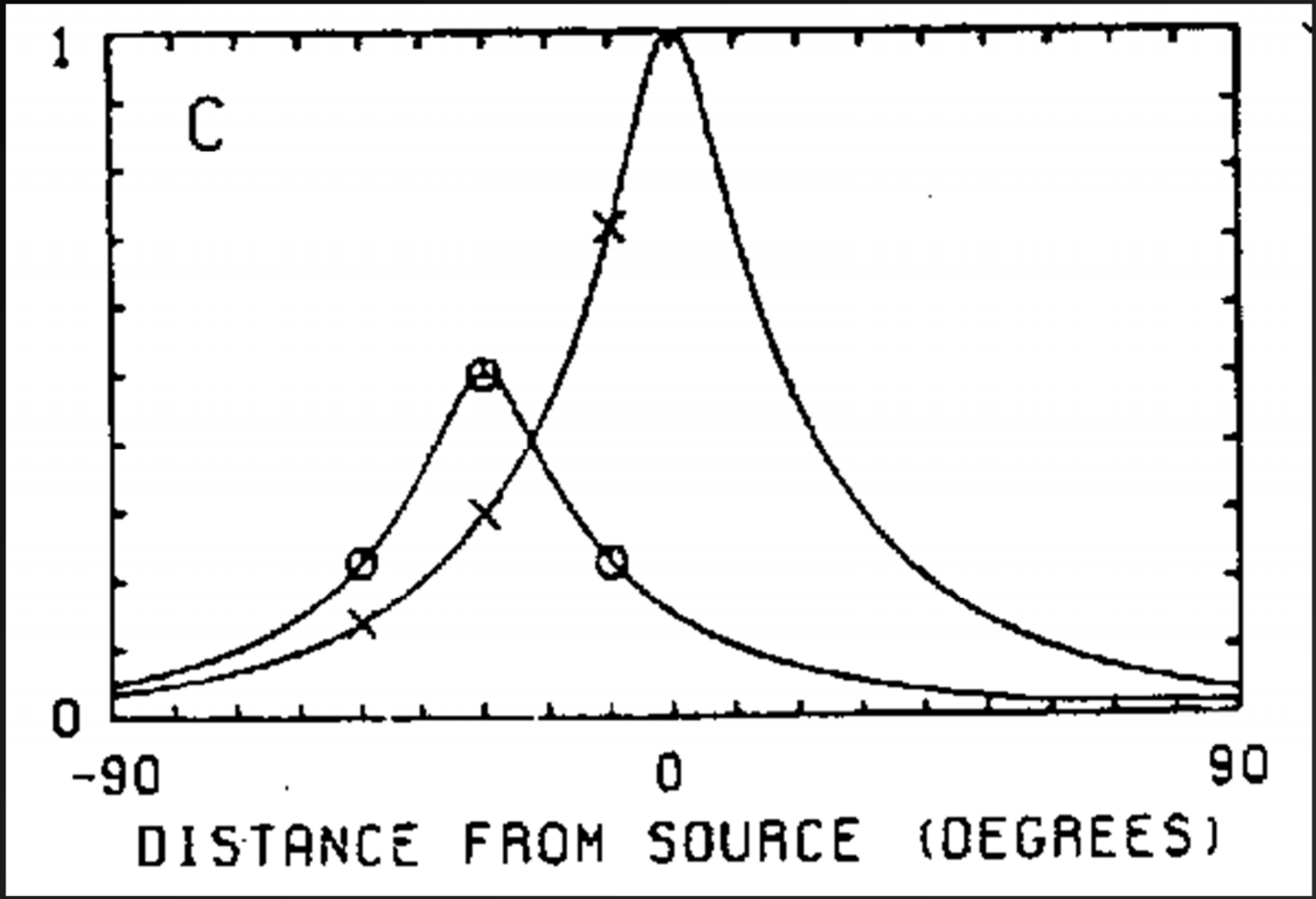
- Seems to do a fair job of improving signal extraction if a few iterations are used and if the original signal itself is singly peaked
- Wastell(1977) reports a decline in the validity of the procedure if numerous iterations are used
- Therefore, unlike averaging, Woody filtering can only improve signal-to-noise ratio over a definite limit
- Suggests also that Woody may not be the solution under conditions of very low signal-to-noise ratio

Using Scalp Topography to Infer Different Generators

- Assumption is that if there are different source generators between, there will be different resultant scalp distributions
- Therefore would expect to find a Scalp site by Condition interaction in ANOVA
- The Problem (Wood & McCarthy, 1985)
  - Potentials do not propagate to scalp in strictly additive manner
  - Same source at different strengths can produce a Scalp site by Condition interaction


1
A
0


.5
B
0
-.5


1
C
0
-90
0
90
DISTANCE FROM SOURCE (DEGREES)

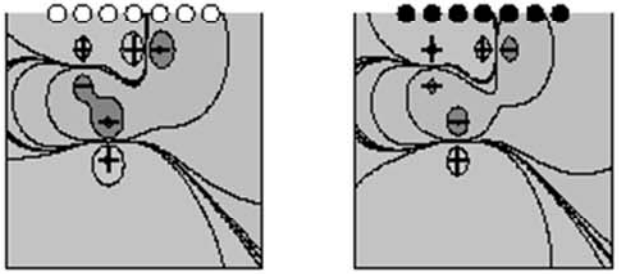
The Solution

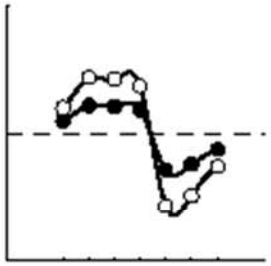
- Normalization
  - For each condition, scale data (e.g. by dividing by site of maximum amplitude)
- Eliminates any overall condition main effect
  - Condition main effect must be assessed in standard (non-scaled) ANOVA
  - Scaled data now lead to an interpretable interaction
  - If interaction survives scaling, then one can reasonably infer different intra-cranial generators

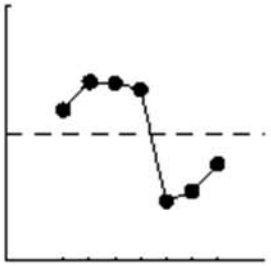
A New Problem

- Urbach & Kutas point out that the solution is not a solution! It's intractable
  - For single point source that is invariant in rotation, perhaps Wood & McCarthy were right
  - But when dipole rotates (e.g. on a gyrus), changes polarity, the W&M strategy will not work
  - When there are multiple generators, with changes in *relative* strength, W&M strategy will not work

A) Generator distributions differing only in overall strength







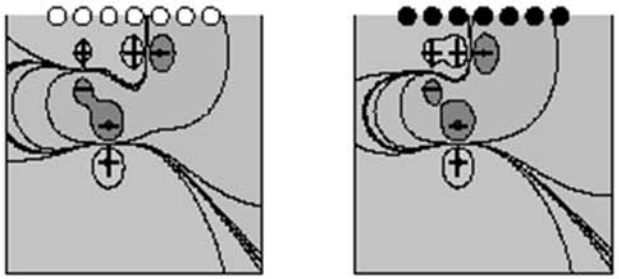
Unscaled potential distributions

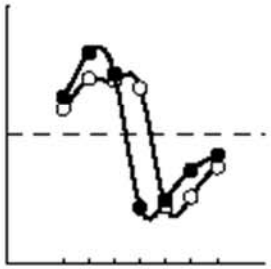
Vector scaled distributions

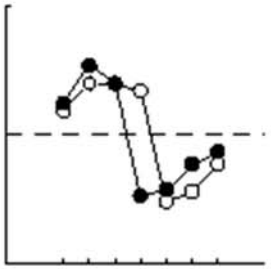
Inference to source configurations

Not different

B) Generator distributions differing in location

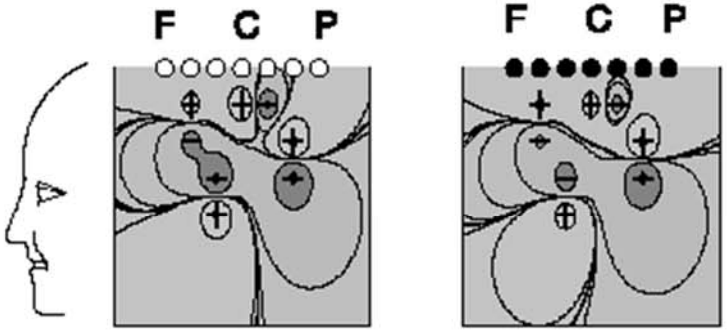


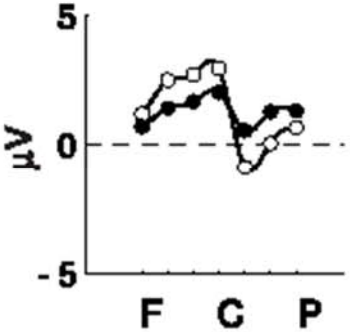


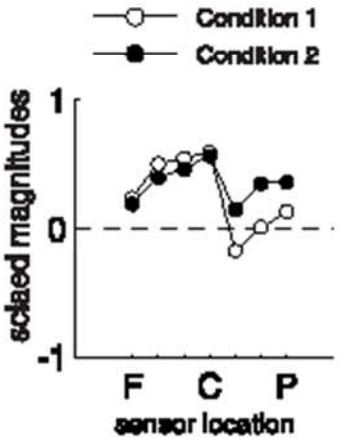


Different

C) Generator distributions differing in relative strength







Different

If, and only if…

- W&M procedure produces valid inferences if and only if two generator distributions *G*1, *G*2, are multiplicatively related
- Two generator distributions are multiplicatively related iff:
  - 1. The locations of the generators are all the same AND
  - 2. The polarities of the generators are all the same AND
  - 3. The intensities of the generators differ in overall strength, not *relative* strength
- But how would you ever know, unless you *knew* where the generators were
  - … in which case you would not be using the W&M procedure!

So, where's that leave us?

- If you scale the amplitudes and there is no interaction between condition and site, then the generators are not different
- But if there is such an interaction, you don't know whether:
  - generators differ in location OR
  - generators differ in polarity OR
  - generators differ in *relative* strength
- So a nonsignificant effect is informative

Principal Components Analysis

- A method for reducing massive data sets
- See Handout for gory details

PCA (1): The Data matrix

$$\mathbf{D}_{Nxn} =$$

Subject #1	$[t_0, t_1, t_2, \ldots, t_{n-1}$
Subject #2	$t_0, t_1, t_2, \ldots, t_{n-1}$
Subject #3	$t_0, t_1, t_2, \ldots, t_{n-1}$
...	...
...	...
Subject #N	$t_0, t_1, t_2, \ldots, t_{n-1}]$

Where N = Number subjects
n = Number sample points per average
t = voltage at time point 0, 1, ...

- Data Matrix above shows only one site – could have multiple sites by adding rows for each subject
- This data matrix is for “temporal PCA” but one could transpose for “spatial PCA”

PCA (2): The Score matrix

$$\mathbf{S}_{Nxm} = \begin{matrix} \text{Subject \#1} \\ \text{Subject \#2} \\ \text{Subject \#3} \\ \ldots \\ \ldots \\ \text{Subject \#N} \end{matrix} \begin{bmatrix} s_1, & s_2, & s_3, & \ldots, & s_m \\ s_1, & s_2, & s_3, & \ldots, & s_m \\ s_1, & s_2, & s_3, & \ldots, & s_m \\ & & \ldots & & \\ & & \ldots & & \\ s_1, & s_2, & s_3, & \ldots, & s_m \end{bmatrix}$$

Where N = Number subjects
m = Number of components
s = score on component 1, 2, ...

- These scores for each subject are optimally weighted composites of the original data, designed to capture as much variance as possible with as few scores as possible.
- But for conceptual ease, imagine 5 scores: P1, N1, P2, N2, P3 amplitude

PCA (3): The Loading matrix (to guess what components mean)

$$L_{mxn} =$$

$$\begin{array}{ll} \text{Component \#1} & [l_0,\ l_1,\ l_2,\ \ldots,\ l_{n-1} \\ \text{Component \#2} & \ \ l_0,\ l_1,\ l_2,\ \ldots,\ l_{n-1} \\ \text{Component \#3} & \ \ l_0,\ l_1,\ l_2,\ \ldots,\ l_{n-1} \\ \ldots & \ \ \ \ \ldots \\ \text{Component \#m} & \ \ l_0,\ l_1,\ l_2,\ \ldots,\ l_{n-1}] \end{array}$$

Where m = Number of components
n = Number sample points per average
l = component loading for time point 0, 1, ...

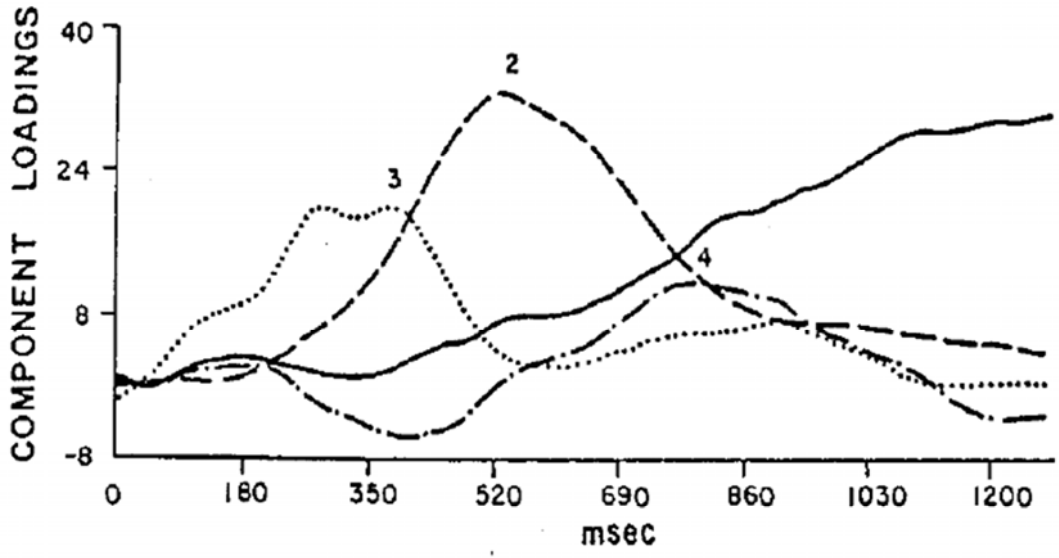
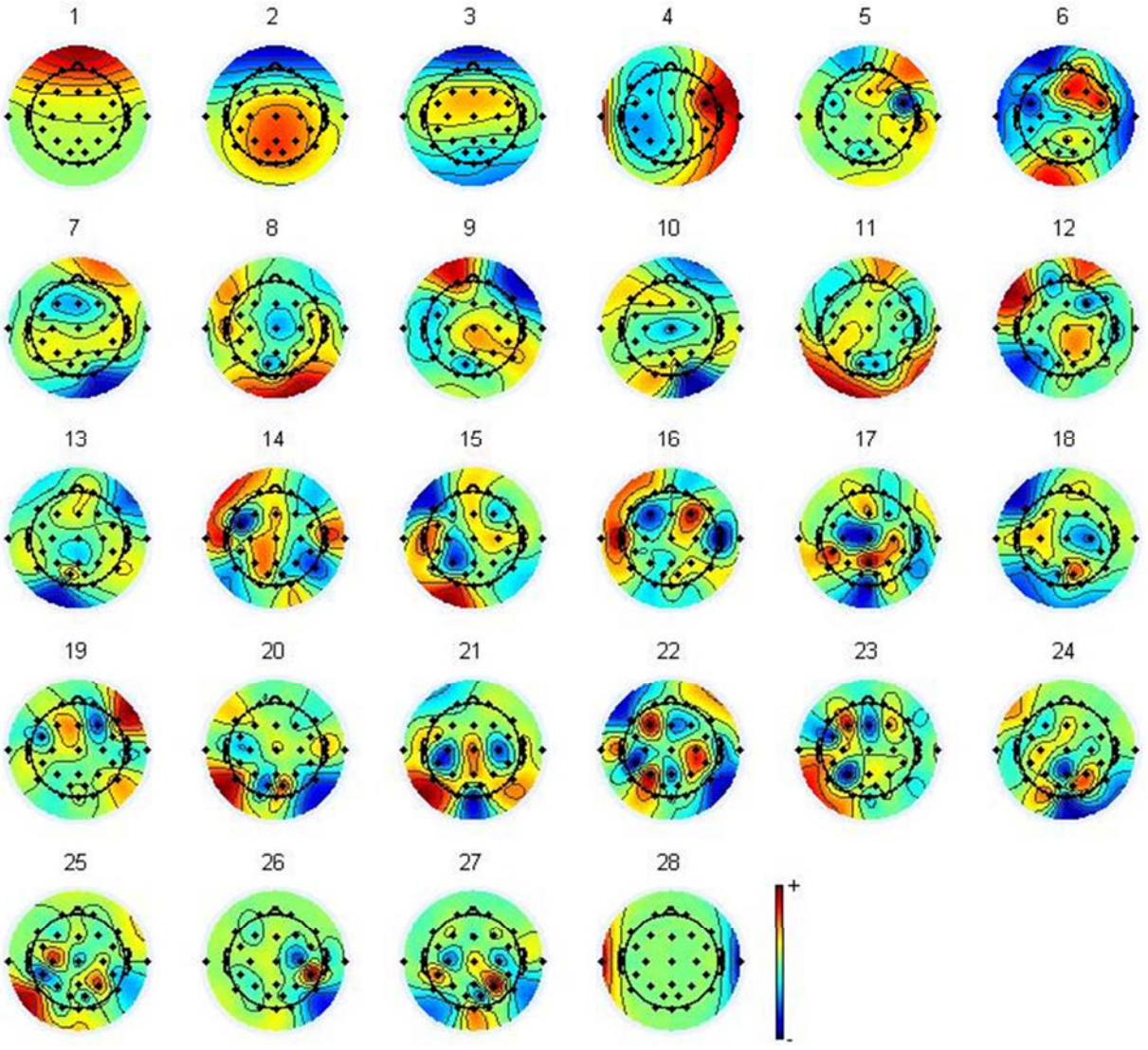
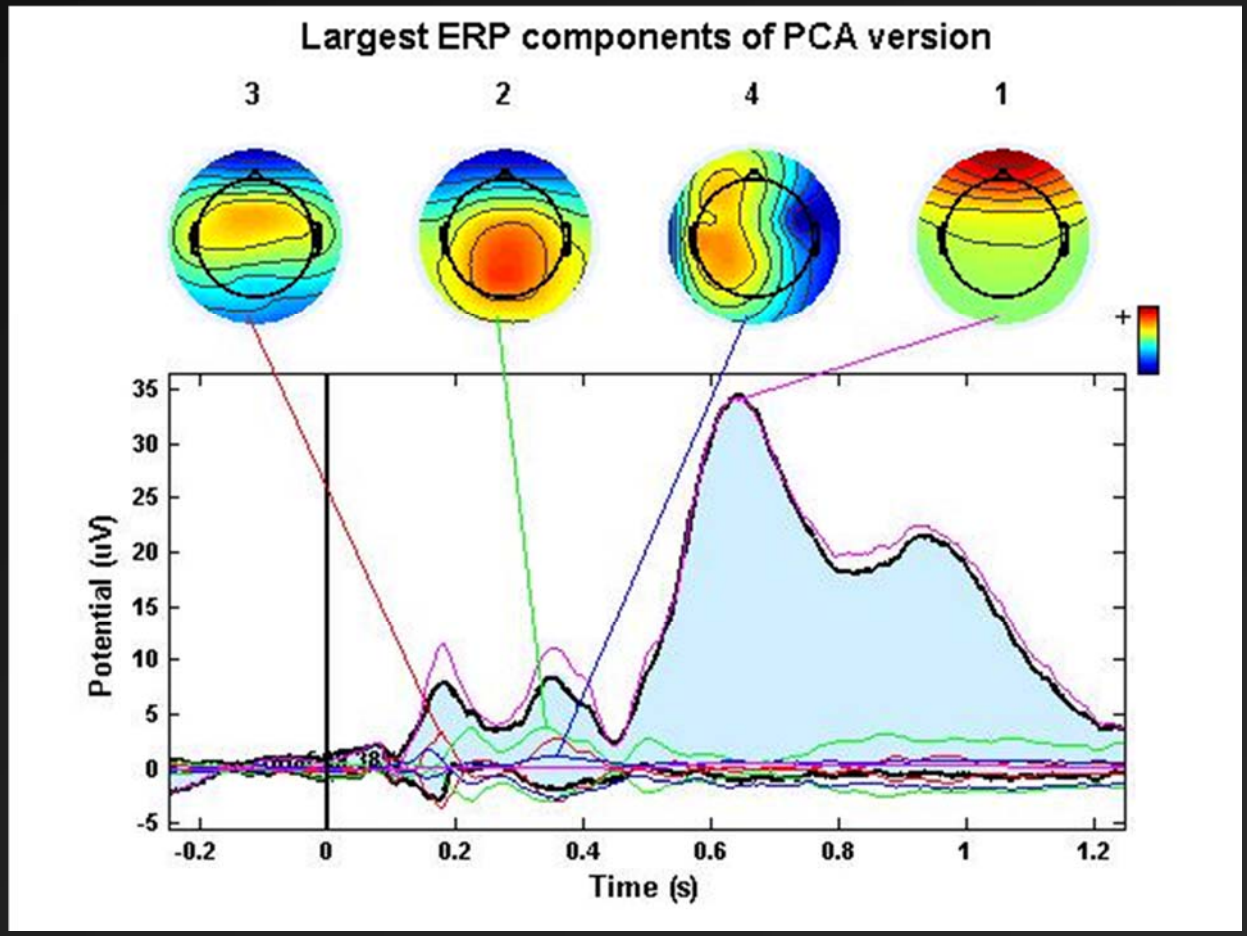


Figure 10-4. Plot of four sets of component loadings derived from a principal-components analysis (PCA) of an ERP data set. Each of the component loading vectors is composed of 128 points corresponding to 128 time points (100-Hz digitizing rate) in the waveforms.

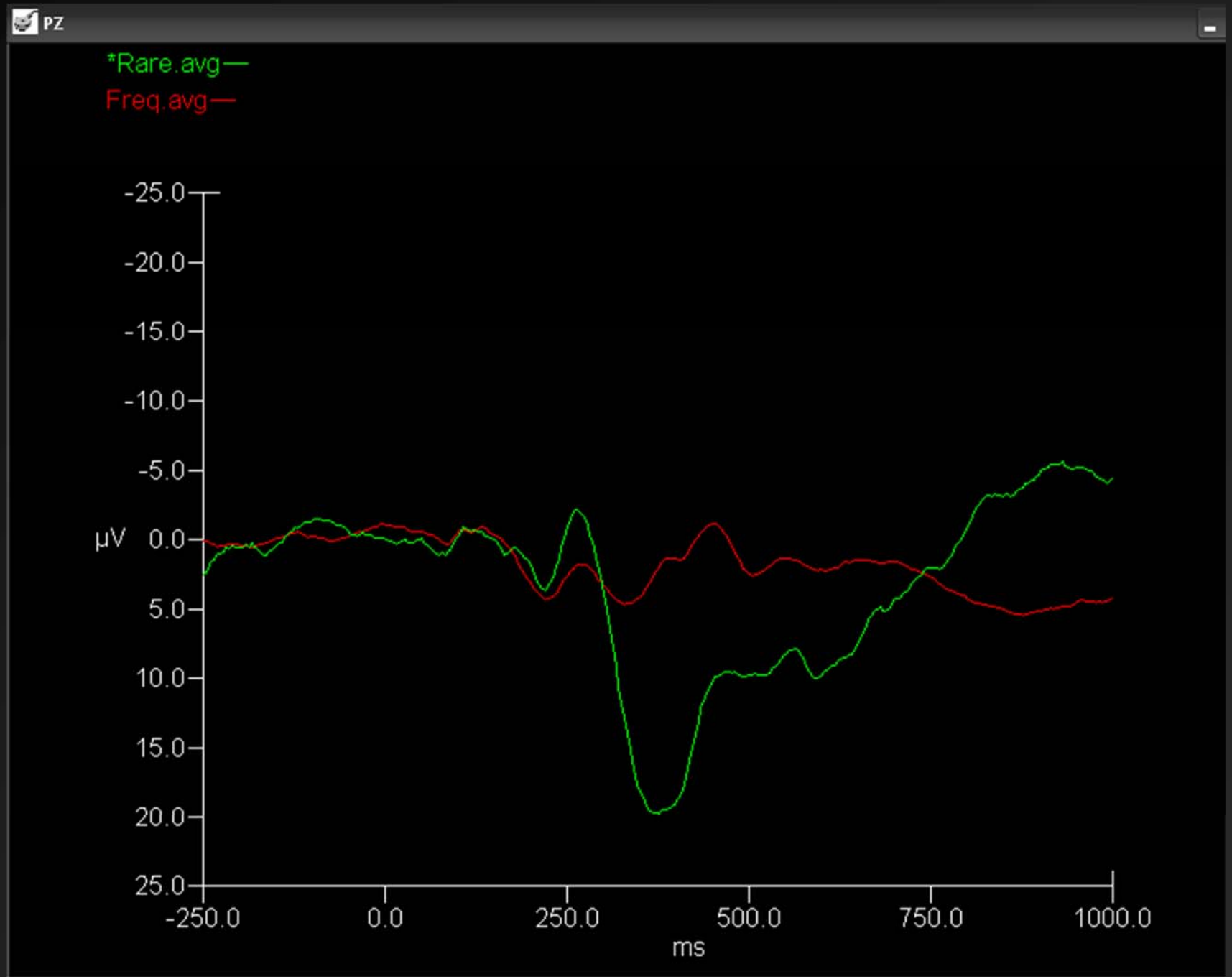
PCA on Sample Data



PCA (3b): The Loading Map (for Spatial PCA)



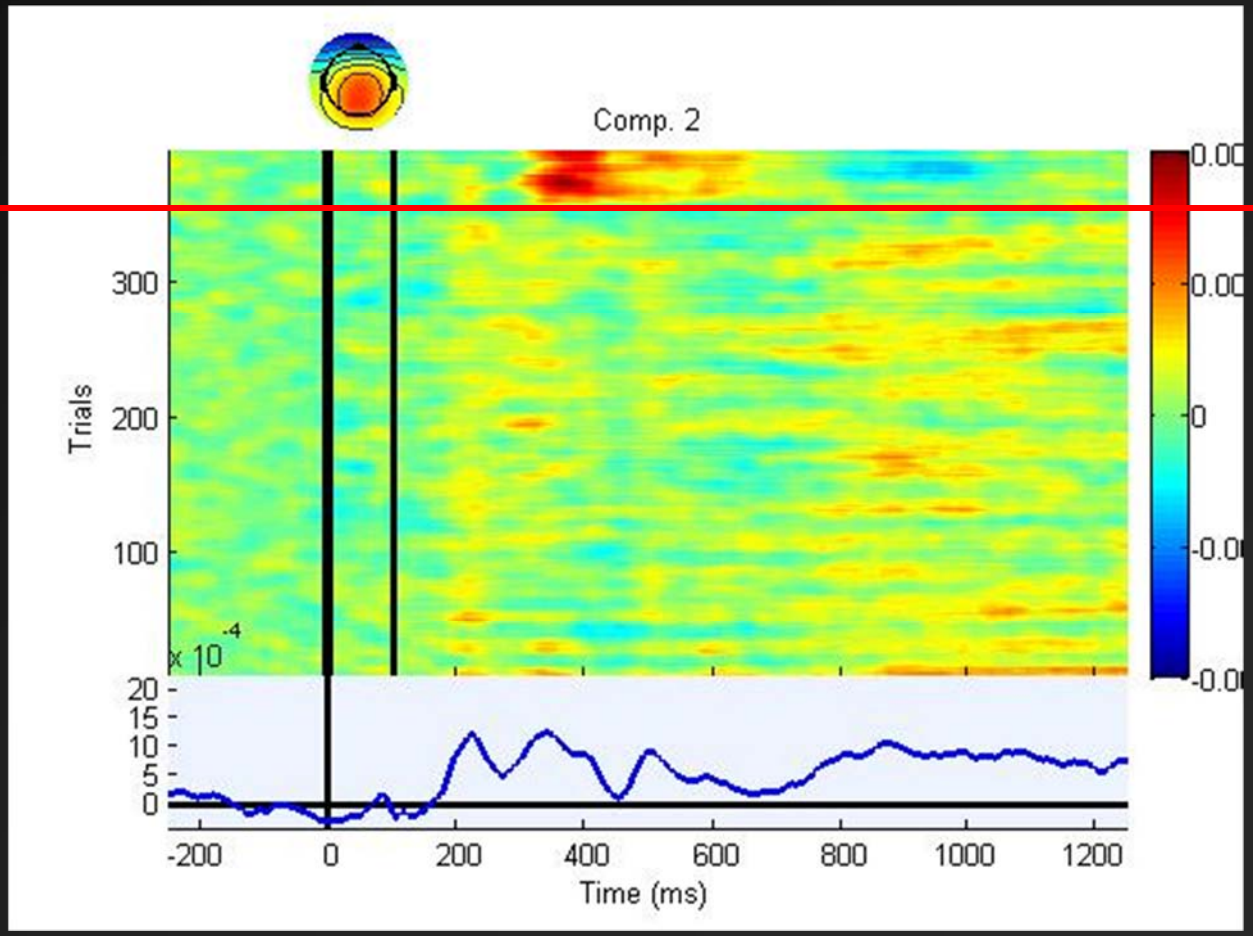
Reminder: The ERP from which it derives



PCA Component 2

Rare

Frequent



PCA (4): Reconstructing Data Matrix

- $\mathbf{D}_{Nxn} \sim= \mathbf{S}_{Nxm} * \mathbf{L}_{mxn}$
- This reconstructed Data matrix will differ slightly from the original Data matrix because not all n components are used.
- To the extent that the m components account for most of the variance in the original data set, the reconstructed data matrix will closely approximate the original data matrix.

PCA (4): Caveat Emptor

- PCA is a linear model; assumes the components sum together without interaction to produce the actual waveform
- Sources of variance are orthogonal; if two sources are highly correlated, may result in a composite PCA component reflecting both
- Component invariability in terms of latency jitter across subjects
  - PCA does not distinguish between variations in amplitude vs variations in latency
  - Especially a problem in comparing control vs pathological groups; pathological groups will typically be more variable
  - Allen & Collins unpublished simulation study:
    - Two groups: Control & Pathological
    - Identical waveforms for each group differed only in latency
    - The two groups differed significantly on three of four principal component scores
    - In other words, if one indiscriminately interprets these as amplitude or morphology differences, one would be WRONG!!!

ICA … better Spatial PCA

- PCA finds orthogonal components
  - First PC accounts for most variance
  - Next PC accounts for most remaining variance
  - Components will have orthogonal scalp distributions
- ICA separates temporally independent components
  - Also known as blind source separation
  - May or may not correspond to brain “hotspots” but do represent functional brain networks
- See:
  http://www.sccn.ucsd.edu/~scott/tutorial/icafaq.html
  http://sccn.ucsd.edu/~arno/ (ICA for Dummies!)

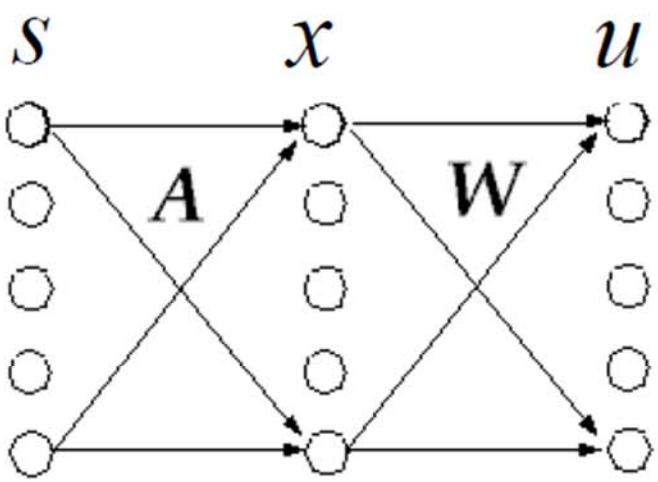
Independent Component Analysis

ICA is a method to recover a version, of the original sources by multiplying the data by a unmixing matrix,

u= Wx,

where x is our observed signals, a linear mixtures of sources,

x= As.



WA after learning:

-4.09	0.13	0.09	-0.07	-0.01
0.07	-2.92	0.00	0.02	-0.06
0.02	-0.02	-0.06	-0.08	-2.20
0.02	0.03	0.00	1.97	0.02
-0.07	0.14	-3.50	-0.01	0.04

While PCA simply decorrelates the outputs (using an orthogonal matrix **W**), ICA attempts to make the outputs **statistically independent**, while placing no constraints on the matrix **W**

From Tzyy-Ping Jung , presented at EEGLab Workshop, Nov 8,2007

$$U = WX$$

ICA activity — Data

Data X

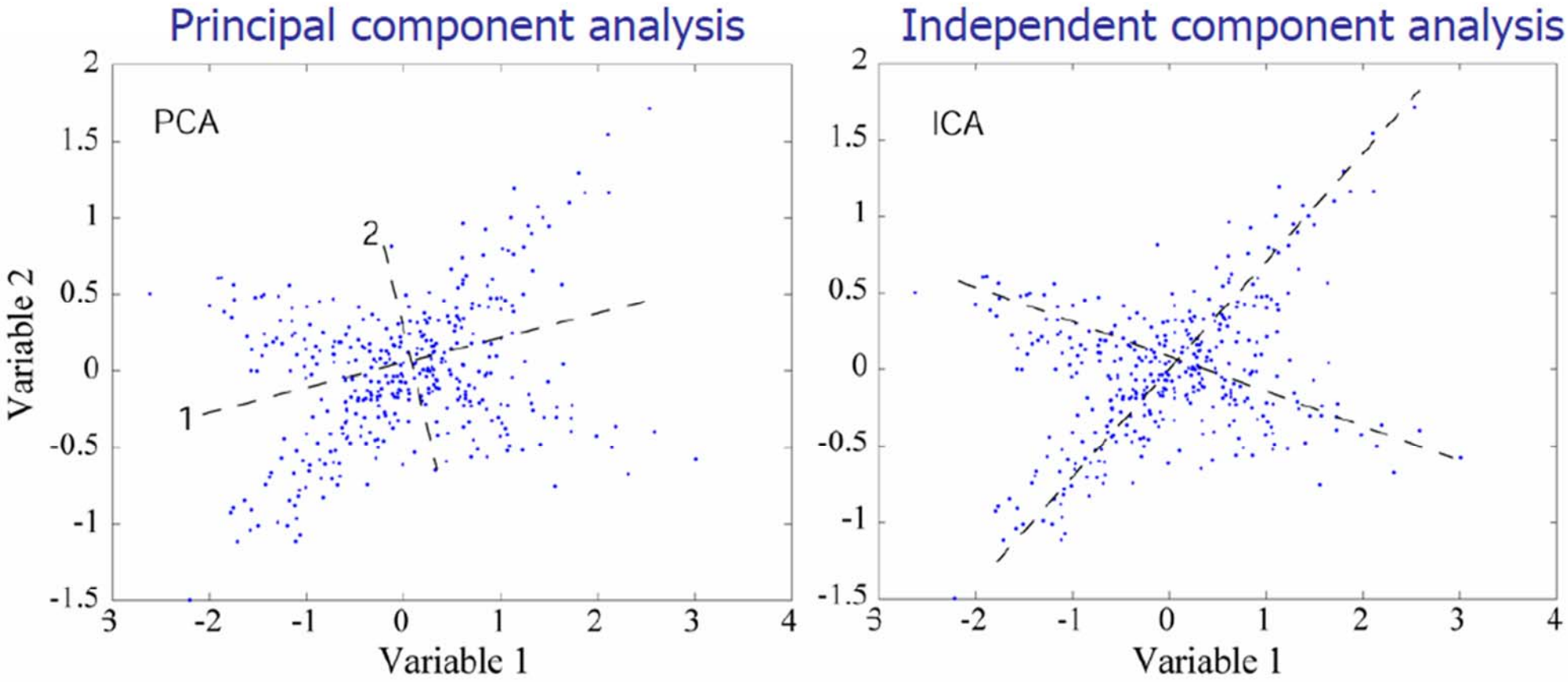
$$\begin{bmatrix} 3 & 2 & 5 & 4 & 3 & 2 \\ 0 & -2 & -5 & -1 & 1 & -1 \\ -1 & 2 & 0 & 1 & 0 & -3 \end{bmatrix} \begin{matrix} \\ \\ \\ \end{matrix} \begin{matrix} \leftarrow \text{Channel 1} \\ \leftarrow \text{Channel 2} \\ \leftarrow \text{Channel 3} \end{matrix}$$

$$\begin{bmatrix} 5 & 3 & -2 \\ 1 & 2 & 4 \\ 0 & -1 & 3 \end{bmatrix} \rightarrow \begin{bmatrix} 3*5+0*3-1*(-2) & 2*5+(-2)*3+2*(-2) & \dots \\ 3*1+0*2-1*4 & 2*1+(-2)*2+2*4 & \dots \\ 5*1-5*2+0*4 & 5*1-5*2+0*4 & \dots \end{bmatrix} \begin{matrix} \leftarrow \text{Comp. 1} \\ \leftarrow \text{Comp. 2} \\ \leftarrow \text{Comp. 3} \end{matrix}$$

Weight matrix W

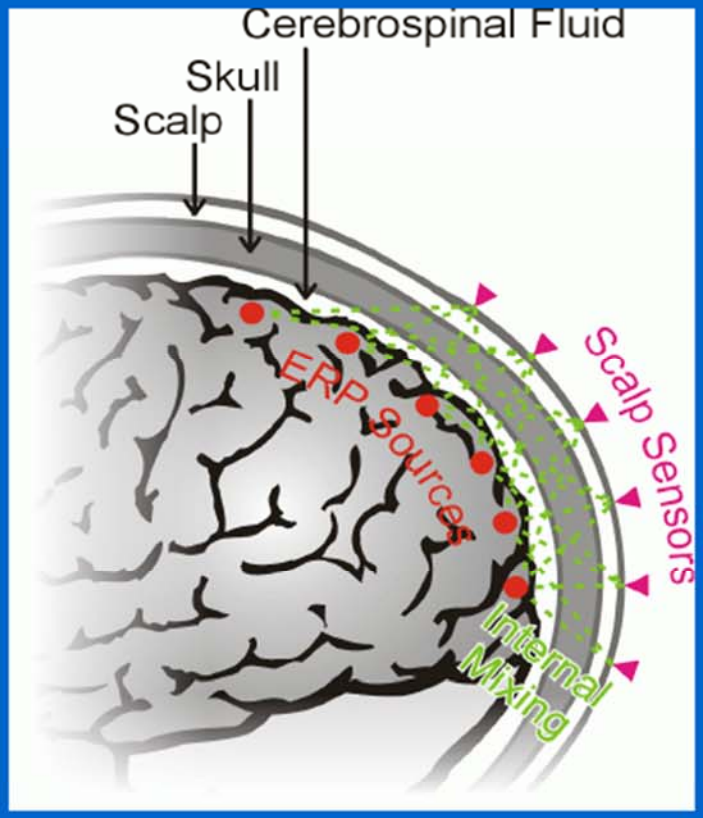
ICA activity U

ICA vs PCA

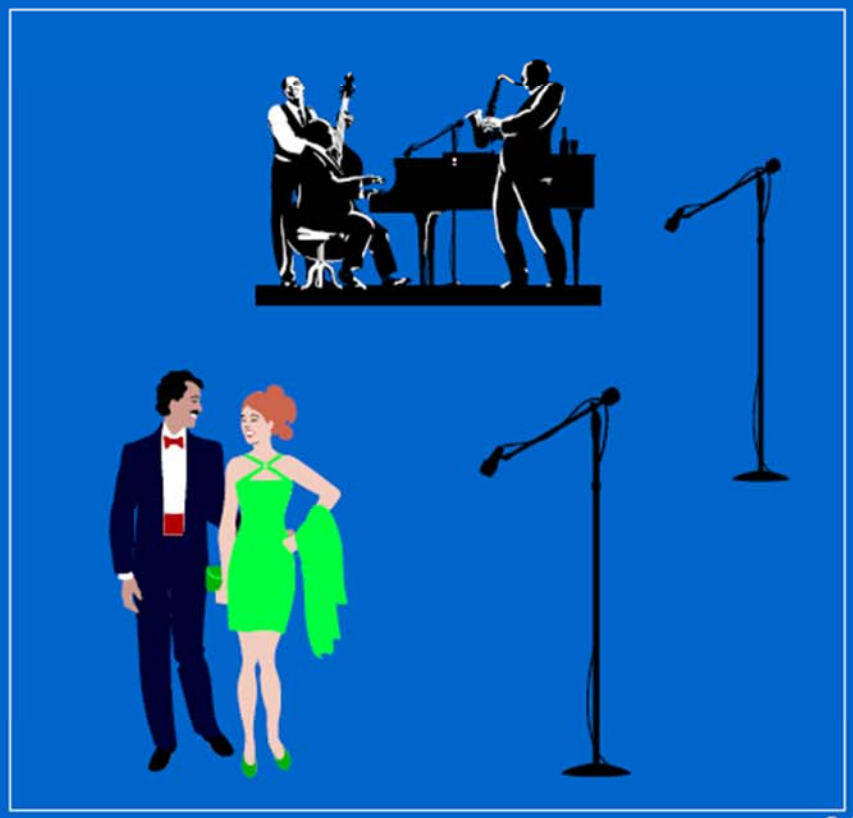


From Tzyy-Ping Jung , presented at EEGLab Workshop, Nov 8,2007

EEG data are mixtures of source signals



Cocktail Party



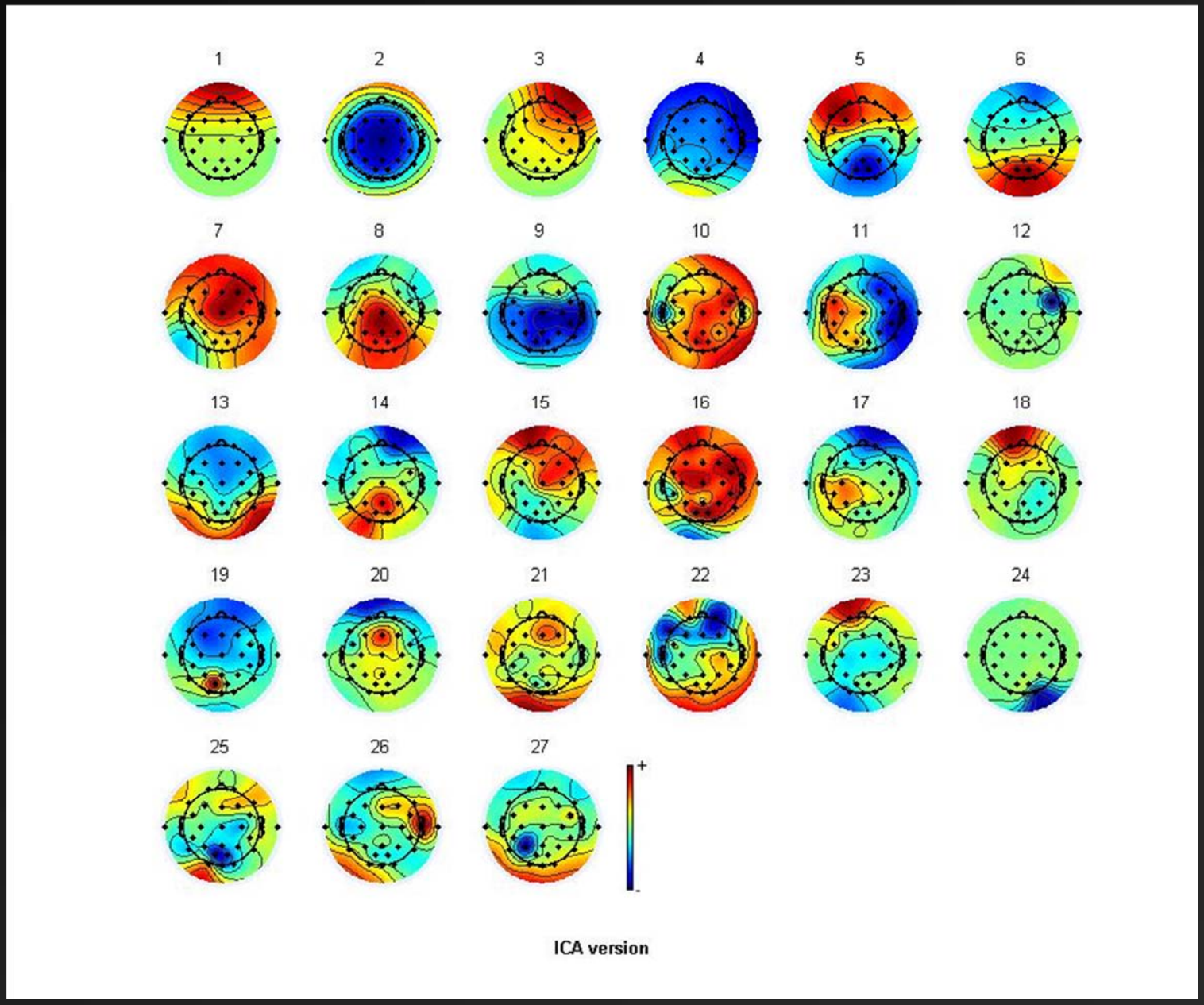
From Tzyy-Ping Jung , presented at EEGLab Workshop, Nov 8,2007

ICA/EEG Assumptions

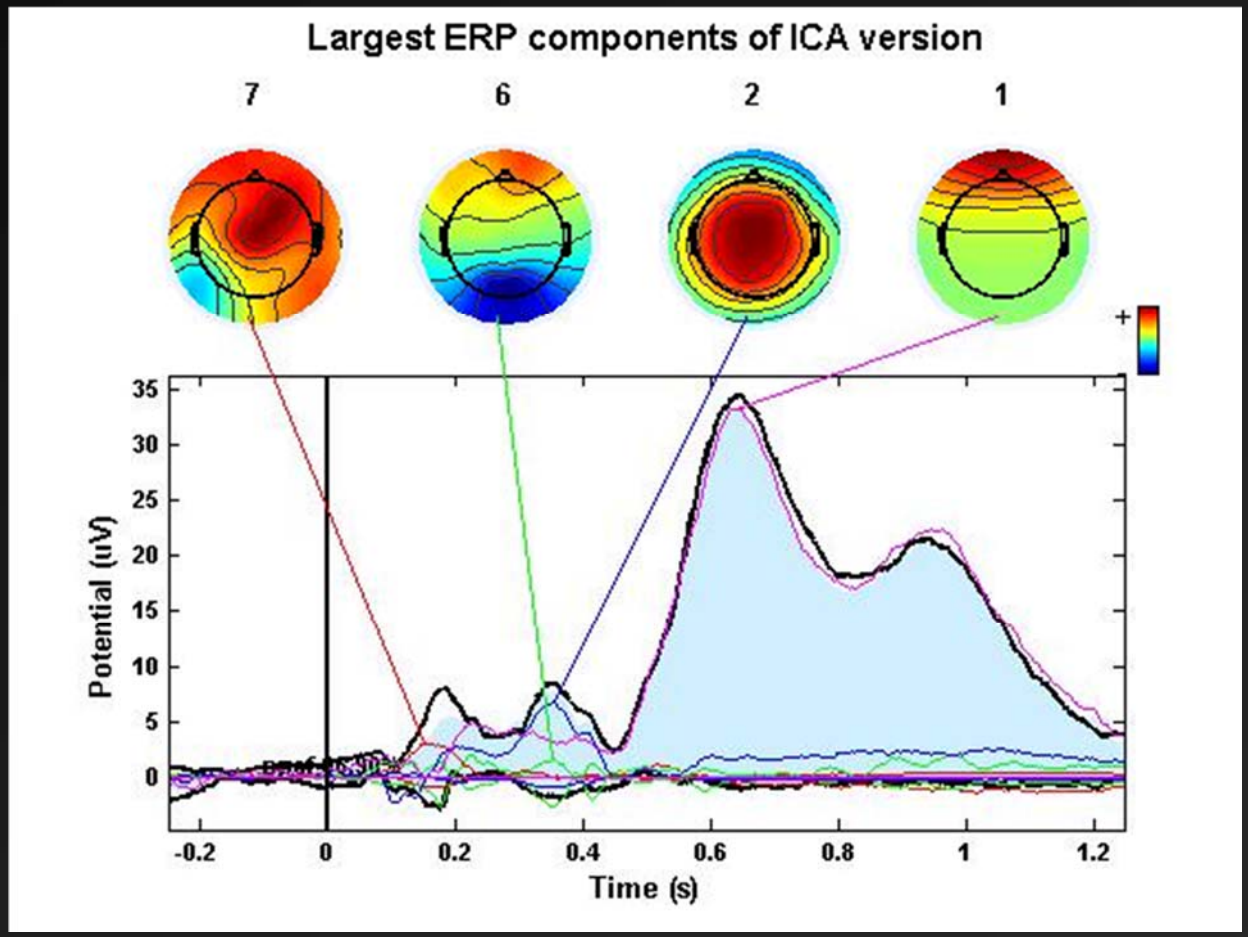
- Mixing is linear at electrodes
- Propagation delays are negligible
- Component time courses are independent
- Number of components $\leq$ number of channels.

From Tzyy-Ping Jung , presented at EEGLab Workshop, Nov 8,2007

ICA: The Projection Map



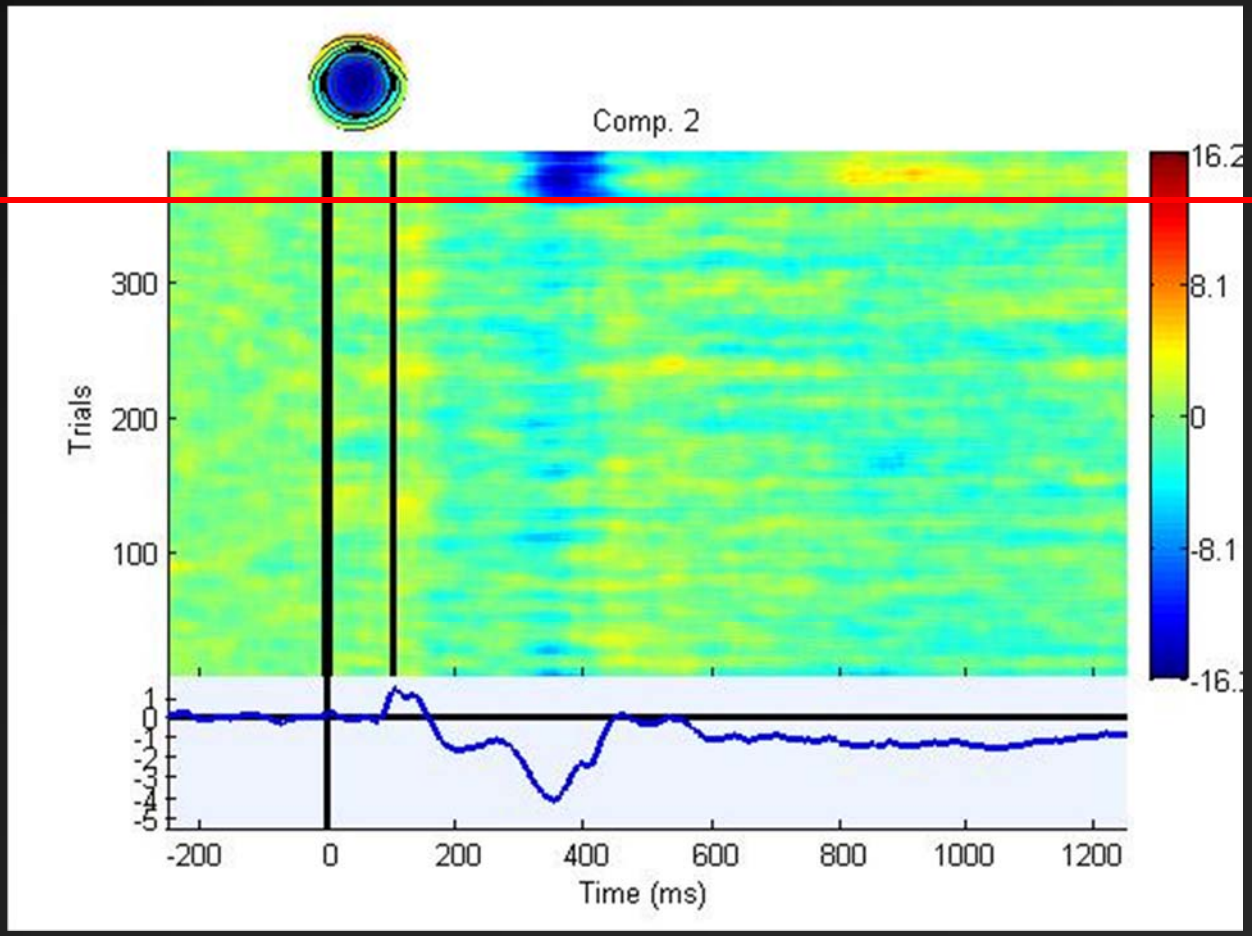
ICA: The Projection Map



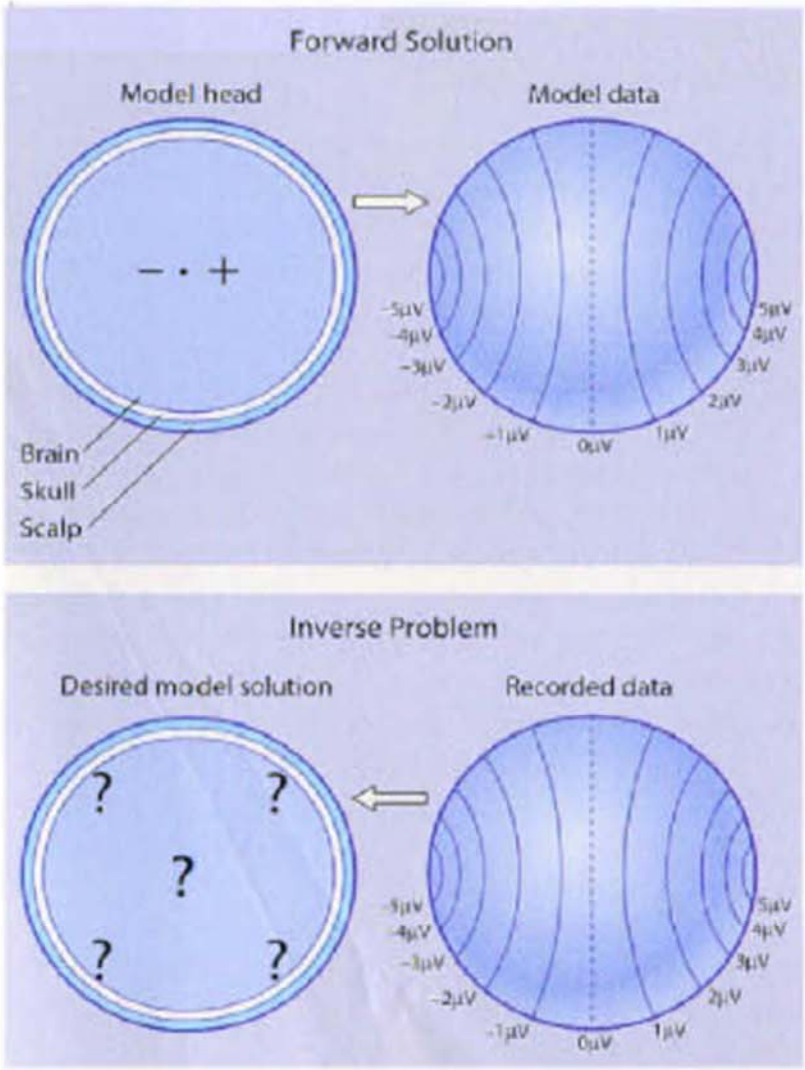
ICA: Trial by Trial IC Projection to Pz

Rare

Frequent



Inverse solution is not unique



A single pattern of neural activity will produce a unique scalp map

BUT …A single scalp map could have been produced by an infinite number of patterns of neural activity

From Tzyy-Ping Jung , presented at EEGLab Workshop, Nov 8,2007

Source Analysis

- BESA -- Brain Electrical Source Analysis
- This is a model-fitting procedure for estimating intracranial sources underlying ERPs
  - Estimate -- if model fits, then data are consistent with these sources; yet there is no unique solution
  - Not for ongoing EEG -- too many sources

BESA

- Imagine a data matrix of ERPs:

$\mathbf{V_{Cxn}}$ (# Channels by # timepoints)

- Note that this is really the result of the subtraction of the activity at the reference from the activity at the these sites; i.e.,

$\mathbf{V_{Cxn} = U_{Cxn} - R_{Cxn}}$

- Note: the reference matrix has identical rows! Thus BESA Presumes that all channels referenced to the same reference!

BESA

- Reconstruct a data matrix that includes not only the original channels, but the implicit channel (reference) as well:

$\mathbf{U_{Exn}}$ (# electrodes = # channels+1),

which represents the activity at each electrode with respect to an average reference (i.e., the average of all channels)

BESA

- Now this matrix $\mathbf{U_{Exn}}$ can be decomposed into
  - a set of sources: $\mathbf{S_{Sxn}}$ (# Sources by # timepoints)
  - a set of attenuation coefficients $\mathbf{C_{ExS}}$
  - so that $\mathbf{U_{Exn}} = \mathbf{C_{ExS}}\ \mathbf{S_{Sxn}}$

BESA

- The attenuation matrix is determined by:
  - the geometry between the source and the electrodes
  - the nature of the conductance of the three-layer head model (Brain, Skull, Scalp);
    - the skull is less conductive than the layers on either side
    - this results in a spatial smearing of potentials as they cross the skull
    - the skull produces the equivalent of a brain that is 60% of the radius of the outer scalp (rather than the "true" figure of ~84%)

Next

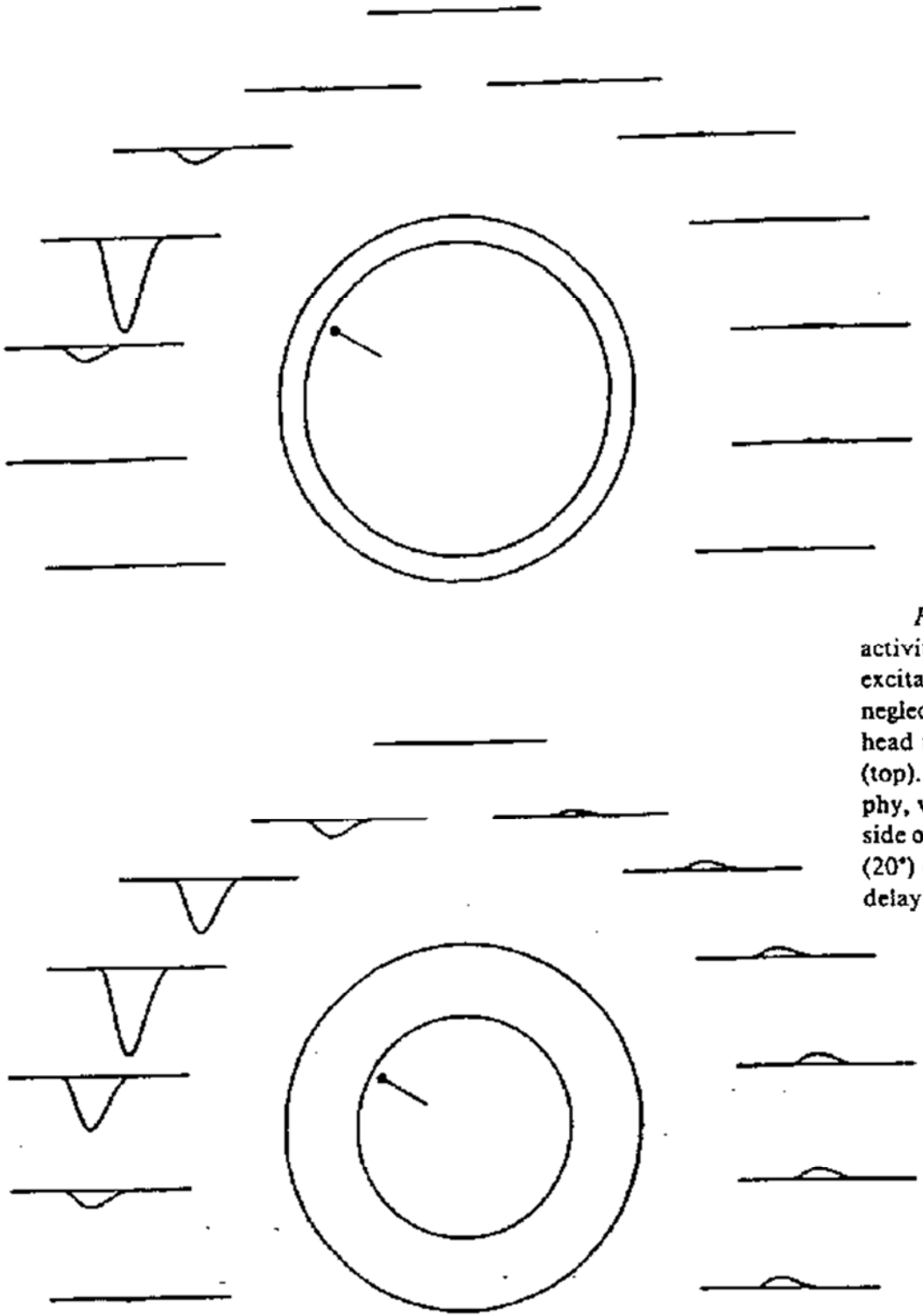
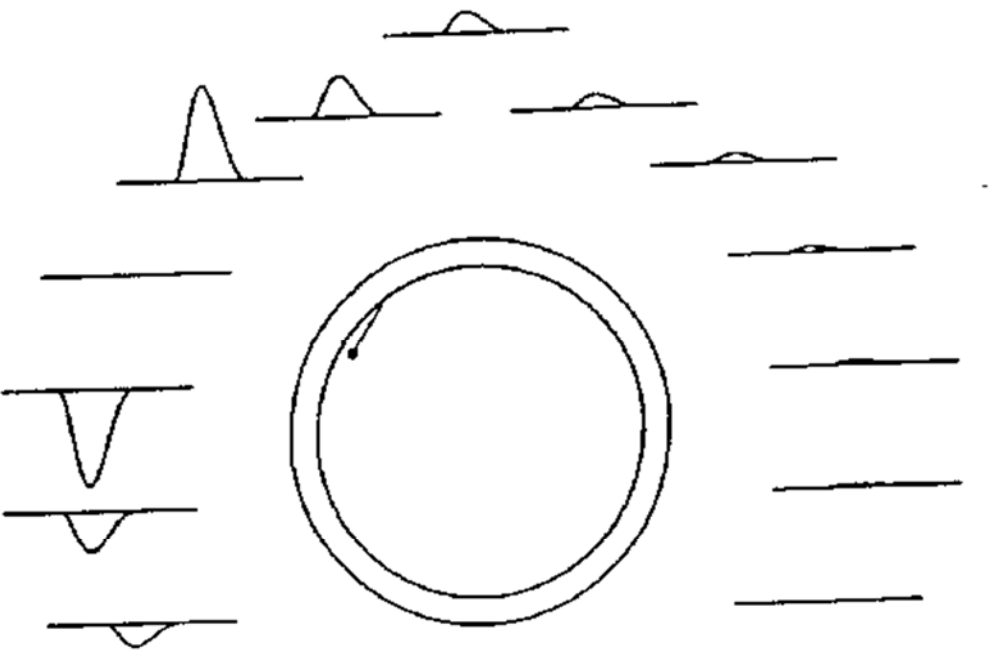


Fig. 4. Coronal scalp potential distribution of a radial equivalent dipole modeling activity of superficial cortex. The dipole is oriented inward to mimic, for example, excitatory pyramidal cell activation at the apical dendrites, producing surface negativity. neglecting the shielding effect, i.e. taking an eccentricity of about 80% in a homogeneous head model, results in a narrow focus, similar to the epicortically recorded topography (top). Adequate reduction of equivalent eccentricity results in a realistic scalp topography, which is much more widespread and exhibits a positive maximum on the opposite side of the sphere (bottom). The simulated waveforms at the vertex (C_z) and at equidistant (20°) electrodes over both hemispheres depict a monophasic activity arising with some delay after stimulus delivery.



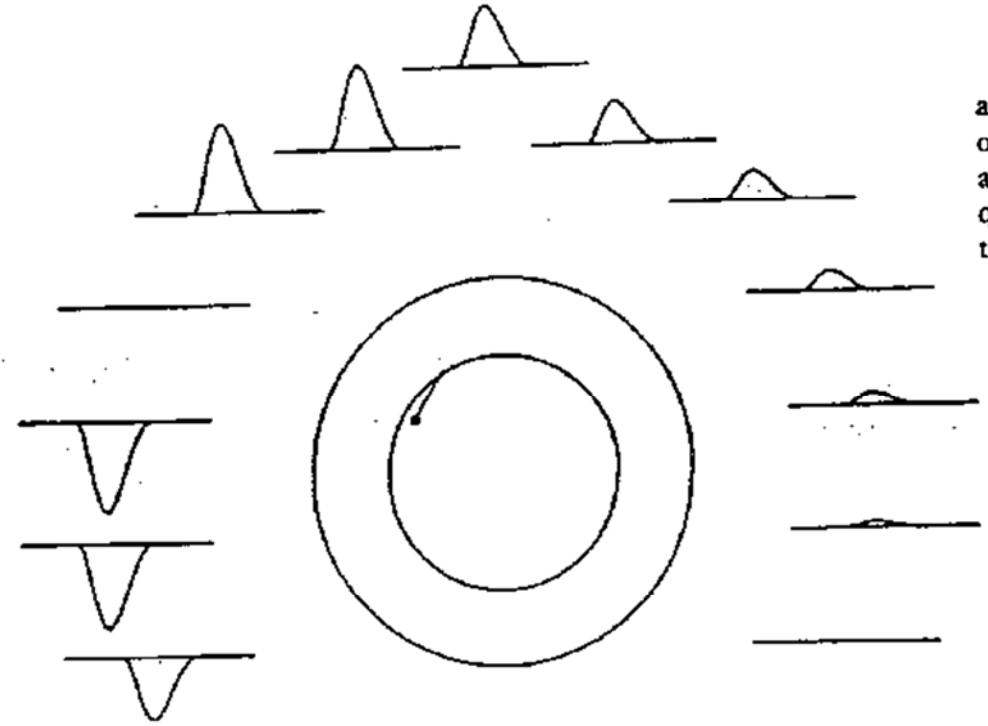
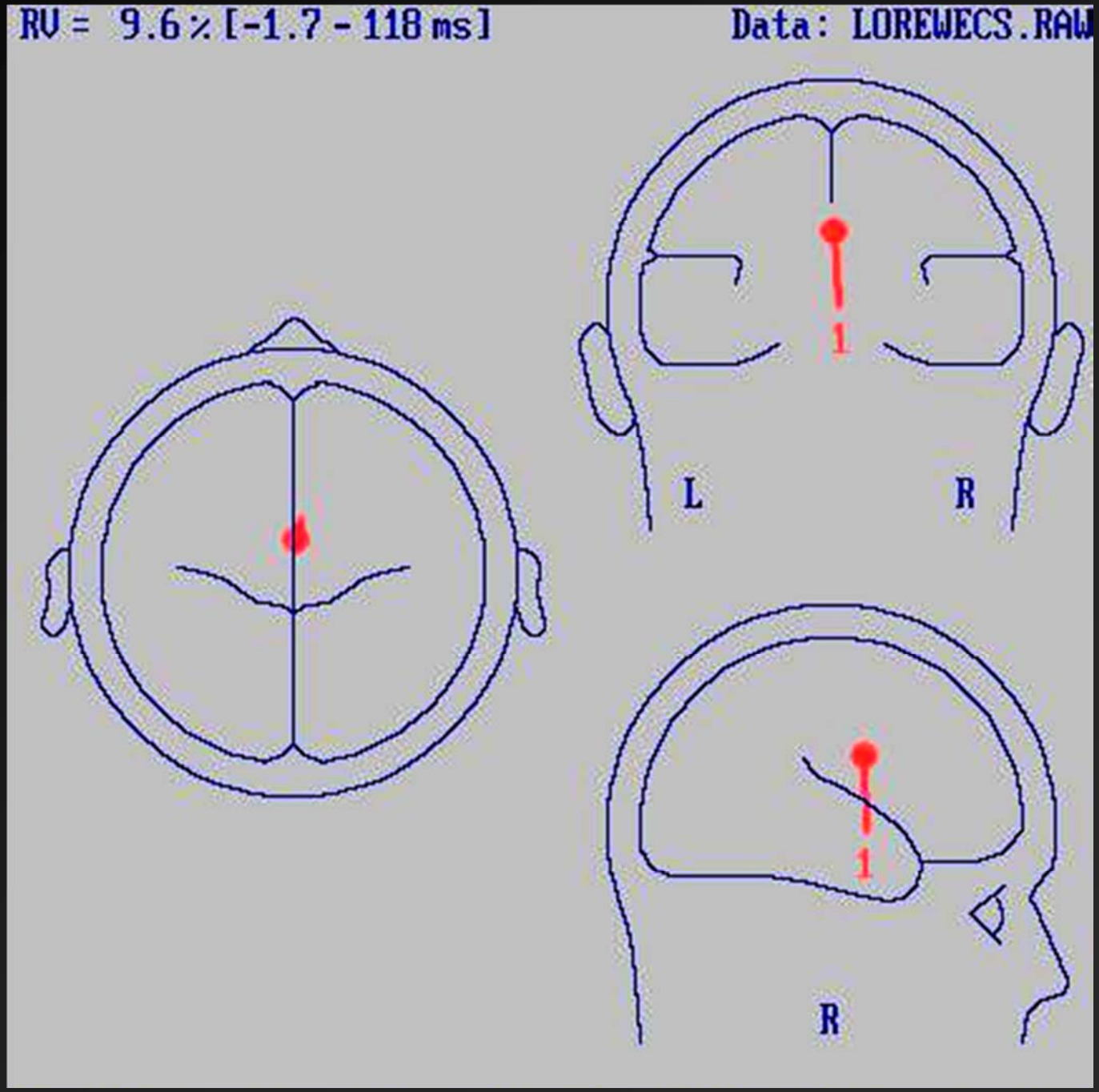


Fig. 5. Coronal scalp distribution of a tangential dipole modeling fissural cortical activity. As explained for figure 4, the correctly transformed ecentricity in the homogeneous head model (bottom) results in a realistic scalp topography with widespread positive and negative maxima to either side of the actual location of the source. Note that in the quasistatic approach a single dipole source contributes the same waveform at all electrodes. Only the attenuation factor and the sign vary with electrode site.

BESA

- Note that the decomposition of **U** into **C** and **S** results in
  - an electroanatomical time-independent matrix (**C**) that reflects that anatomical substrates do not move around in the head
  - a time-variant dipole source potential matrix that represents the change in activity of each source over time


RV = 9.6% [-1.7 - 118 ms]
Data: LOREWECS.RAW
1
L
R
1
R

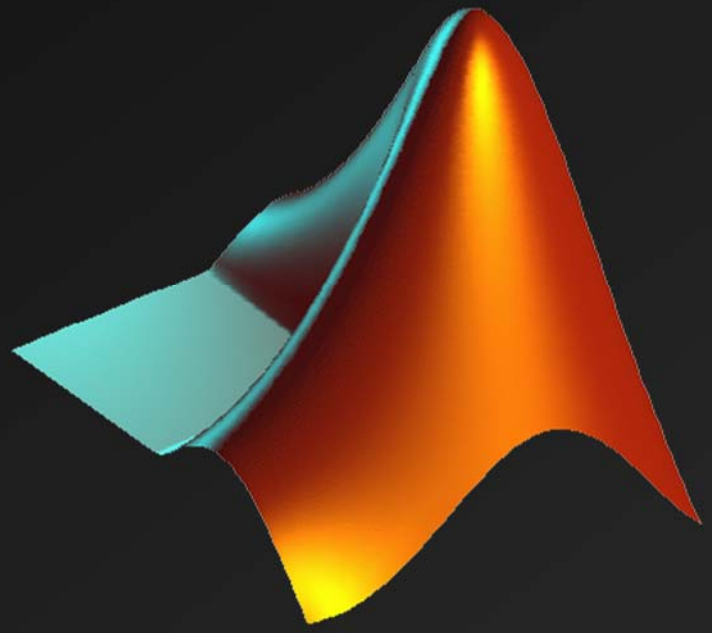
BESA Vs PCA Vs ICA
(a battle of acronyms)

- This decomposition is akin to PCA/ICA
  - PCA and ICA have sources and propagation coefficients
  - PCA solutions are constrained by orthogonality of components, and by those that account for greatest common variance
  - ICA constrained to find temporally independent components
  - BESA solutions are constrained by the geometry of the head, the volume conduction of the dipoles, and the anatomical constraints dictated by the user (e.g., inside the head, symmetrical, not in the ventricles, must not be in the brainstem after a certain point in time, etc...)

BESA Vs PCA Vs ICA continued

- Like PCA/ICA, the reconstruction of the original data set will be imperfect
  - With all methods. better chance of reconstructing the original matrix if data are reliable
  - If you capture the important sources, the reconstruction should be very good (i.e., small residual variance)
  - It is useful to attempt to upset a solution by inserting another source and seeing if:
    - the original solution is stable
    - the new source accounts for any substantial variance
- Can do dipole localization (BESA) on an IC!

Let’s Try!



Implementations

- BESA can be used:
  - in a strict hypothesis-testing manner by designating sources a priori and testing the fit
  - in an exploratory/optimizing manner by allowing the program to iteratively minimize the residual variance (between observed and reconstructed waveforms) by:
    - moving dipoles
    - changing the orientation of dipoles
    - altering the time-by-activity function of the dipoles

BESA – Did it work?

- In the end, the adequacy of your solution will be judged by
  - stability of your solution:
    - against insertion of additional dipoles
    - across multiple subjects
  - anatomical feasibility
  - follow-up tests with patients with lesions
  - your reviewers!

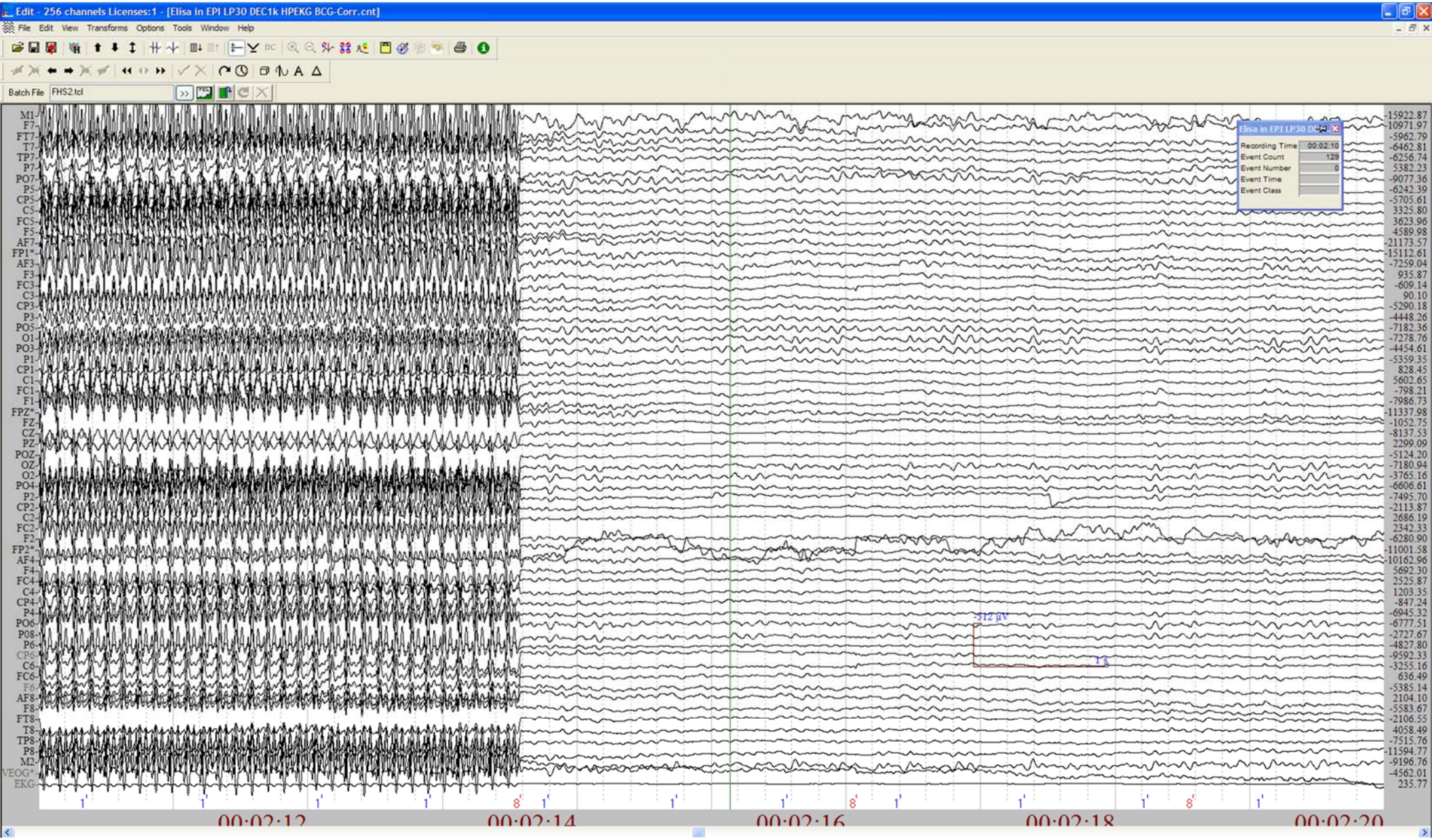
Recording EEG in fMRI environments: Oodles of Issues

- EEG can be bad for fMRI
  - Wires and electrodes can be ferromagnetic = TROUBLE
  - Wires and electrodes can be paramagnetic = less trouble
  - Solution: Non-paramagetic cap
- MRI and fMRI can be bad for EEG
  - RF pulse creates huge artifact for EEG
  - Movement in Magnetic fields creates current in any conductive medium (e.g. wires!)
  - High frequency current can make wires HOT and RF is 127.68 MHz at 3T – that's fast, and can create mega-hurts!

Next

Carbon fiber Cap

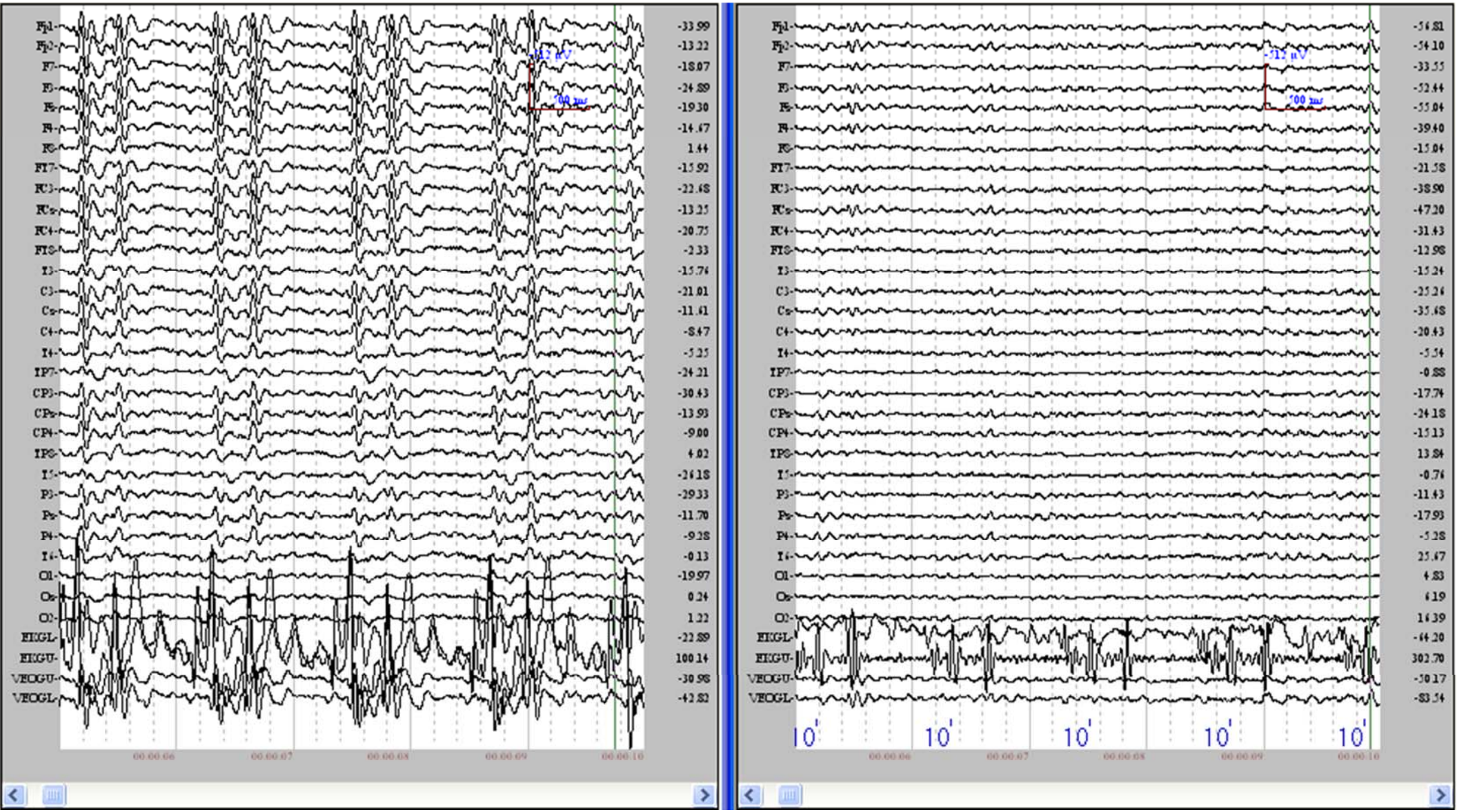
- Conductive
- Will not heat up
- Will not pose hazard in strong magnetic field
- Includes 5Kohm inline resistor to prevent any induced current from reaching the subject
- Includes Styrofoam head at no charge



Spontaneous EEG data obtained from a 3T scanner, with data on the left side shown prior to correction for the rf-pulse, and data on the right reflecting the correction.

By linking the trigger for the rf pulse with the EEG acquisition system, and knowing the rf pulse sequence parameters, software can model and remove the artifact, with the EEG signal preserved despite the large artifact that appears to overwhelm it.

Other artifact: Movement in the Magnetic Field



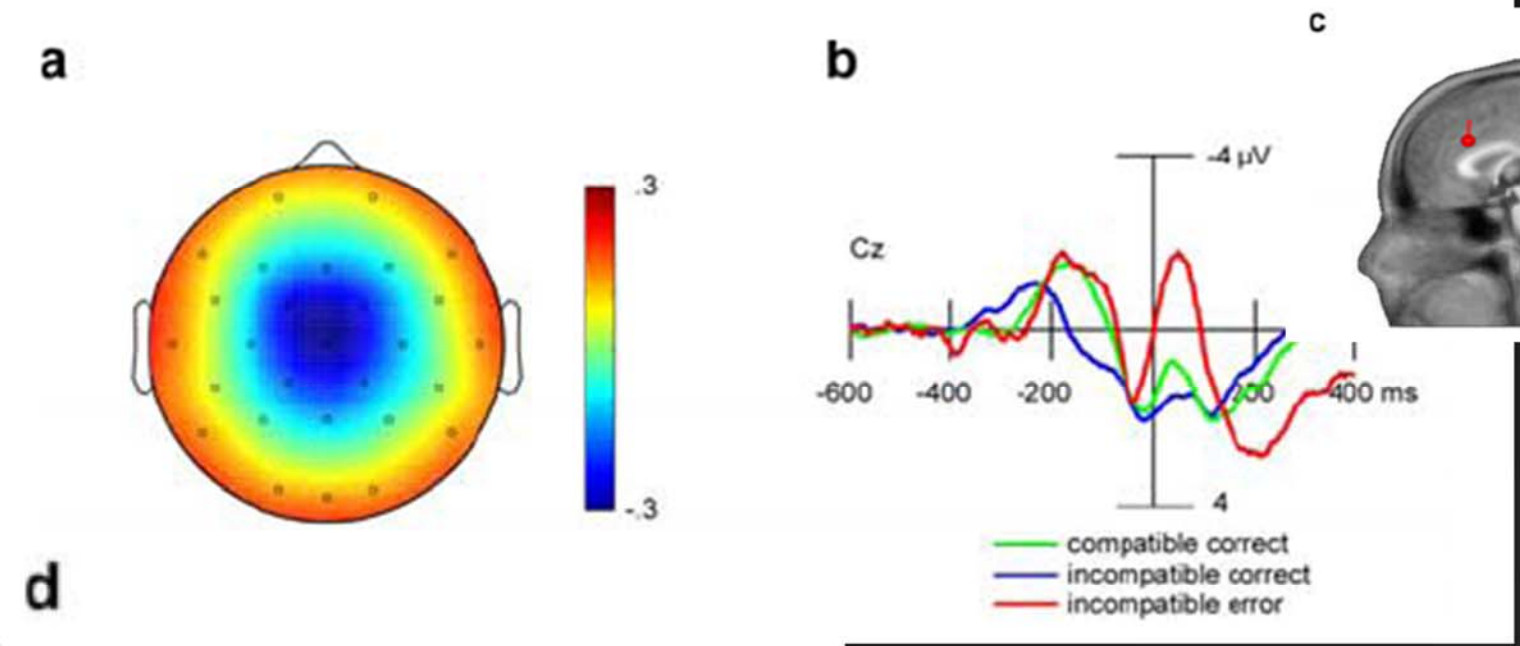
Pulsatile changes in blood flow with each heart beat create motion in the strong magnetic field that induces electrical current. Uncorrected spontaneous EEG data displayed on the left show clear ballistocardiogram artifact. On right, same data following ballistocardiogram artifact reduction. Note uncorrected EKG channel near the bottom of the panel.

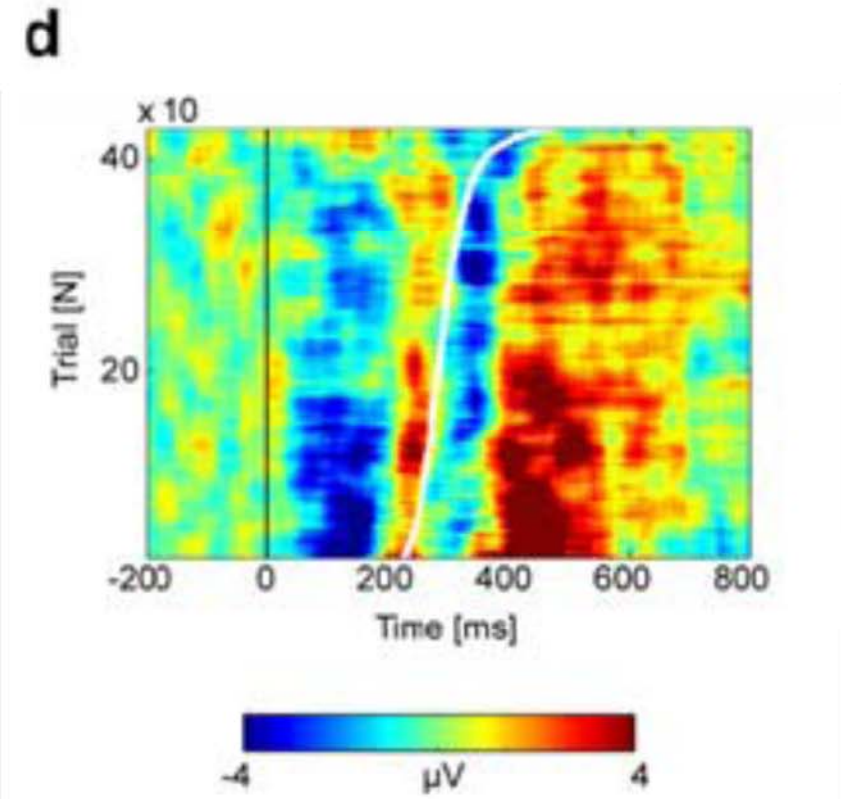
Recording EEG in fMRI environments: Really making use of the two technologies

- Could easily correlate ERP amplitude with fMRI (BOLD) signal
- This is potentially suboptimal:
  - If done on average, this neglects trial-to-trial fluctuations
  - Confounds between versus within-subject effects
  - Correlation addresses whether *people* with bigger ERP component amplitudes have larger BOLD signal
  - We wish to know whether *variations within people* from trial to trial underlie both ERP and BOLD changes within subjects

ICA of ERN Data:

The IC corresponding to the ERN for three conditions, with dipole model fit





- ICs exist for every raw trial!
- At left is ERP-image plot of IC incompatible error trials at vertex electrode (Cz) aligned to stimulus onset
- Sorting the trials by reaction time visualizes the ERN–reaction time relationship
- ERN is visible, without stimulus locking the trials!

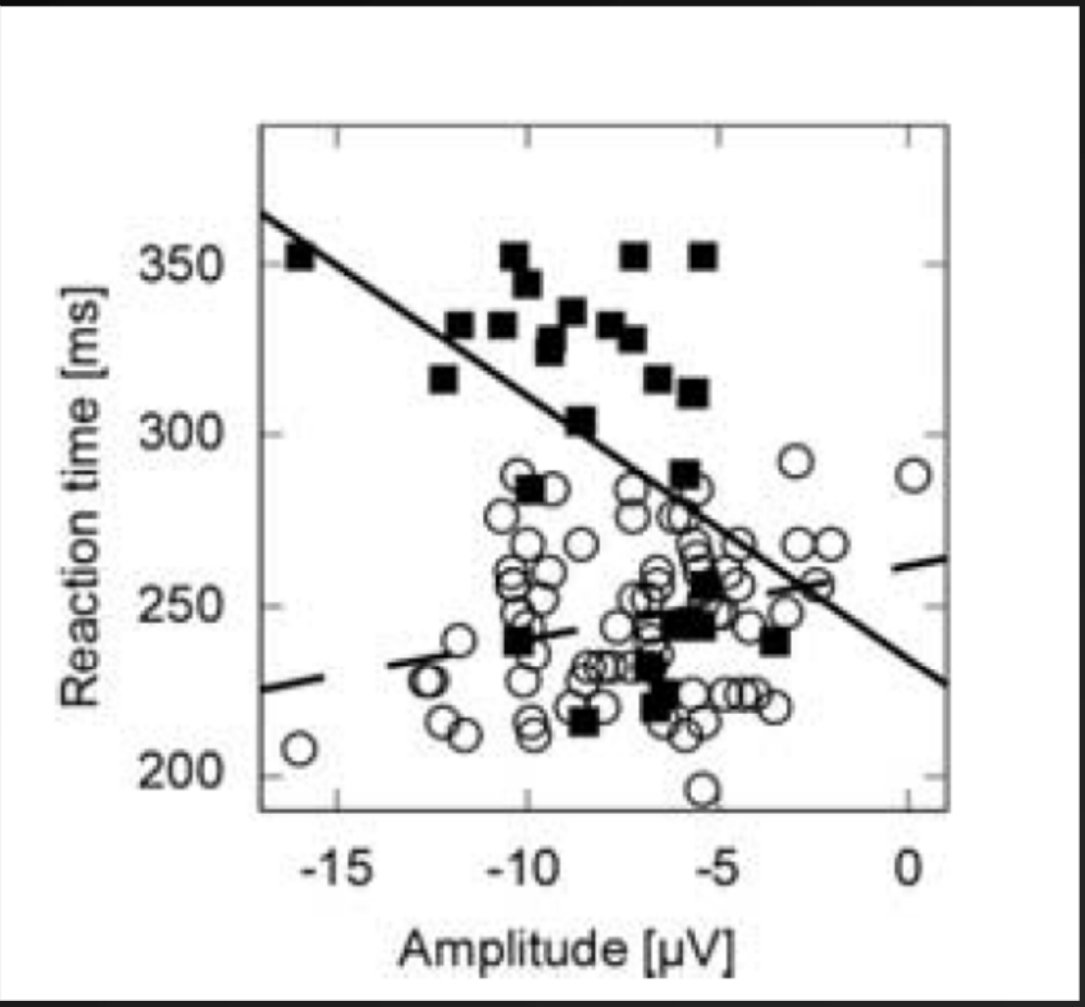
Debener, Ullsperger et al *J Neurosci* 2006

ICA on ERP with fRMI!

- Single-trial error-related negativity of the EEG is systematically related to behavior in the subsequent trial
- This trial-by-trial EEG measure of performance monitoring predicted the fMRI activity in the rostral cingulate zone (aka ACC!)

Debener, Ullsperger et al *J Neurosci* 2006

Single Trial ERN IC related to trial-to-trial variations in behavior!



- Single-subject example
- Incompatible error condition
- Relationship between single-trial IC amplitude and reaction time, separately for the current trial (open circles; dashed regression curve) and for the reaction time of the following trial (filled squares; solid regression curve).

Debener, Ullsperger et al *J Neurosci* 2006

fMRI activations to Errors

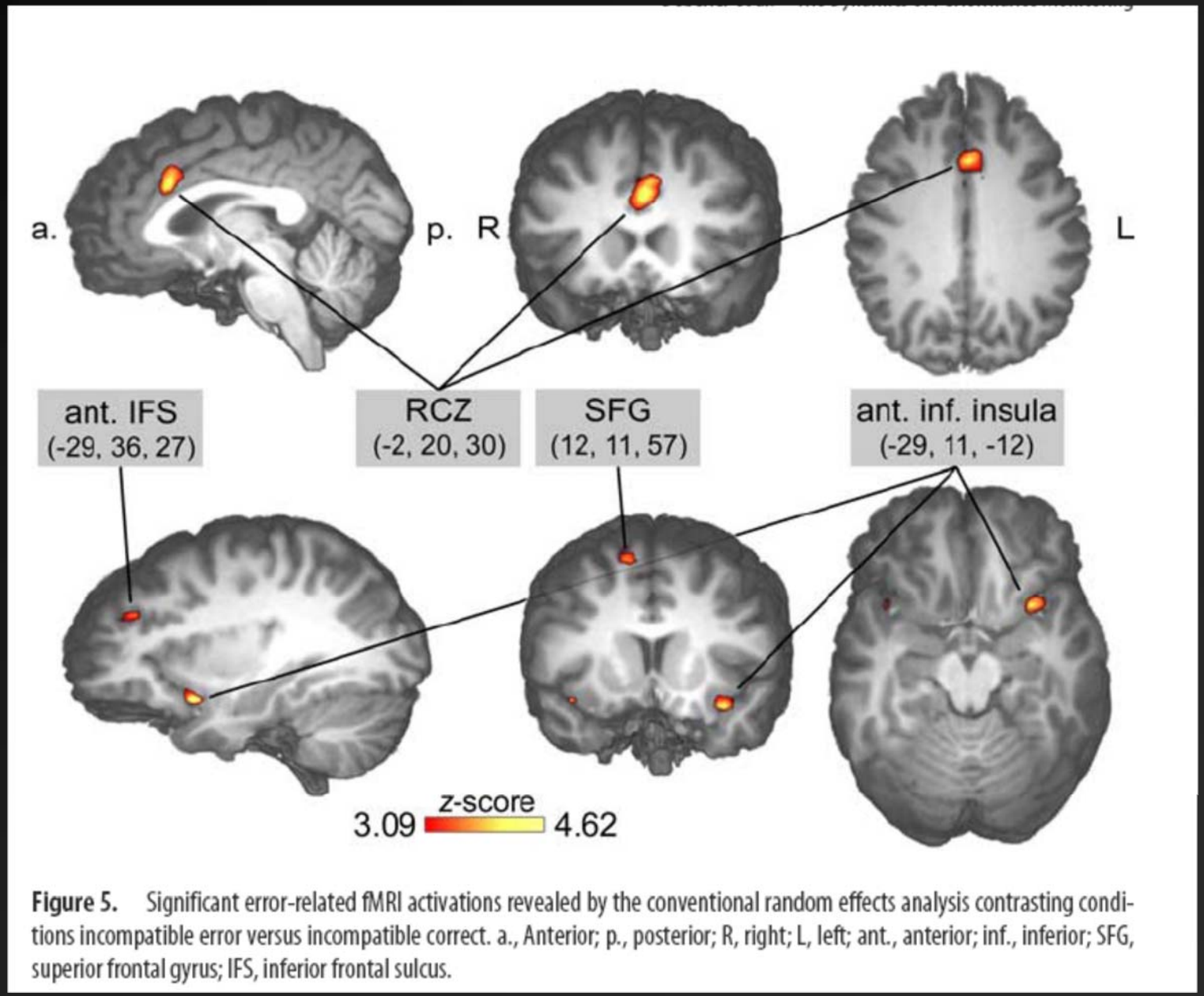


Figure 5. Significant error-related fMRI activations revealed by the conventional random effects analysis contrasting conditions incompatible error versus incompatible correct. a., Anterior; p., posterior; R, right; L, left; ant., anterior; inf., inferior; SFG, superior frontal gyrus; IFS, inferior frontal sulcus.

Debener, Ullsperger et al *J Neurosci* 2006

Regions related to ERN IC activity

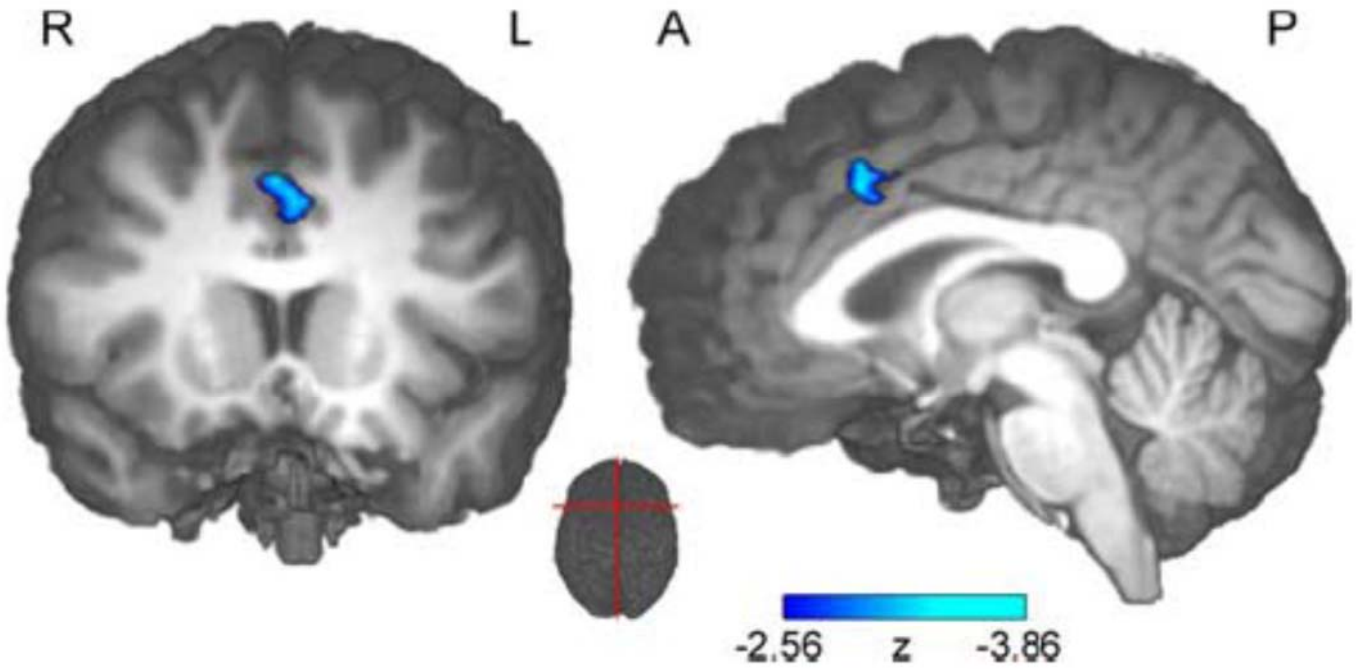


Figure 4. Result of the EEG-informed parametric fMRI analysis based on IC single-trial amplitudes, plotted on an individual brain. fMRI signals correlated with single-trial amplitudes solely in the RCZ along the banks of the cingulate sulcus [center of gravity at coordinates (x, y, z) = 0, 17, 42; $z = -3.86$]. The left part shows coronal view; the right part shows the sagittal view on the right hemisphere. The red lines on the middle top view inset indicate slice sections. R, Right; L, left; A, anterior; P, posterior.

Debener, Ullsperger et al *J Neurosci* 2006

Psychophysiology -- Synopsis

- Psychophysiology is inherently interdisciplinary, and systemic
- Psychophysiology based on dual assumptions (Cacioppo, Tassinary, & Berntson, 2007)
  - Human behavior and experience are embodied and embedded phenomena
  - Physiological responses of brain and body – when studied within the context of an appropriate experimental design – can illuminate aspects of behavior and experience.

Psychophysiology -- Synopsis

- Ultimately we obtain correlates of behavior and experience
  - Psychophysiological Correlates are not privileged; they are no better, no worse, than any other correlate of behavior and experience
- The utility of these correlates – like any correlates in science – hinges upon:
  - good experimental design
  - strong theoretically driven hypothesis testing
  - the development of a nomological net, a set of inter-relationships among tangible measures and constructs that place the findings in a larger theoretical context, and lend construct validity to the measures and findings