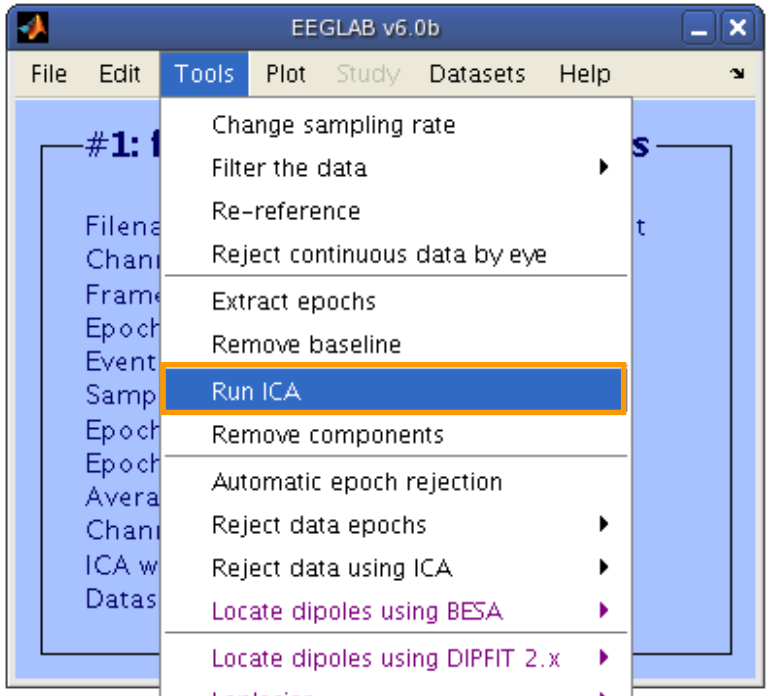
ICA decomposition practicum and practical Q&A

Makoto Miyakoshi

25th EEGLAB Workshop at Tokyo

Sept 25, 2017

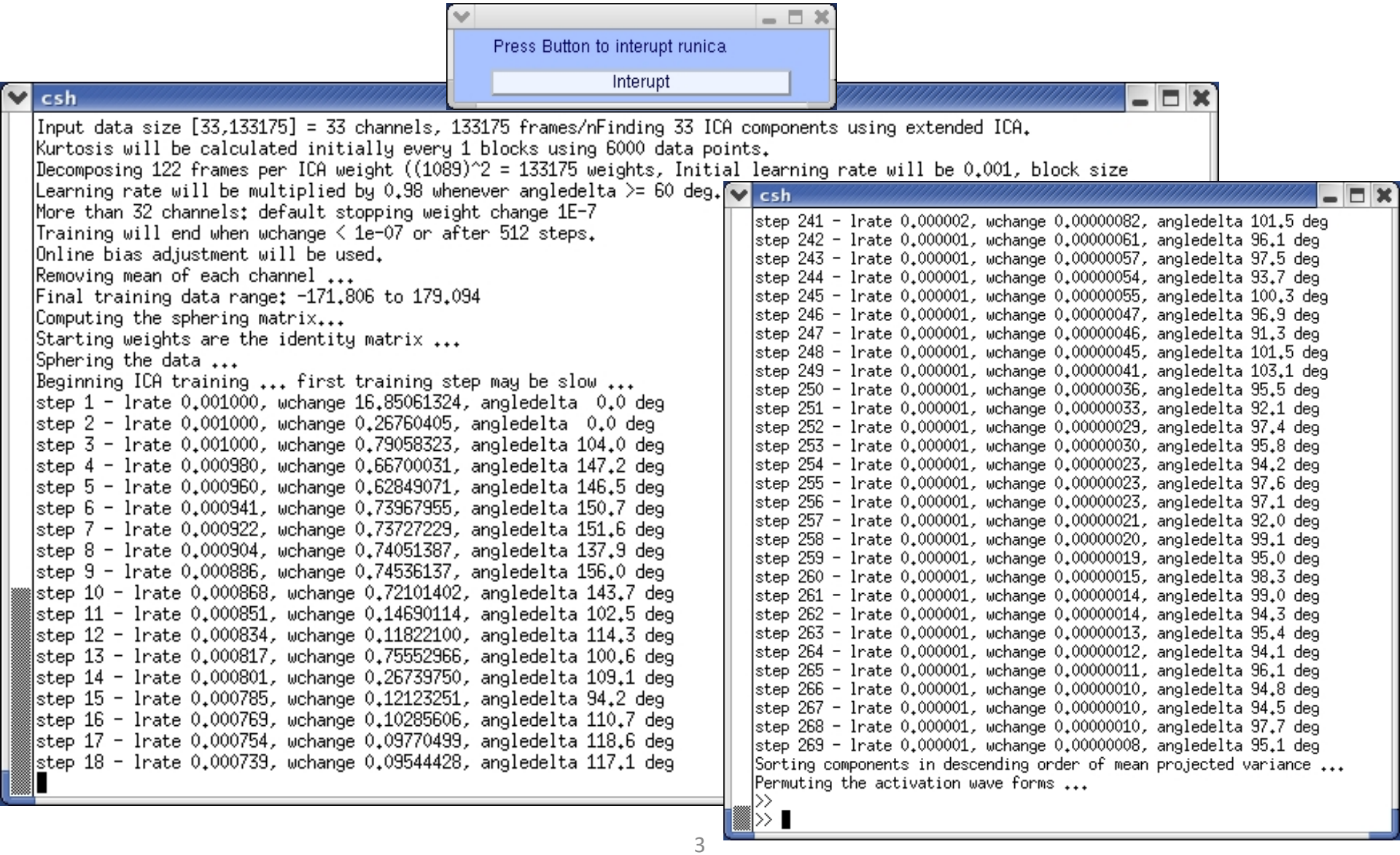
Perform Independent Component Analysis



Option	Default	Comments
'extended'	0	1 is recommended to find sub-gaussians
'stop'	1e-7	final weight change → stop
'lrate'	determined from data	too small → too long… too large → wts blow up
'maxsteps'	512	more channels → more steps
'pca'	0 or EEG.nbchan	Decompose only a principal data subspace

Other algorithms:
binica, amica, cudaica, beamica

Runica() in progress



Practicum finished!
Thank you for your attention.

Supplementary Material
Practical Q & As about ICA

Easy

1. ICA vs. PCA?
2. ICA after IC rejection cleans data? What is rank?
3. What if I have 10,000 'boundary' events in data?

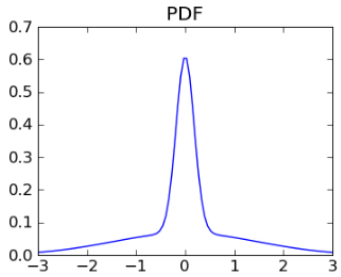
Normal

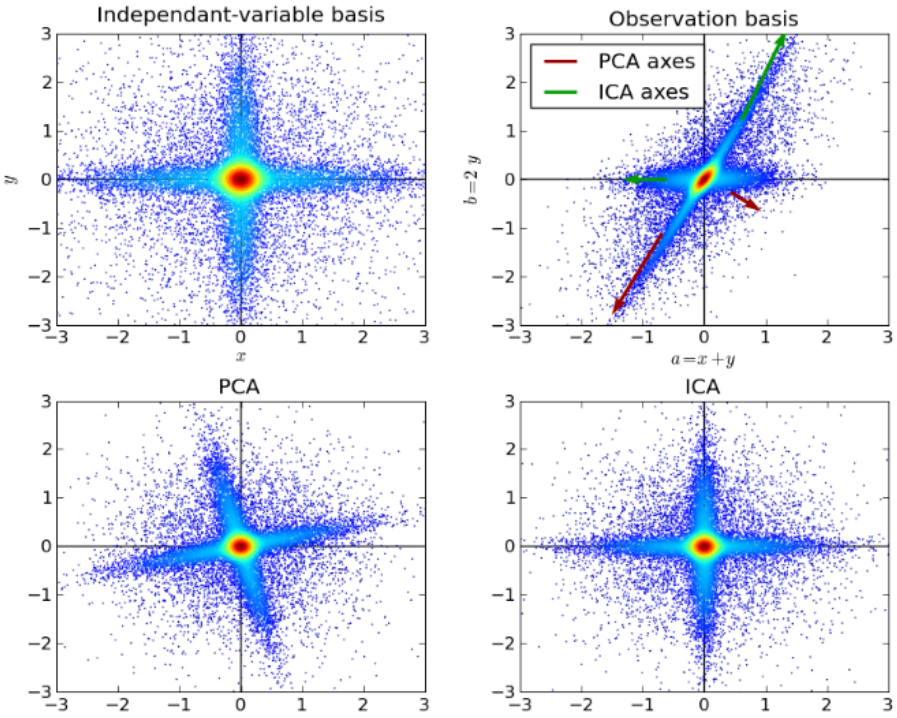
4. How to measure ICA's goodness of fit?
5. ICA algorithms? Infomax vs. AMICA?
6. How many channels and datapoints do we need?
7. How does ICA model physiology?
8. Why do connectivity analyses work after ICA?

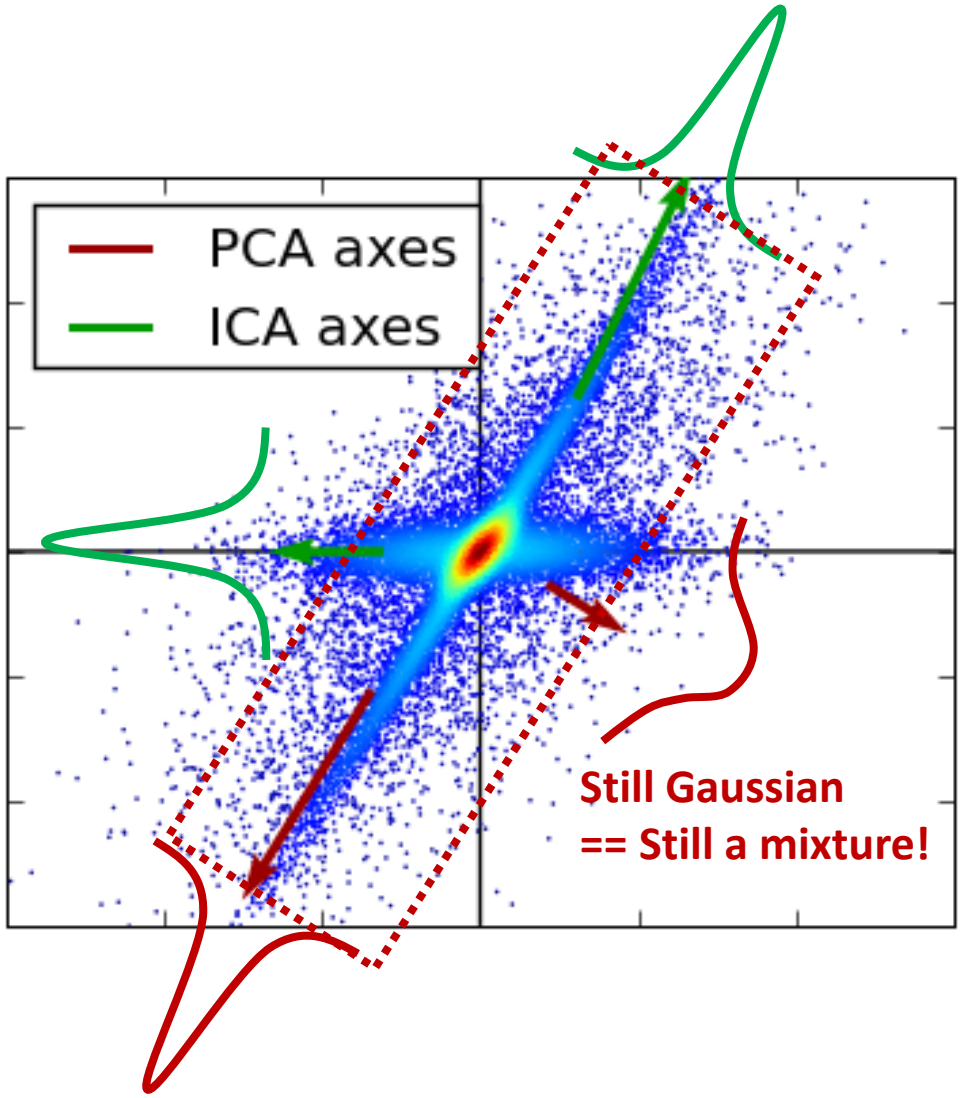
Hard

9. Why are similar ICs found? What is a subspace?
10. What are limitations of ICA+dipfit approach?

1. ICA vs. PCA?

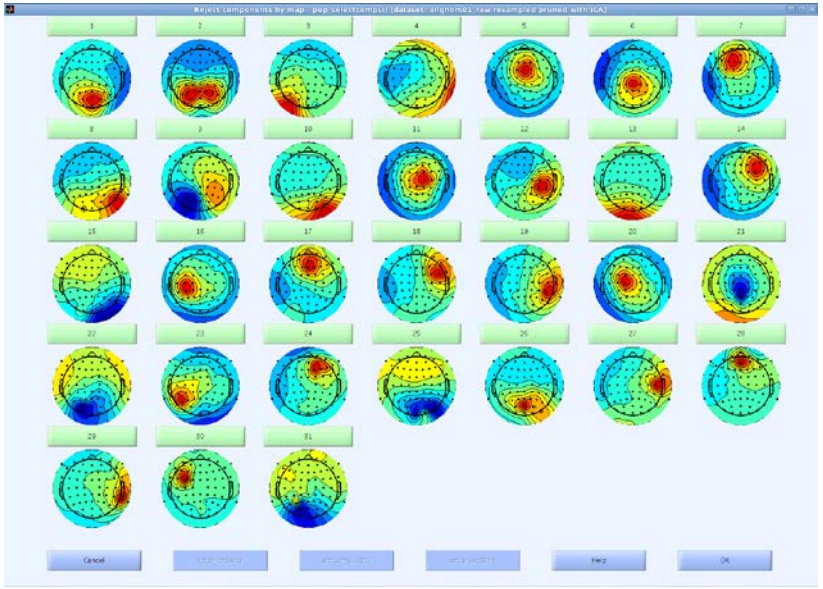


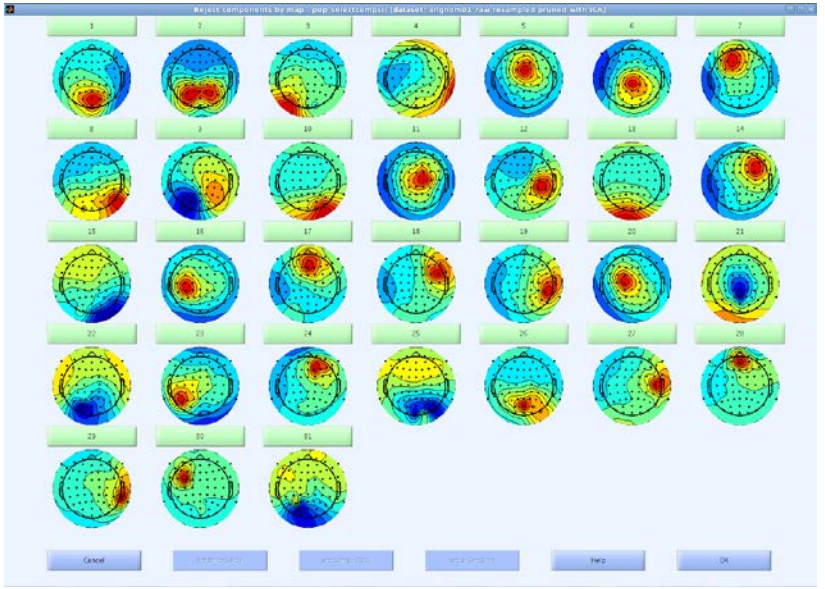




Figures by Gaël Varoquaux
http://gael-varoquaux.info/science/ica_vs_pca.html

2. ICA after IC rejection cleans data? What is rank?





64ch data, 31 ICs selected -> Backproject to channels -> ICA again -> same 31 ICs show up, not the new 64 ICs! This is becaue ICA computes data *rank* to set the number of ICs to calculate. This is a very common misunderstanding, and people expect ICA cleans data in this way.

Data rank of the following equasion is 2 (because the first and the third are linearly dependent) This is called rank deficient.

$$2x + 3y - 5 = 0$$

$$3x + 5y + 3 = 0$$

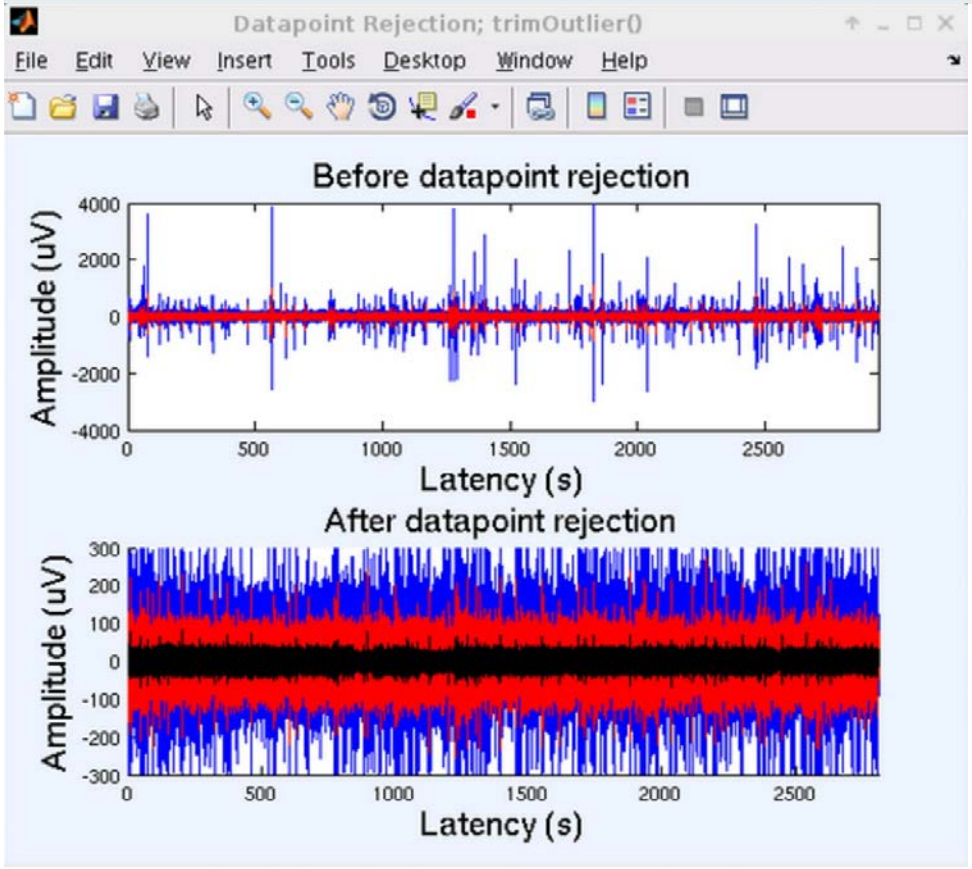
$$4x + 6y - 10 = 0$$

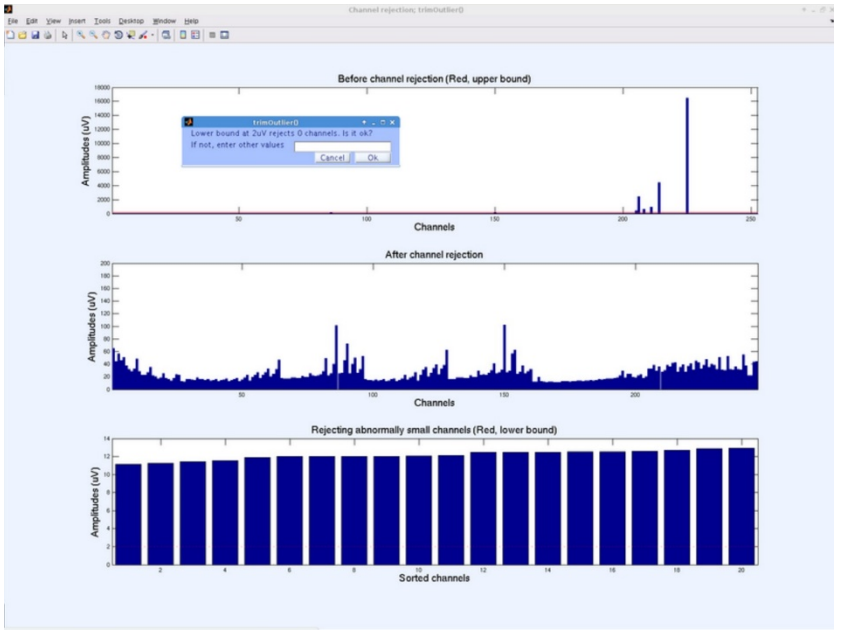
3. What if I have 10,000 ‘boundary’ events in data?

- To ICA, chronologically ordered and randomly shuffled EEG are equivalent. This is becaue ICA processes each time point individually.

- When you start ICA, it randomizea all datapoints across time, and it’s repeated for every iteration.
- So 10,000 ‘boundary’ events does NOT affect ICA.

If having 10,000 ‘boundarys’ does not matter...

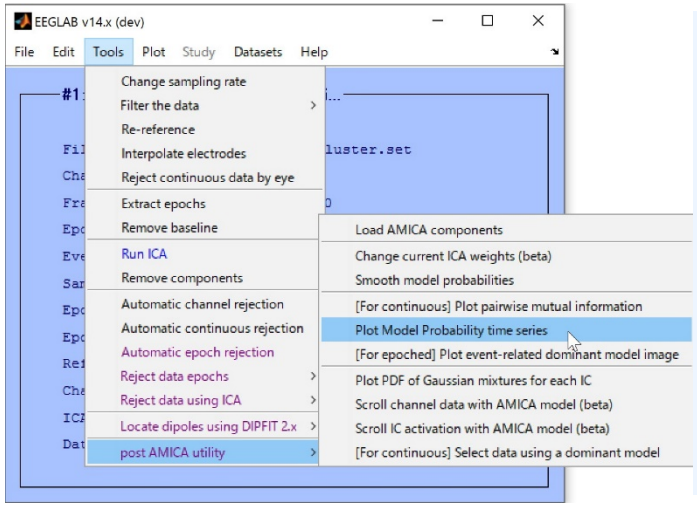


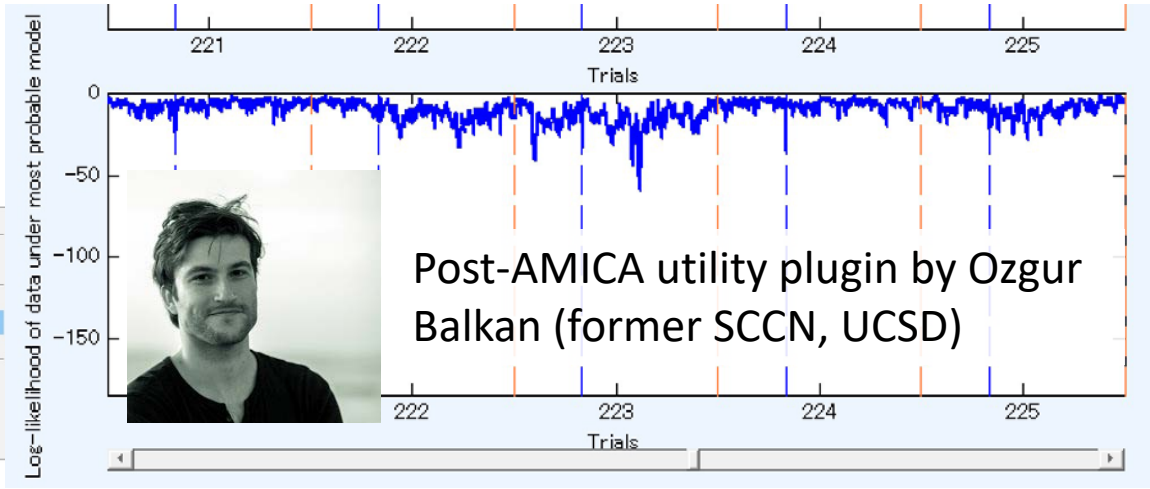


Use *trimOutlier()* to force all-channel EEG to be +/-300 uV, for example, ICA, and apply the weight matrix to midly and decently processed data.

4. How to measure ICA's goodness of fit?

- ICA returns no variance measure. How to measure Goodness of fit (GOF)?
- AMICA returns log likelihood time series, which represents datapoint-by-datapoint change in GOF (therefore explains data non-stationarity).





Can we measure goodness of fit in infomax?

Infomax ICA [Lee et al. 1999]:

$$\boldsymbol{W} \leftarrow \boldsymbol{W} + \eta[\boldsymbol{I} - \boldsymbol{f}(\boldsymbol{y}) \cdot \boldsymbol{y}^{\boldsymbol{T}}]\boldsymbol{W}$$

Gradient

Source activity: $\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x}$

Non-linear function: $\boldsymbol{f}(\boldsymbol{y}) = tanh(\boldsymbol{y}) + y$

Model Deviation Index (MDI)

Cross-talks errors

$$MDI\ (\boldsymbol{W}_0) = \frac{\left\|\langle \boldsymbol{f}_{\mathrm{i}} \cdot \boldsymbol{y}_{\mathrm{j}}^{\mathrm{T}} \rangle_{\mathrm{i}\neq\mathrm{j}}\right\|_{\mathrm{F}}}{\left\|\langle \boldsymbol{y} \cdot \boldsymbol{y}^{\boldsymbol{T}} \rangle\right\|_{\mathrm{F}}}$$

Source Power

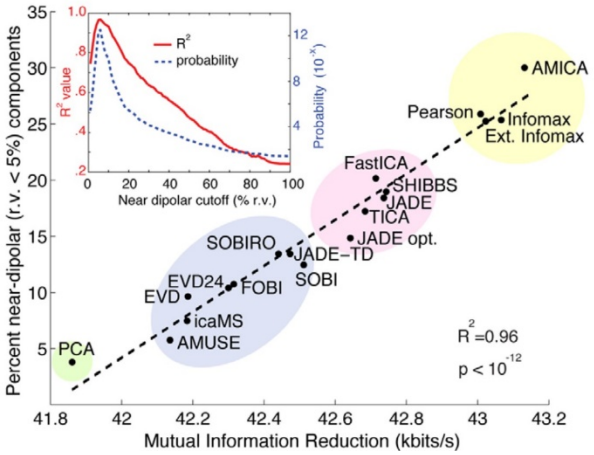
Sheng-Hsiou (Shawn) Hsu
(SCCN, UCSD)

- If it is worthwhile, it can be made into a plugin...?

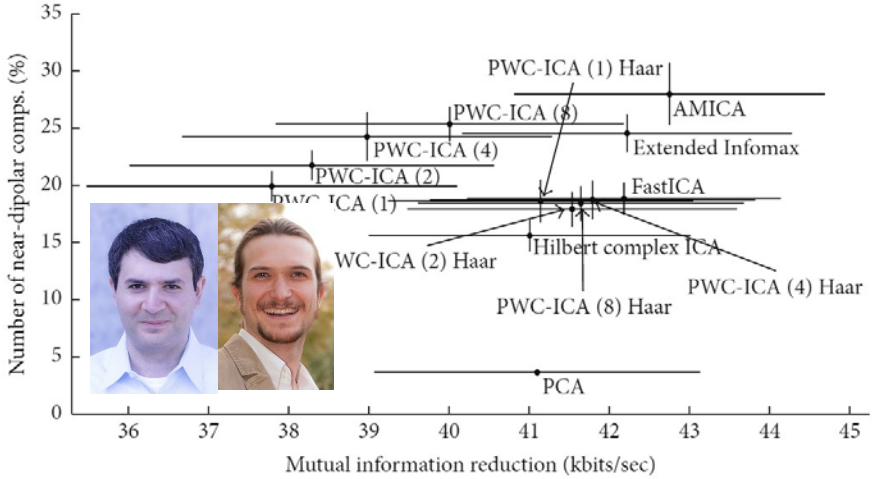
Hsu & Jung, 2016 (under review)

5. ICA algorithms? Infomax vs. AMICA?

Infomax == Makes ALL probability density functions into super-Gaussian.
Extended infomax == Makes them either super- or sub-Gaussian as necessary.
Adaptive Mixture ICA == Makes them sum of n-Gaussian distributions (default: 3)



Delorme et al., 2012



Ball, Bigdely-Shamlo, Mullen, Robbins, 2016

BSS algorithm	Amari indices Experiment 1: static uncoupled oscillators	Experiment 2: static coupled sources	Experiment 3: dynamic coupled oscillators
AMICA	0.31	0.37	0.25
Extended Infomax	0.31	0.38	0.25
FastICA	0.32	0.30	0.26
PWC-ICA (1)	0.35	0.29	0.28
PWC-ICA (2)	0.34	0.41	0.26
PWC-ICA (4)	0.35	0.41	0.24
PWC-ICA (8)	0.30	0.41	**0.23** ($p < 2.3E-3$)
PWC-ICA (1) Haar	0.30	0.32	0.25
PWC-ICA (2) Haar	**0.21** ($p < 1.5E-11$)	0.35	**0.23** ($p < 4.5E-4$)
PWC-ICA (4) Haar	**0.21** ($p < 1.5E-10$)	0.40	**0.23** ($p < 4.4E-2$)
PWC-ICA (8) Haar	**0.27** ($p < 1.5E-5$)	0.38	0.24
Hilbert complex	**0.28** ($p < 1.5E-4$)	0.38	**0.23** ($p < 6.7E-3$)
Average baseline	0.36	0.42	0.36

Amari index

$$= \frac{1}{2n(n-1)} \sum_i \left(\left(\sum_j \frac{|P_{ij}|}{\max_k |P_{ik}|} \right) - 1 \right) + \frac{1}{2n(n-1)} \sum_j \left(\left(\sum_i \frac{|P_{ij}|}{\max_k |P_{kj}|} \right) - 1 \right).$$

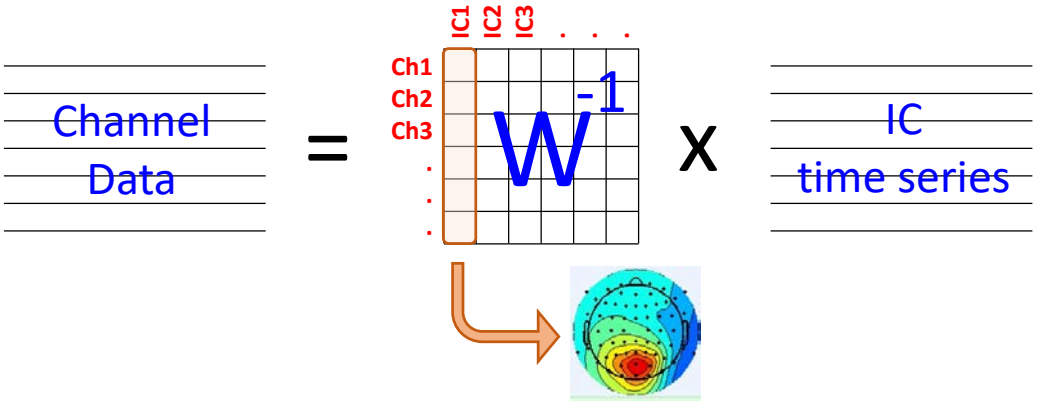
An Amari index of zero for a nonsingular matrix **P** indicates that **P** is a permutation of a diagonal matrix, while an index of one indicates that the entries of matrix **P** are the same constant. Thus, a lower value of the Amari index indicates that **W** is better at isolating the individual sources.

6. How many channels and datapoints do we need?

- Rule of thumb formula: channels^2 x k, k = 20~30 for 30 channels when sampling rate is 250 Hz. The constant k should increase as the number of channels increases.
- Downsampling to 100-128Hz does not seem to influence decomposition quality (it could be even better, since cutting off high frequency of non-interest; at least, much faster.)
- By the same token, upsampling does not help lack of datapoints.
- There is neither theoretical nor empirical evidence available for these numbers.

7. How does ICA model physiology?

- Scott qualitatively formalized it using the concept of 'near synchronous patch' in Onton and Makeig (2006) in *Progress in Brain Research*. That's where ICA met physiology.
  - This makes ICA more than 'just one of linear transform algorithms', and 'mixing matrix == spatial filter' was formalized.
- EEG.data = EEG.icawinv * EEG.icaact



IC1 Scalp Tophgraphy!

For example, consider Freeman's model of EEG source dynamics, based on his observations of mammal brains with small (sub-millimeter spaced) electrode grids, of circular wave patterns that spread across small areas of cortex like pond ripples produced by throwing a small rock into a pond (Freeman, 2004b). What field dynamics on the scalp should be produced by such activity active at, e.g., 10 Hz? At a nominal traveling velocity of 2 m/s, and assuming a cortical domain diameter of as much as 3 cm, the 10-Hz phase difference between the focal center of the 'pond rippling' potentials and the edge of the active 'ripple' area (1.5 cm from the center) would be only

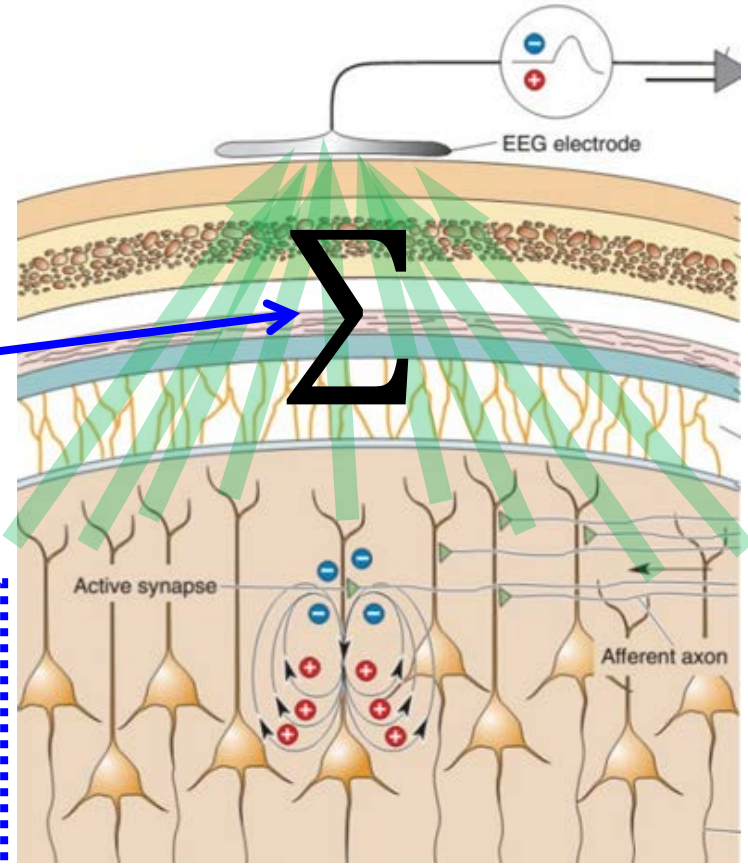
$$\frac{1.5\ \text{cm}}{0.002\ \text{m/ms} \times 100\ \text{cm/m}} / 100\ \text{ms/cycle} \times 360^\circ/\text{cycle} = 27^\circ$$

Thus, the outer edge of the pond-ripple pattern would lead (or follow) the center by less than a 13th of a 10-Hz cycle, and mean local-field potentials within the patch (and at the scalp electrodes) would change from positive to negative and back again *nearly* synchronously. Unless the cortical

Onton J, Makeig S. 2006. Information-based modeling of event-related brain dynamics. Prog Brain Res. 159: 99-120.

What makes ICA special among other signal processings?

	ICA's assumptions	Biophysical facts of EEG
Mixing Process	Linear	[Probably linear]
Mixing Speed	Instantaneous	Near the speed of light
Source PDF	Non-Gaussian	Super Gaussian
Definition of the 'source'	Temporally independent of each other	Unknown: empirical and analogical evidence by electrophysiology
Source spatial stationarity	Stationary	Non-stationary (traveling waves)
Source temporal stationarity	Stationary	Non-stationary (task-dependent)



Benign/acceptable violations
(e.g. Using ERP paradigm helps to increase stationarity)

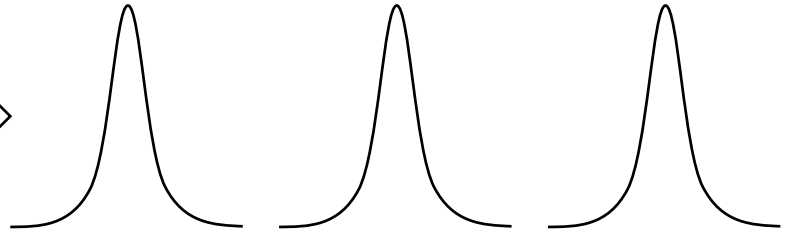
I-D identity indirectly proves it.

Independence-Dipolarity identity *(I-D identity, IDID)*

Scalp-channel signal PDF

ch1 ch2 ch3

Independent Component PDF



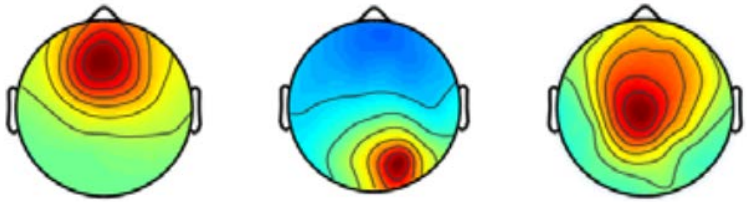
IC1 IC2 IC3

Temporally maximally independent
--> Tautological, due change.

'It's a non-trivial nature of the *I-D identity* that a spatial solution emerges from a temporal solution. Solving a temporal problem solves a spatial problem.'

Hirokazu Tanaka

Independent Component Scalp amplitude distribution



They are unexpectedly highly *dipolar.*
--> Unintended change!

8. Why do connectivity analyses work?

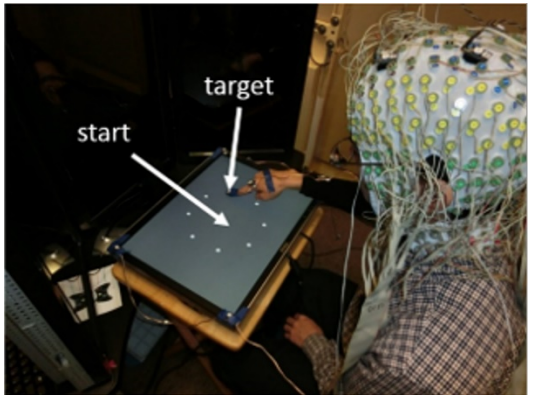
- Again, to ICA, chronologically ordered and randomly shuffled EEG are equivalent. This is becaue ICA processes each time point individually.

Captures only ONE datapoint !

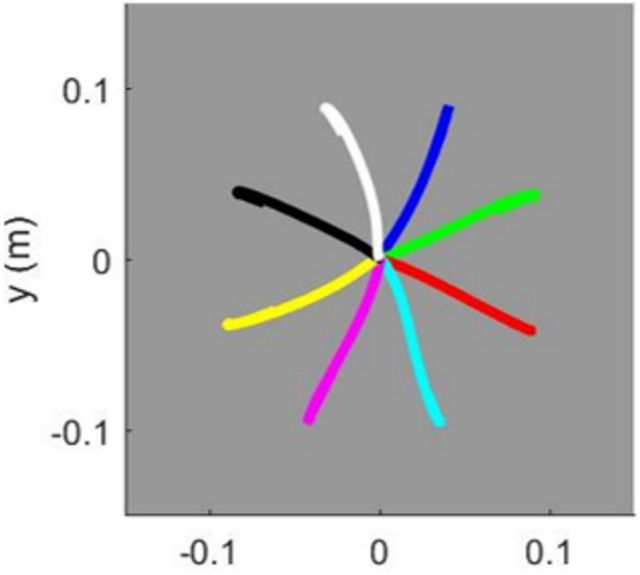
- ICA has no sensitivity to time-delayed dynamics between sources. ICA preserves across-source dynamics, which can be exploited later to study effective connectivity (e.g. Granger Causality).

Unpublished evidence of IC dynamics dependency (this is why SIFT works)

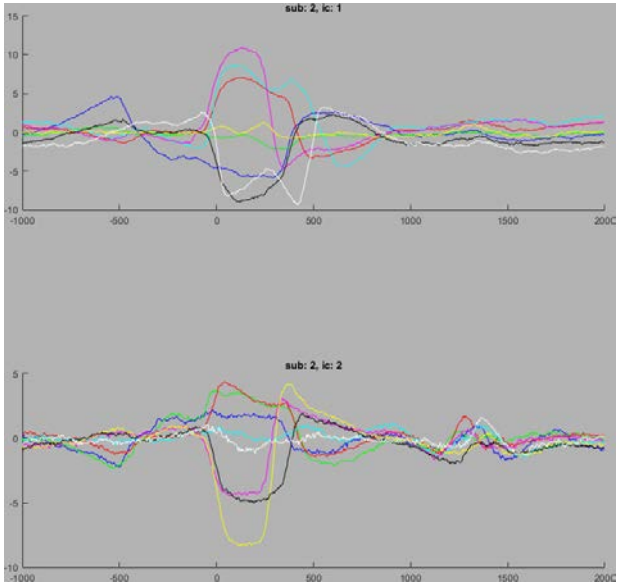
Takana et al. (submitted)
Reaching task using MoBI system at SCCN.



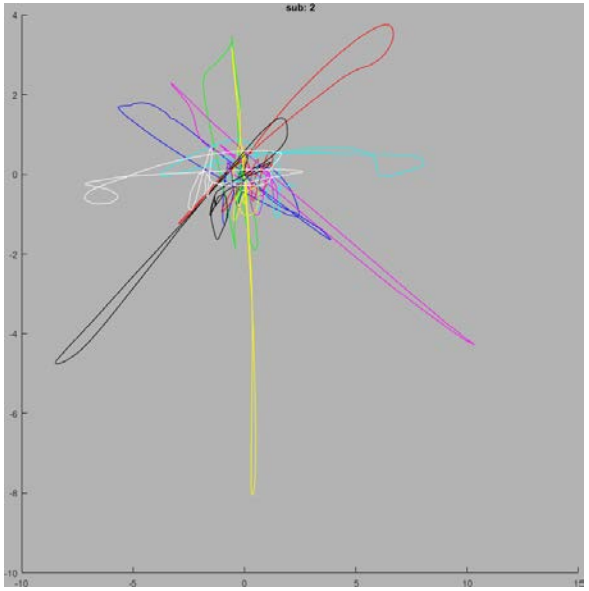
Hirokazu Tanaka
(JAIST)



Trajectory of finger point by motion capture



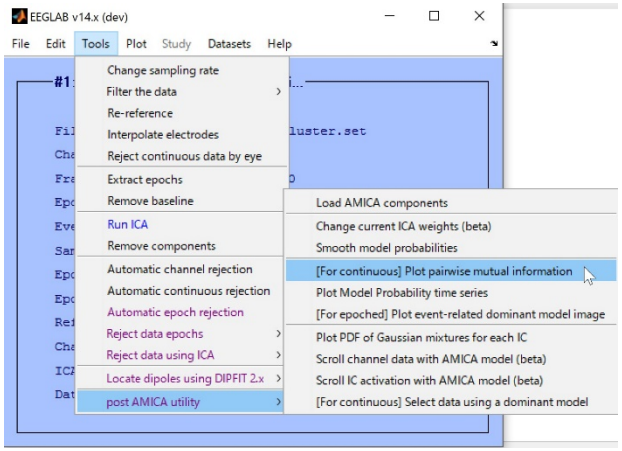
Horizontal and vertical EOGs: IC ERPs



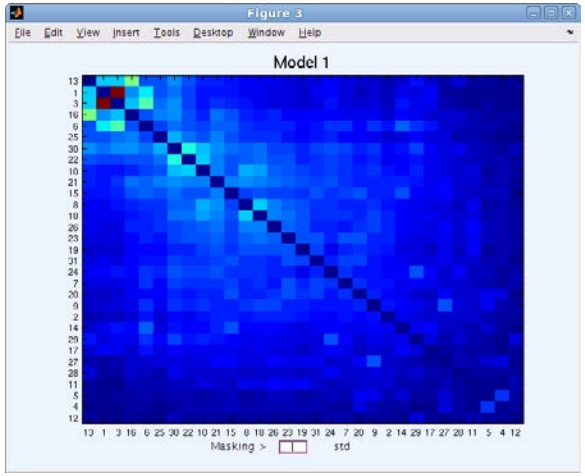
Lissajous figure by the two EOG IC ERPs

9. Why are similar ICs found? What is a subspace?

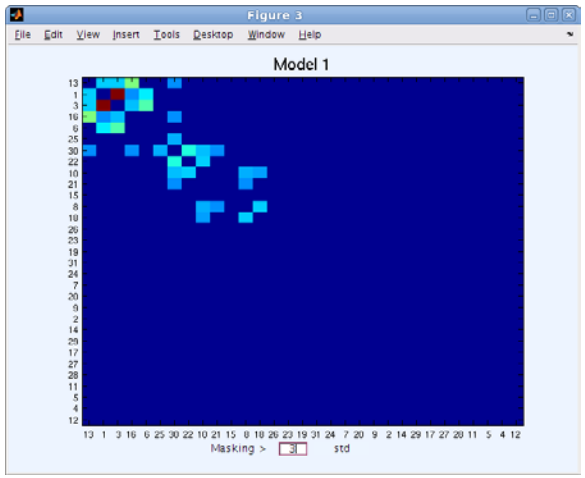
- Independent subspace
- Independent subspace is by definition a group of ICs that are intra-dependent but inter-independent.
- Dependence can be measured by computing pairwise mutual information (PMI) across ICs.



postAmicaUtility() plugin



Pairwise Mutual Information across ICs



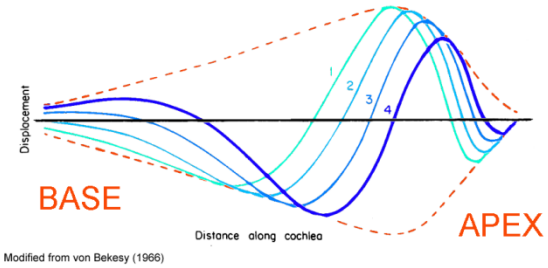
Masked with 3SD. IC1 and IC3 most likely forms subspace.

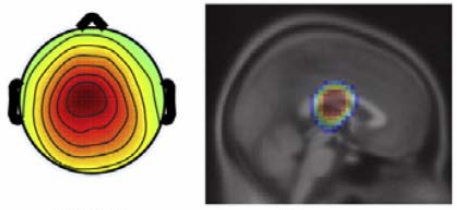
Do independent subspaces result from 'overfitting', or is there physiological significance?

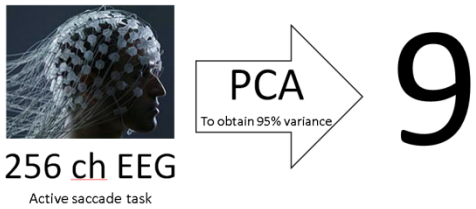
- Himberg et al. (2004) considered the subspace as ICA's model overfitting, and proposed a method to reduce data rank until subspace is eliminated.
- We have unpublished simulation study that moving dipole was decomposed into a subspace (Maki, in prep.)
- Does ICA subspace have physiological meaning? This is an open question.

10. What are limitations of ICA+dipfit approach?

- **Traveling waves in ECoG**
- **Poor source localization validity**
  - ICA's fault
  - Dipfit's fault
- Low 'true degrees of freedom'—only 10-20 good Ics
  - This is probably the nature of scalp EEG and not ICA+dipfit's limitation.



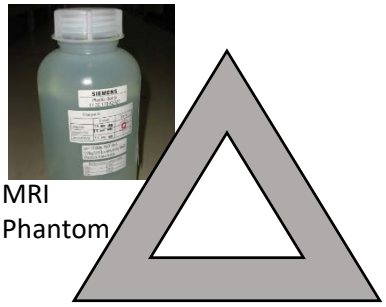




ICA model is NOT the ground truth itself—hence 'effective' sources.

http://www.d.umn.edu/~jfitzake/Lectures/DMED/InnerEar/CochlearPhysiology/PlacePrinciple.html

Vacation of the ground truth in EEG



Ground truth to scan

Imaging device
(X-ray, MRI)

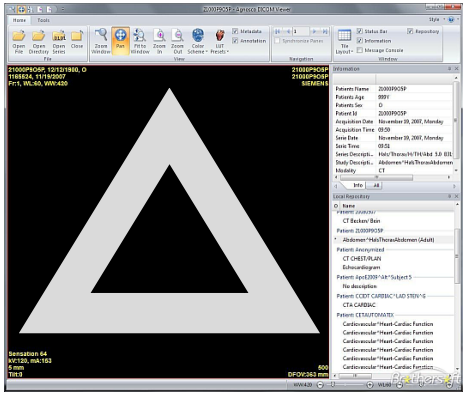
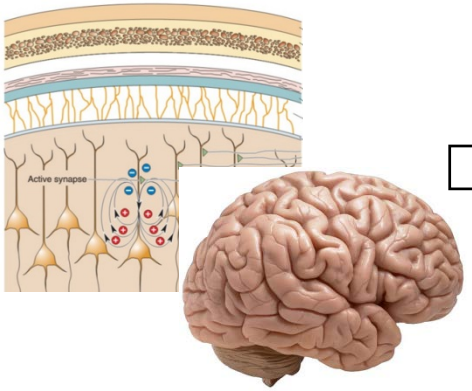
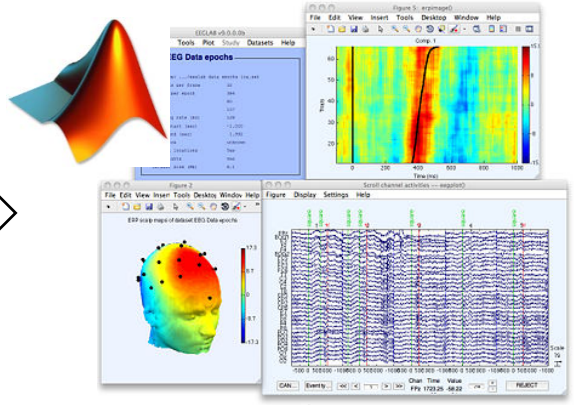


Image
(reconstructed truth)



Unknown ground truth shuttered into micro (single unit), meso (LFP), and macro (ECoG/EEG) scales.

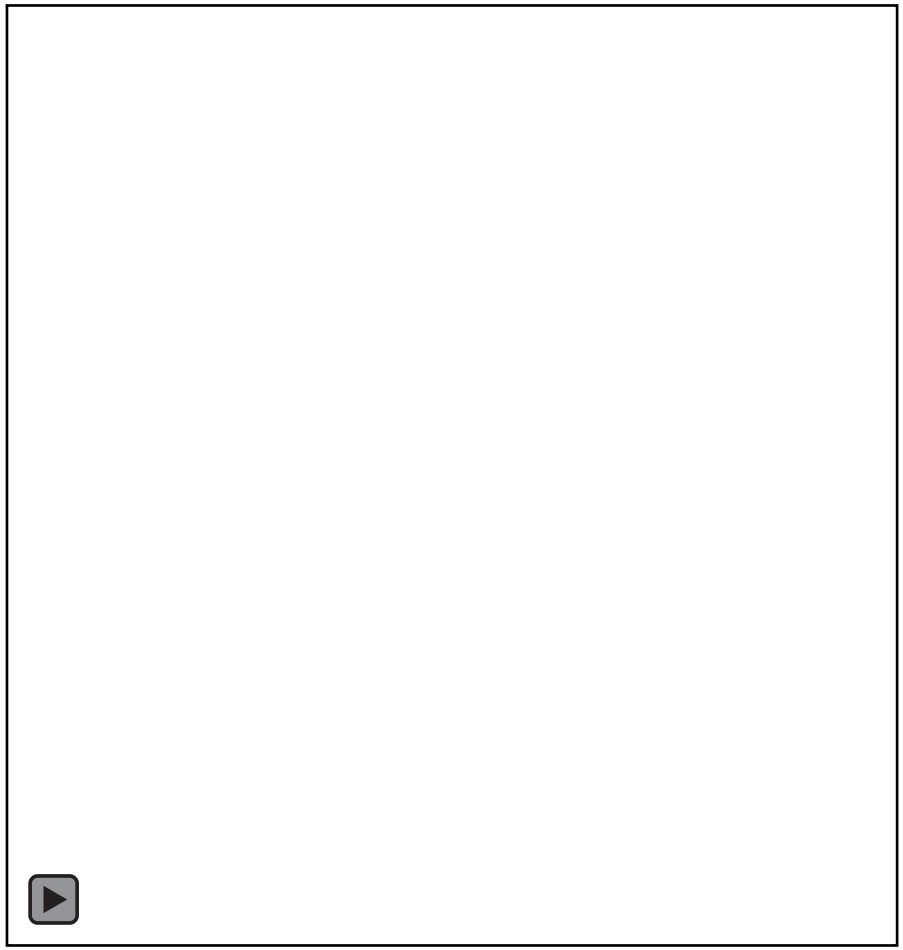
Recording device
(ECoG/EEG)



‘Colorful’ visualizations;
Is this reconstructed truth?

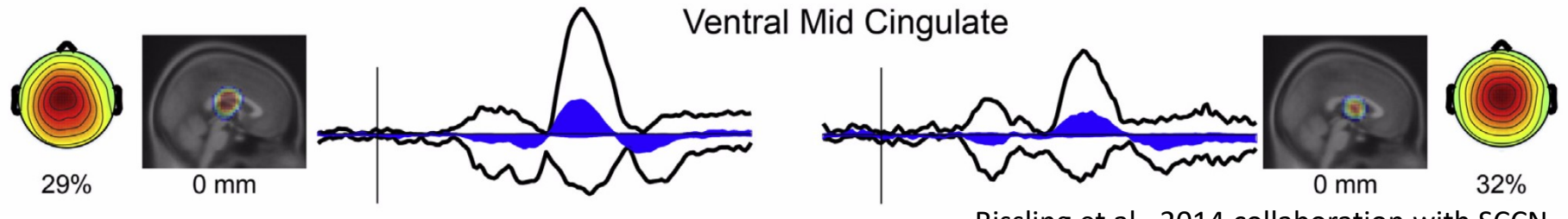
http://www.ocmr.ox.ac.uk/_asset/image/siemens-3t.png
http://img.brothersoft.com/screenshots/softimage/a/agnosco_-_free_dicom_viewer-368151-1274324869.jpeg
https://i.ytimg.com/vi/bBIBLJ7pqTg/maxresdefault.jpg

Traveling waves in ECoG: one ground truth that is necessarily missed by ICA due to assumption of *spatial stationarity*



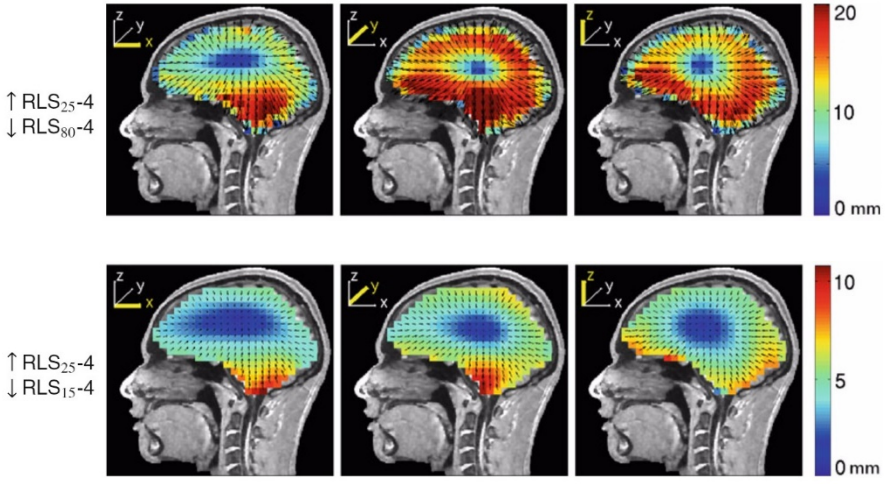
https://www.eurekalert.org/multimedia/pub/127004.php

Mystery of deep dipoles in EEGLAB



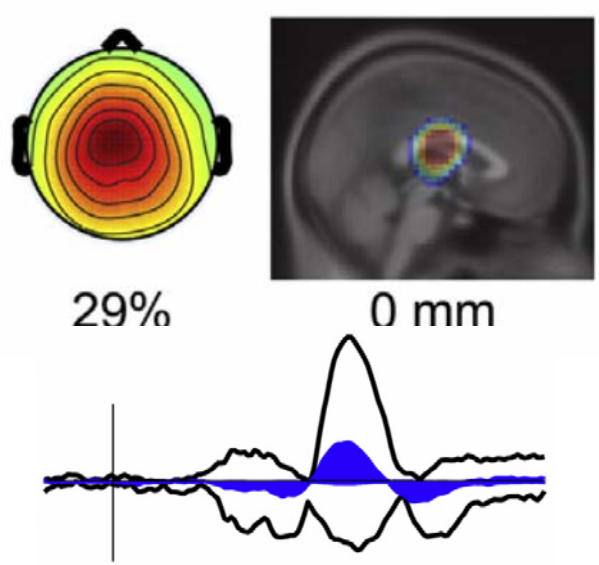
Rissling et al., 2014 collaboration with SCCN

- Brain-to-Skull conductivity ratio. Simulation Model, 25:1 (model parameter.) Top, 80:1 (EEGLAB default; skull is underconductive) Bottom, 15:1 (skull is overconductive).
- It does not make sense that EEGLAB head model fits dipoles too deeply. Should be opposite.

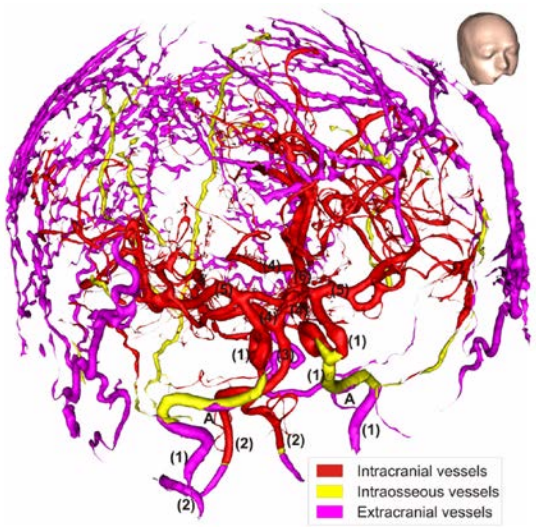


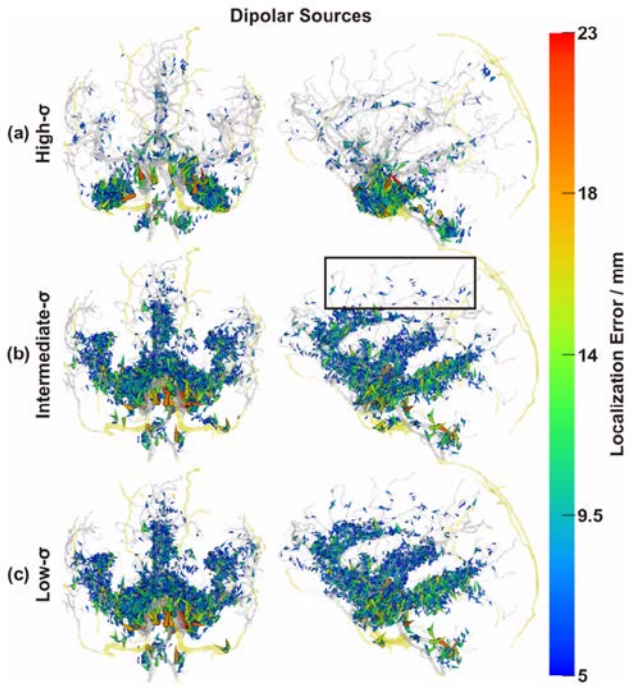
Zeynep Akalin Acar
Forward-model specialist

Dipfit's fault: Poor electric head forward model



(Rissling et al., 2014)



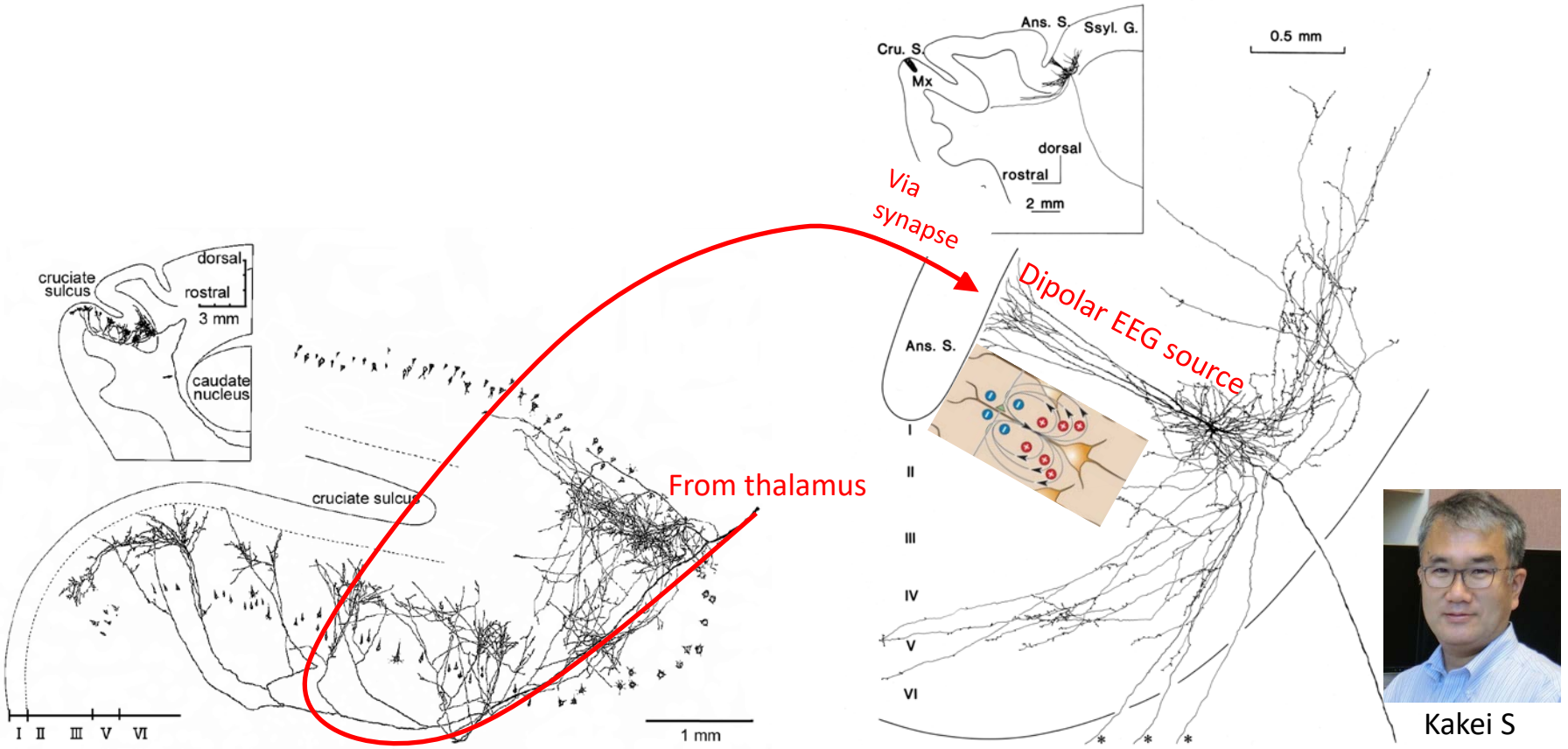


(Fiederer et al., 2016)

Hypothetical reason for deep dipoles

1. ICA resolves *sheet dipoles* correctly and returns very broad scalp projection map (ICA source ≠ point source!)
2. DIPFIT tris to fit a single dipole to explain the broad scalp map **with a point source**—by walking back to depth?
3. In doing so, dipfit uses simple physiological model—this is another source of inaccuracy.

Voice of ground truth: EEG sources could have quite an area on the cortex

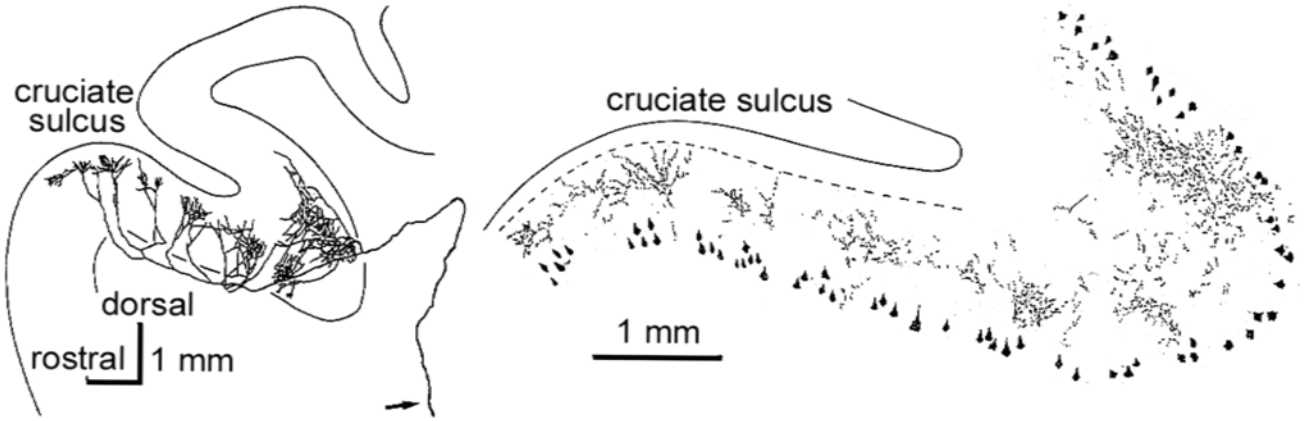


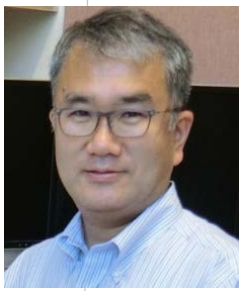
Thalamo-cortical axon (brings inputs)

Pyramidal neuron (forms dipolar EEG source)

Figures courtesy of Shinji Kakei and Yoshikazu Shinoda

Over a cm scale of synchronous source patch is possible (e.g. An entire gyrus/sulcus)





Figures courtesy of Shinji Kakei and Yoshikazu Shinoda

How to interpret and report deep dipoles?

- There are patterns in deep dipoles: For example, they show up near mid-cingulate or inferior occipital regions.
- Such a bias toward depth should come from *systematic structure*, such as spatiotemporal source patch size, *which is rather good*.
- In reporting, describe these deep dipoles as 'Upper/Lower basal' or even as 'deep dipoles' and do not mention their anatomical labels. But it is still encourraged to report their coordinates.
- When estimating their corresponding cortical source locations, move them to the surface along with radial axis (i.e. Corpus callosum -> midcingulate cortex.)

Excercise—Test your understandings

- Q1. What happens when you perform ICA after IC rejection?
- Q2. Does ICA care waveforms?
- Q3. Which exploits mutual information better, Infomax or Amica?
- Q4. How many datapoints do you need to run ICA?
- Q5. How does ICA correspond to physiology? [hint: dipolarity]