The Neural Engineering Data Consortium: ‘Déjà Vu All Over Again’[1]

Iyad Obeid, Associate Professor
Joseph Picone, Professor
Department of Electrical and Computer Engineering
Temple University
September 7, 2012

1. "It's déjà vu all over again.” Berra explained that this quote originated when he witnessed Mickey Mantle and Roger Maris repeatedly hit back-to-back home runs in the Yankees' seasons in the early 1960s.

Abstract

It is hard to underestimate the impact the evaluation-driven research paradigm has made on human language technology (HLT). Yet, some fields, including the bioengineering community, continue to operate in a mode where research results cannot be easily verified and performance claims are often overly optimistic. The brain machine research community, in particular, is mired in this type of disorganization at a time when mass media and entrepreneurial interests in these technologies are at an all-time high.

Hence, we are developing, in collaboration with the Linguistic Data Consortium, a center at Temple University focused on the development of resources to advance brain machine interfaces. The center will build on the best practices developed by LDC, and is expected to eventually address a wide range of data needs in the bioengineering community. Our first corpus will be the release of over 10,000 EEG recordings conducted at Temple Hospital in Philadelphia, constituting the largest publicly available corpus of its type in the world. Machine learning technology developed on this data is expected to have both clinical and engineering research impact.

Specious Claims

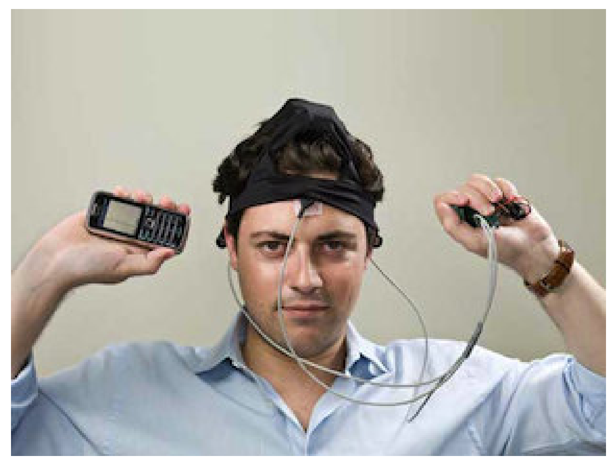
YAHOO! NEWS

iBrain can 'read your mind'; enlists Stephen Hawking

By Eric Pfeiffer | The Sideshow – Mon, Apr 9, 2012

Email | Recommend 1k | Tweet 231 | Share 35 | +1 23 | Print

A team of California scientists have developed the world's first portable brain scanner, and it may soon be able to "read a person's mind," playing a major role in facilitating medical breakthroughs.

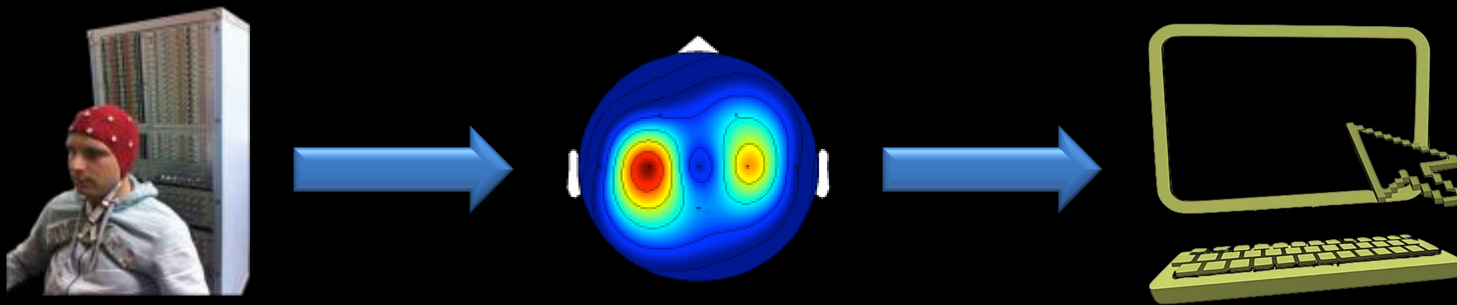


Dr. Philip Low wearing the "iBrain" (Misha Gravenor/TechnologyReview.com)

"This is very exciting for us because it allows us to have a window into the brain. We're building technology that will allow humanity to have access to the human brain for the first time," said the project's leader, Phillip Low.

KGTV reports that the device, created by San Diego-based NeuroVigil, and dubbed the iBrain, fits over a person's head and measures unique neurological patterns connected to specific thought processes.

Déjà Vu?



- Group X reports algorithm breakthrough; publishes numerous extensions
- Group Y can't replicate the results
- Group X had to manage particular algorithmic details in order to succeed (pre-processing, training, choice of subjects, inclusion criteria for data, etc.)
- Group Y spends more time replicating the results but still finds the algorithm doesn't work on different data
- *This is the state of algorithm development for neural engineering today*
- *Does this sound familiar?*
- *Not conducive for overall progress in the field*

Brain Machine Interfaces

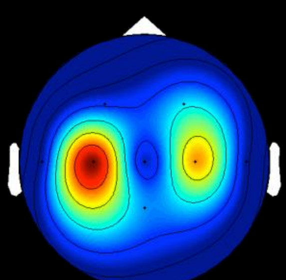
Amplifier
- Gain
- Bandwidth
- Impedance
- Channel Count
- CMRR
- Electrode Technology

Referencing & Montage
- Common Average Reference
- Laplacian Reference
- Common Spatial Filtering
- Unipolar (Fp1)
- Bipolar (Fp1-F3)

Signal Processing
- Frequency bands
- Energy
- Phase
- Coherence
- Source Localization
- etc

Model Building
- Number of participants
- Health of participants
- Number of Test Trials
- Number of Training Trials
- *Sufficient variability?*
- *What is normal?*

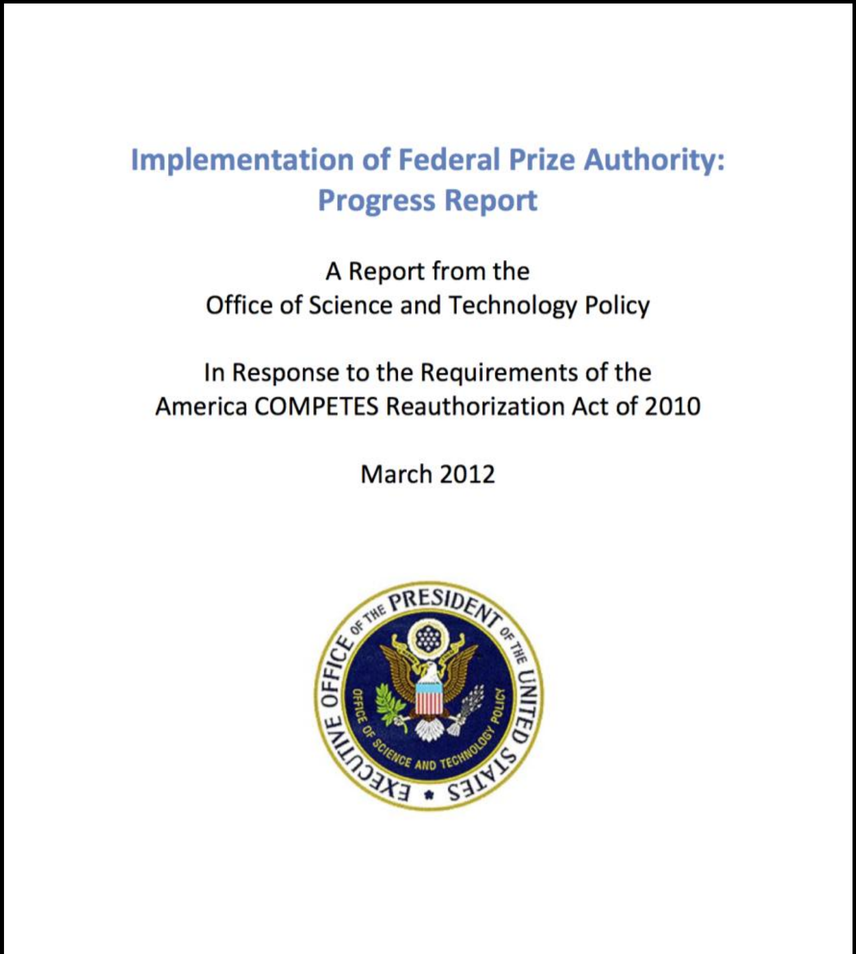
Results
- Generalizability
- Repeatability



Machine Learning

- Neural signal processing cannot leverage most contemporary machine learning algorithms because the data sets are too small
- Applications like speech and image processing can exploit “found data” because the signals are external; neural signals require sophisticated and sometimes invasive data collection techniques, which makes the cost of collection high
- Only by pooling community resources and expertise can you build large databases that drive research; this collaboration model has been successful in a number of fields including high-energy physics, HLT and data mining
- Potential for creating new research opportunities with multimodal data
- In recent years there has been a shift to unsupervised algorithms that require orders of magnitude more data than supervised learning; neural databases are nowhere close to supporting this type of research

Proposed Solution
Neural Engineering Data Consortium



Implementation of Federal Prize Authority: Progress Report

A Report from the
Office of Science and Technology Policy

In Response to the Requirements of the
America COMPETES Reauthorization Act of 2010

March 2012

EXECUTIVE OFFICE OF THE PRESIDENT OF THE UNITED STATES
OFFICE OF SCIENCE AND TECHNOLOGY POLICY

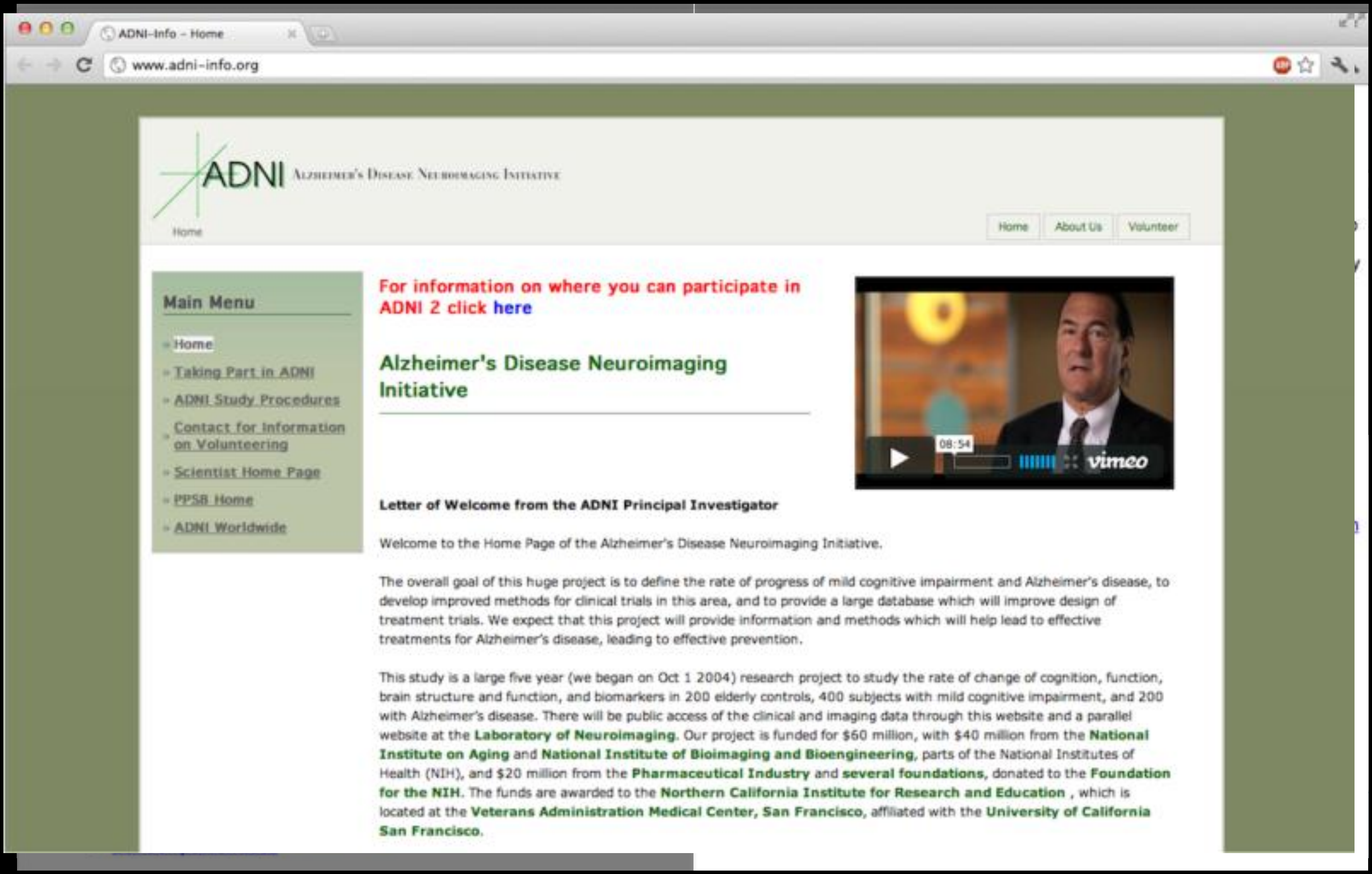
- Centrally organized data generation & distribution
  - Prize-Based Competitive Research
  - Massive Data Corpora
  - Common Evaluations
- *Level the playing field*
- Proven paradigm
  - Neural Data
  - Linguistics Data
- Temple is an ideal host

Demonstrated Community Interest

- BCI Contest
  - Based at Berlin Institute of Technology
  - Four contests to date
  - Bragging rights
  - Contributed data
  - Strong interest from unfunded, non-traditional labs

1.	Qingguo Wei	91%	Tsinghua University, Beijing
2.	Paul Hammon	87%	University of California, San Diego
3.	Michal Sapinski	86%	Warsaw University, Poland
3.	Mao Dawei	86%	Zhejiang University, P.R.C.
3.	Alexander D'yakonov	86%	Moscow State University
3.	Liu Yang	86%	National University of Defense Technology Changsha, P.R.C.
7.	Florian Knoll	84%	TU Graz
7.	Zhou Zongtan	84%	National University of Defense Technology Changsha, P.R.C.
9.	Jianzhao Qin	83%	South China University of Technology, China and Institute for In
10.	Matthias Krauledat	82%	Fraunhofer FIRST, Berlin
11.	Kiyoung Yang	81%	University of Southern California
11.	Martin Hieden	81%	TU Graz, Austria
13.	Archis Gore	79%	Fergusson College, Pune
13.	Elly Gysels	79%	CSEM, Neuchatel
15.	Xiaomei Pei	69%	Xi'an Jiaotong University, P.R.C.
16.	Ehsan Arbabi	67%	Sharif University of Technology, Tehran
17.	Florian Popescu	66%	Fraunhofer FIRST, Berlin
18.	Hyunjin Yoon	65%	University of Southern California
18.	Guido Nolte	65%	Fraunhofer FIRST
20.	Timothy Uy	60%	University of California, Irvine (?)
21.	Wit Jakuczun	59%	Warsaw University of Technology
22.	Ken Wong	58%	Stanford University
23.	Xi-Chen Sun	54%	Peking University
24.	Nanying Liang	50%	Nanyang Technological University, Singapore
25.	Bin An	48%	University of Science and Technology of China, Hefei
26.	Miharu Nishino	44%	Univ. of Tokyo
27.	Yan Ning	22%	University of Science and Technology of China, Hefei

Existing Data Repositories



Our Plan

Phase 0: Exploratory

- Develop concept
- Identify Collaborators & Stakeholders

Phase 1: Proof of Concept

- EEG-based data bank
- Temple Neurology
- Funded by Temple University Hospital

Phase 2: Planning Grant

- NSF with support from DARPA, NIH
- Build support within community
- Research goals
- Administrative structure

Phase 3: Center Grant

- NIH, NSF, DARPA
- International Partners

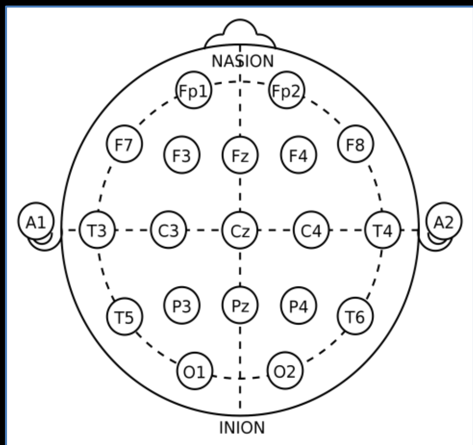
Phase 0 (One Year): Exploratory

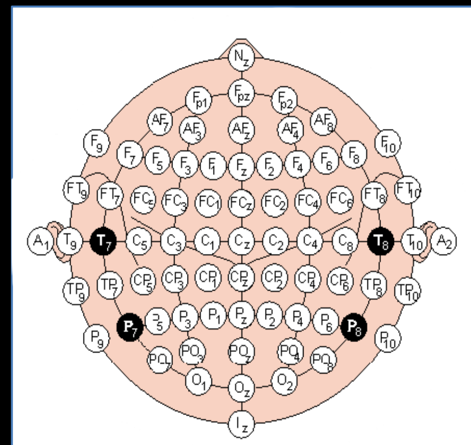
- July 2011 visit to LDC (Philadelphia)
- November 2011 visit to potential sponsors
  - NSF, DARPA, NIH
- Spring 2012 – meeting with potential collaborators throughout Temple Hospital /Medical School / Bioengineering Department
- Pursuing industry interest (e.g., Google)
- Seed money from DARPA, Temple Office of Research and Temple College of Engineering for Phase 1 data preparation

Phase 1 (9 mos.): EEG Data Corpus

- Temple Neurology: over 12,000 EEGs in digital archive (over ten years of data)
  - Accompanying patient summaries (MS Word)
  - Some have accompanying MRI
- De-identify, sort, mark-up, label data
  - Medical condition (pathology, anatomy)
  - Medication
  - Behavior
- Make data corpus freely available online

EEG 10-20 System

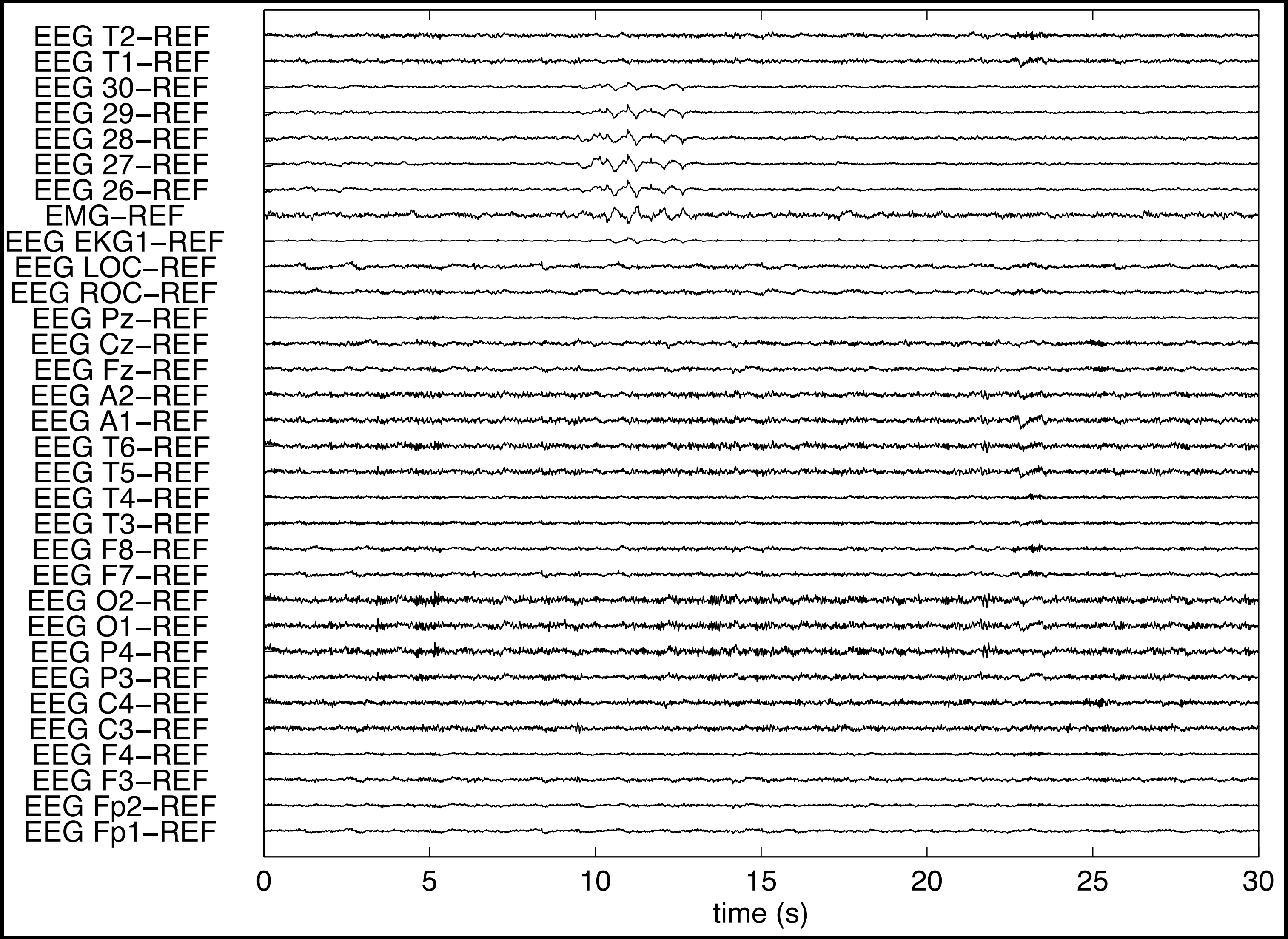




Frontal
Temporal
Central
Parietal
Occipital

- Each Channel :
  - Sampling Rate: 250 Hz x 2 bytes per sample
  - Data Rate: 500Bps/per channel
- Channel count varies by application:
  - Clinical: 24-32
  - Research: up to 256 channels
- Typical 40 minute EEG = approximately 40MB, plus video
- EKG, EMG, EOG often recorded as well
- EEG Corpus is approximately 500GB

Typical EEG Montage



Phase 1: Outcomes

- Demonstrate value of massive biomedical data corpus (machine learning)
- Demonstrate our ability to generate, curate, & disseminate data of this magnitude
- Establish industry-standard baselines for existing technology

Phase 2 (One Year): NSF Planning Grant

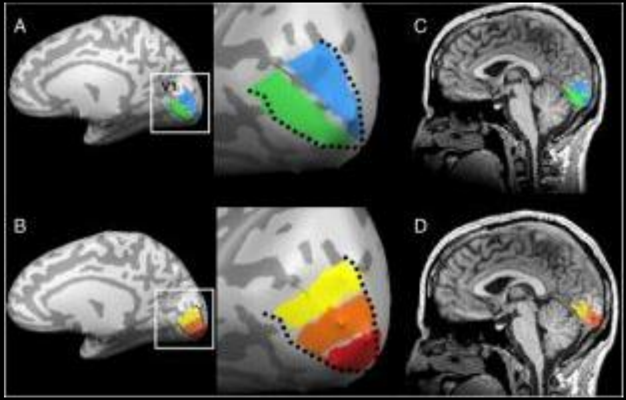
- Mobilize scientific community
  - Conference symposia
  - Build “Working Group” of community experts
  - Plan first two to three corpora in detail
- Establish management plan
  - Membership, Costs
  - Physical infrastructure
- Expand on first corpus
- October due date – Funding through 2014
  - NSF has encouraged us to submit

Phase 3 (Two Years): Center Grant

- Joint proposal to be funded by NSF, DARPA and NIH
- Consortium members: academia, medical, industry
- Clinical Benefits
  - Epilepsy, TBI – we are in premiere position to write proposals for new brain technology
  - Computer assisted patient evaluation
  - Improved understanding of brain function
- Located at the Temple School of Medicine (adjacent to Temple Hospital)
- Administered through TUMS
- Temple Benefits
  - Leverages Temple's strengths in clinical medicine and neuroscience, as well as their clinical infrastructure

Longer-Term Opportunities

- Simultaneous collection of neural, voice and video data
- Integration of other types of bioengineering data (e.g., clinical recordings of vital signs, other forms of imaging)
- Applications of machine learning to multimodal biomedical signals is in its infancy due to a lack of data



Temple – Ideal Host Institution

- Established relationship with LDC
  - Ability to leverage HLT experience
- Temple Hospital
  - Clinically oriented research
  - Diverse patient population
  - Large clinical scope
  - Interested in large-scope research endeavors
  - Administratively viable

Summary: Why Common Resources?

- Common resources:
  - Accelerates research progress
  - Increases participation, particularly of under-resourced research groups
  - Increases technology transfer
  - Promotes collaboration
  - Leverages investments
- A much broader community of supporters: research hospitals (globally), federal research facilities (CDC, military research labs, national labs, etc) are all potential first-tier members