DATA MANAGEMENT & TYPES OF ANALYSES OFTEN USED

Dennis L. Molfese
University of Nebraska - Lincoln
DATA MANAGEMENT

- Backups
- Storage
- Identification
- Analyses
Data Analysis

- Pre-processing
- Statistical Analysis
- Written report
Backup at End of Test

• Always:
  • check data files immediately after completion of the test
  • make 2 backup copies immediately
  • check backup copies immediately for errors
  • Store in 2 different locations
How To Analyze Electrodes

- Pre-selected sites
- All sites
- Problems
  - Formidable at 32+ electrodes
  - Electrode site variance (Mocks Correction)
  - Interpreting Hemisphere Differences
  - Interpreting Anterior vs. Posterior effects
Channel Clustering

Curran (1999)

Fig. 1. Sensor locations on the 128-channel Geodesic Sensor Net. The approximate sensor locations were projected onto a 3-dimensional head model from which these 2-dimensional images were taken. Sensors appear more closely spaced at the edges because depth is lost in the 2-dimensional images, but actual electrode spacing is approximately equidistant throughout. Different symbols are used to denote channels within each of the 8 spatial regions used in ANOVAs. The tables define each symbol along with the abbreviations used for each region. Midline electrodes are denoted with diamond-shaped symbols. VR = vertex reference.
Channel Clustering

Infant Clusters (6)

Adult Clusters (5)
Data Analysis

"I think you should be more explicit here in step two."
More Popular Data Analysis Approaches

- Peak amplitude/latency measures
- Area under the curve
- PCA-ANOVA
- ICA
- Source localization (BESA)
- Photogrammetry + fMRI
Analysis Approaches

Filtering

Superimposition
(Dawson, 1947)

Averaging based on mean or median
(DAWSON, 1951)

Filtering rediscovered

Single trial analyses
(John, 1967)

Adaptive filters (Woody, 1967)

Peak amplitude & latency

Frequency analyses (FFT - Fourier)

Time-series

Area under curve

String Method

Multivariate approaches
- PCA
- ICA

Growth Curve

Lisrel modeling

BESA + fMRI + Photogrammetry
Analyses based on 2 points (stimulus or peak onset & target peak latency)

Baseline & average of multiple points around latency point of interest
Woody Filters (1967)

- brain recognizes as SIGNALS the neural events which are either not locked to a stimulus or have variable waveforms.
- Cross-correlate each data sample with some "template" waveform.
  - correlation coefficient is maximal for time shift at which the template and data samples are MOST similar.
  - by knowing these time shifts, the data can then be lined up in time PRIOR to sampling.
  - Better resolution can be gained by replacing the template with this new average and repeating the process until there are no differences between the template and the generated average.

Gupta & Molfese Filtering

- ~ Woody Filtering but
- Line up each peak, stretching/contracting waveform
- IEEE Transactions on Biomedical Engineering, 1996, 43, No. 4, 348-356.
Multivariate Analyses

ANOVA on successive data points by conditions

Discriminate function

PCA (Donchin, 1968; 1969)

PCA - discriminate function (Chapman, 1979)

PCA - ANOVA (Molfese, 1975; 1976)

Cluster Analysis (Ruchkin; John et al, 1975)

PCA - MANOVA (Samar, 1986)

PCA - ANOVA - REGRESSION (Molfese & Molfese, 1986)

PCA - ANOVA - Canonical Correlation (Molfese & Molfese, 1979)

Growth curves (Espy, Molfese & Molfese, 2004)
PCA

• Data reduction tool
  • Reduce 1,000,000 ERPs to 4-7 factors
  • Identifies regions of variability
  • Subsequent analyses determine whether variability changes systematically (ANOVA, MANOVA, Discriminant Function, Regression, GC, Modeling)
Principal Component Analysis (PCA)

- Like other ERP measures, assume ERP regions (peaks, slopes) reflect DISCRETE neural events
- PCA arrives at number of independent factors MINIMALLY required to describe any one waveform.
- Shows strong inter-lab replications
Principal Component Analysis (PCA)

- Some disadvantages comparable to other approaches:
  - various factors do not necessarily have anatomical or physiological bases
  - misallocated variance for adjacent peaks
  - requires some statistical expertise
PCA

Strong data reduction tool
Blind to subjects and conditions
Identifies areas of high variability = factors
Various strategies for selecting # of factors:
Scree test (Cattell, 1966)
Select factors accounting for a % of total variance
Eigen value criterion (> 1.0)
Varimax rotated factor scores used as DVs for later analyses
PCA

Decisions regarding approaches
Covariance matrix vs. correlational matrix
Number of factors to resolve
To rotate or not to rotate factors
Inclusion of baseline data or not
Digitizing window length
Sampling rate
Masking by electrode effects (% variance)
PCA

- PCA and ICA only analysis techniques that allow reconstruction of original ERPs from end analyses.
- High replication across labs.
Rotation of Factors

Regardless of factors constructed, exact configuration of factor structure NOT unique.

Many statistical equivalent ways to define the underlying assumptions of the same data set.

Not all solutions equally meaningful or interpretable.

Must choose rotation method that best meets theoretical & practical needs
PCA + ANOVA

CONDITION: CC CI, TIME=14:55:49, DATE=10/02/00, MAG=3, S=1, TOTAL GAIN=40000
H-SCALE=2, STARTING POINT=21, ENDING POINT=160, TOTAL POINTS=140

CORRECT WORD

INCORRECT WORD

1

2

Disk file name: CONDITIONS TXT

MSec Calibration = 2.5 mV
PCA Centroid + Factors

12-Year-Old
n=68

Centroid
Factor 1
Factor 2
Factor 3
Factor 4
Factor 5
Factor 6

Total Variance
31%
15%
8.4%
6.6%
5.7%
4.7%

ms 0 100 200 300 400 500 600 700
GAIN SCORES

GAIN SCORES

+0.5  =  

-2.0  =  

+0.5  =  

+2.0  =  

-1.0  =  
PCA Analyses

To Recreate Data from PCA:

Multiple each Factor by Gain Scores
Add all products to Centroid
Recreated ERPs include only variance contributed by the summed Factors to the Centroid.
ICA

- **Independent Component Analysis**
  - single trial analysis
  - removal of artifacts
  - identification of the *number* of sources
    (not *location*)
ICA and Noise removal

Observing signals, assumed to be mixture of underlining source signals

Source signals after ICA decomposition

Noise components can be removed and observational signals can be rebuilt with uncontaminated components
ICA Assumptions

(Jung, Makeig, Westerfiled, et al 2000)

- # of sources is not larger than # of channels
- Data sources are temporarily independent and they interact with each other linearly
- Data sources are spatially stationary and the spatial spread of electric current from sources by volume conduction does not involve significant time delay
- Enough data points (# of data points $\geq$ the square of # of sources)
Advantage of ICA

(Jung, Makeig, Westerfiled, et al 2000)

- Based on the characteristics of the data; requires no knowledge of signals
- Preserve data from all trials
- Preserve data from all scalp channels
- More power to do single-trial analysis & study trial-to-trial variations.
Limitation of ICA

(Jung, Makeig, Westerfiled, et al 2000)

- N to N; cannot solve when number of sources is larger than N
- Sources in the brain may not be independent
- Source(s) in the brain may not be spatially stationary
- A sufficient amount of data is needed (Delorme, &. Makeig, 2004)
ICA Toolbox—EEGLAB
ERP Source Analyses
Source Analysis

- ERPs do not identify sources
- Use Brain Electromagnetic Source Analysis (BESA) to determine sub-scalp sources of the observed electrical activity in the brain
- Inverse Problem

Prerequisites:
- High-density spatial sampling
- Hypotheses about source locations
- High quality data
Inverse Problem

- Sources within brain generates current recorded at scalp electrodes
- Problem - how to use scalp recordings to identify where within the brain the currents originate
Dipoles

a) Dipole - theory of ERP generation.

b) Dipoles perpendicular to surface (since cortex folds, not necessarily perpendicular).

c) Reflect differences in soma and dendrite ion flow across cortical layers.

d) Models activity.

e) Activity at scalp not necessarily result of ion movements below electrode.

f) Caution: Dipoles generated in one hemisphere may generate higher shifts in other hemisphere.
Dipoles

- Likely positioned in cortex
- Positive pole on one side and negative pole on other side
- Nearness to scalp indicated by side of current spread at scalp
  - Small spread - close to scalp
  - Large spread - far from scalp
Source analysis with BESA 2000

EEG Review
Process raw data

Artifact detection
and correction

Define conditions
and average

3D coordinates
MRI coregistration

Cortical current
density mapping

Multiple source
scan and fit

Calculate source
activity waveforms

Use MRI /fMRI to
test / seed sources
High GRTR Scores
2-dipole Model 200 ms

Match

Mismatch
Dipole Analysis Strategies

- Use neurophysiology & neuropsychology information to place dipoles.
- Need to decide on # of dipoles
- Use % of variance or residual variance as guide
- Developmental caution
- Tasks must be tightly designed
Source Localization Limits for Pediatric Populations

- Baby and Child Brains morphologically and structurally different from Adult Brains
- Developmental issues (neurogenesis, myelination changes)
Source Localization Limits for Pediatric Populations

Temporal Lobe
Source Localization Limits for Pediatric Populations
Source Localization Limits for Pediatric Populations
Source Localization Limits for Pediatric Populations

Infant Source Solutions Models Based on Adult Volume Conduction, Skull Thickness, Impedance, Integrity and Brain/Skull Relationships Are Limited.
ERP Source Analyses
ERP Source Analyses
ERP Source Analyses
ERP Source Analyses
ERP Source Analyses

![Inverse Matrix Editor: Tucker SLORETA](image)

- **Source-Imaging Spec Name:** Tucker SLORETA
- **Forward Head Model**
  - Sun-Stok 4-Shell Sphere
  - Finite Difference Model
- **Inverse Model**
  - Constraint: SLORETA
  - Regularization: Tikhonov
    - Value: 1 x 10^-2
ERP Source Analyses
ERP Source Analyses
ERP Source Analyses
ERP Source Analyses

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control.nsw</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thu, Sep 23, 2010, 3:13:00 PM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calculate measure for regions marked with event fav1, extending region 500 ms after event and 750 ms after event.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean amplitude of individual channels in the group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>favor1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My New Regions of Interest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L Anterior Cing R Anterior Cing L Cingulate Gy R Cingulate Gyrus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERCyberball_3</td>
<td>1.21834</td>
<td>1.03924</td>
<td>0.22826</td>
<td>0.31522</td>
<td></td>
</tr>
<tr>
<td>ERCyberball_3</td>
<td>0.75086</td>
<td>0.96786</td>
<td>0.7455</td>
<td>0.7078</td>
<td></td>
</tr>
<tr>
<td>Cyberball_302</td>
<td>0.72798</td>
<td>0.63217</td>
<td>0.2698</td>
<td>0.19303</td>
<td></td>
</tr>
<tr>
<td>Cyberball_302</td>
<td>0.51497</td>
<td>0.8596</td>
<td>0.61726</td>
<td>0.65459</td>
<td></td>
</tr>
<tr>
<td>childcyberball</td>
<td>1.30115</td>
<td>1.07712</td>
<td>0.57325</td>
<td>0.48815</td>
<td></td>
</tr>
<tr>
<td>3048m_Cyber</td>
<td>0.79452</td>
<td>0.44782</td>
<td>0.51696</td>
<td>0.35792</td>
<td></td>
</tr>
<tr>
<td>3070m_cyberl</td>
<td>0.92566</td>
<td>0.61692</td>
<td>0.52686</td>
<td>0.49165</td>
<td></td>
</tr>
<tr>
<td>3071f_cyberb</td>
<td>0.95198</td>
<td>0.60417</td>
<td>0.28385</td>
<td>0.36937</td>
<td></td>
</tr>
<tr>
<td>3077m_cyberl</td>
<td>0.53129</td>
<td>0.52029</td>
<td>0.41268</td>
<td>0.41578</td>
<td></td>
</tr>
<tr>
<td>3078m_cyberl</td>
<td>1.54282</td>
<td>0.87209</td>
<td>0.79006</td>
<td>0.85809</td>
<td></td>
</tr>
<tr>
<td>ERCyberball_3</td>
<td>0.80551</td>
<td>0.80725</td>
<td>0.42154</td>
<td>0.35918</td>
<td></td>
</tr>
<tr>
<td>3081f_ERCyb</td>
<td>0.82742</td>
<td>0.83708</td>
<td>0.35529</td>
<td>0.39623</td>
<td></td>
</tr>
<tr>
<td>3082f_ERCyb</td>
<td>0.66937</td>
<td>0.70975</td>
<td>0.37551</td>
<td>0.28563</td>
<td></td>
</tr>
<tr>
<td>3097f_cyberb</td>
<td>1.78567</td>
<td>1.6326</td>
<td>2.65075</td>
<td>2.47149</td>
<td></td>
</tr>
<tr>
<td>3096f_ERCybe</td>
<td>1.4584</td>
<td>1.22511</td>
<td>1.11647</td>
<td>0.85741</td>
<td></td>
</tr>
<tr>
<td>3099m_ERCybe</td>
<td>1.01164</td>
<td>0.84152</td>
<td>0.71082</td>
<td>0.73372</td>
<td></td>
</tr>
</tbody>
</table>
ERP Source Analyses

Data Transferred to SAS/SPSS, etc. for analyses to test for statistical significance
Data Presentation (Pubs)

- Report:
  - Gain setting
  - Sampling rate
  - Filter settings
  - Electrode Reference(s)
  - Artifact rejection rates & strategies
  - Analysis procedure
  - Subject attrition rate
  - Power & Effect size
Questions ???
Data Preprocessing

• Additional filtering (Remember-Refiltering additive)
• Isolate data segments of interest
  • Set pre-stimulus baseline and post-stimulus epoch length
  • Adjust for stimulus offsets
• Review data for artifacts and bad channels
• Remove trials with artifacts
• Average / Re-reference
• Perform baseline correction
• Format & output files to statistical software
Reference


