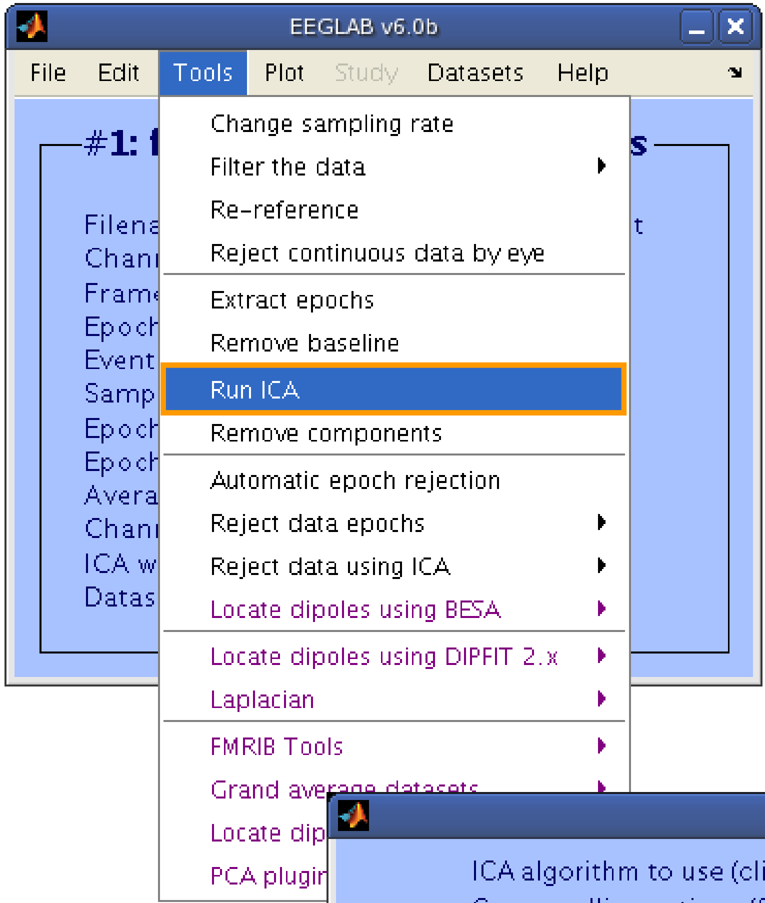
ICA decomposition of EEG data practicum

Makoto Miyakoshi

23rd EEGLAB Workshop at Mysuru

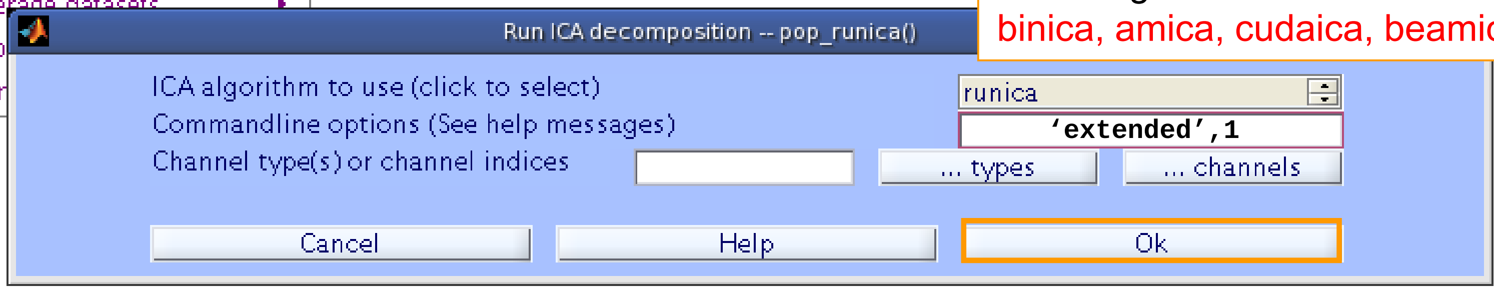
Jan 16, 2017

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

Press Button to interupt runica

Interupt

csh

```
Input data size [33,133175] = 33 channels, 133175 frames/nFinding 33 ICA components using extended ICA.
Kurtosis will be calculated initially every 1 blocks using 6000 data points.
Decomposing 122 frames per ICA weight ((1089)^2 = 133175 weights, Initial learning rate will be 0.001, block size
Learning rate will be multiplied by 0.98 whenever angledelta >= 60 deg.
More than 32 channels: default stopping weight change 1E-7
Training will end when wchange < 1e-07 or after 512 steps.
Online bias adjustment will be used.
Removing mean of each channel ...
Final training data range: -171.806 to 179.094
Computing the sphering matrix...
Starting weights are the identity matrix ...
Sphering the data ...
Beginning ICA training ... first training step may be slow ...
step 1 - lrate 0.001000, wchange 16.85061324, angledelta  0.0 deg
step 2 - lrate 0.001000, wchange 0.26760405, angledelta  0.0 deg
step 3 - lrate 0.001000, wchange 0.79058323, angledelta 104.0 deg
step 4 - lrate 0.000980, wchange 0.66700031, angledelta 147.2 deg
step 5 - lrate 0.000960, wchange 0.62849071, angledelta 146.5 deg
step 6 - lrate 0.000941, wchange 0.73967955, angledelta 150.7 deg
step 7 - lrate 0.000922, wchange 0.73727229, angledelta 151.6 deg
step 8 - lrate 0.000904, wchange 0.74051387, angledelta 137.9 deg
step 9 - lrate 0.000886, wchange 0.74536137, angledelta 156.0 deg
step 10 - lrate 0.000868, wchange 0.72101402, angledelta 143.7 deg
step 11 - lrate 0.000851, wchange 0.14690114, angledelta 102.5 deg
step 12 - lrate 0.000834, wchange 0.11822100, angledelta 114.3 deg
step 13 - lrate 0.000817, wchange 0.75552966, angledelta 100.6 deg
step 14 - lrate 0.000801, wchange 0.26739750, angledelta 109.1 deg
step 15 - lrate 0.000785, wchange 0.12123251, angledelta 94.2 deg
step 16 - lrate 0.000769, wchange 0.10285606, angledelta 110.7 deg
step 17 - lrate 0.000754, wchange 0.09770499, angledelta 118.6 deg
step 18 - lrate 0.000739, wchange 0.09544428, angledelta 117.1 deg
```

csh

```
step 241 - lrate 0.000002, wchange 0.00000082, angledelta 101.5 deg
step 242 - lrate 0.000001, wchange 0.00000061, angledelta 96.1 deg
step 243 - lrate 0.000001, wchange 0.00000057, angledelta 97.5 deg
step 244 - lrate 0.000001, wchange 0.00000054, angledelta 93.7 deg
step 245 - lrate 0.000001, wchange 0.00000055, angledelta 100.3 deg
step 246 - lrate 0.000001, wchange 0.00000047, angledelta 96.9 deg
step 247 - lrate 0.000001, wchange 0.00000046, angledelta 91.3 deg
step 248 - lrate 0.000001, wchange 0.00000045, angledelta 101.5 deg
step 249 - lrate 0.000001, wchange 0.00000041, angledelta 103.1 deg
step 250 - lrate 0.000001, wchange 0.00000036, angledelta 95.5 deg
step 251 - lrate 0.000001, wchange 0.00000033, angledelta 92.1 deg
step 252 - lrate 0.000001, wchange 0.00000029, angledelta 97.4 deg
step 253 - lrate 0.000001, wchange 0.00000030, angledelta 95.8 deg
step 254 - lrate 0.000001, wchange 0.00000023, angledelta 94.2 deg
step 255 - lrate 0.000001, wchange 0.00000023, angledelta 97.6 deg
step 256 - lrate 0.000001, wchange 0.00000023, angledelta 97.1 deg
step 257 - lrate 0.000001, wchange 0.00000021, angledelta 92.0 deg
step 258 - lrate 0.000001, wchange 0.00000020, angledelta 99.1 deg
step 259 - lrate 0.000001, wchange 0.00000019, angledelta 95.0 deg
step 260 - lrate 0.000001, wchange 0.00000015, angledelta 98.3 deg
step 261 - lrate 0.000001, wchange 0.00000014, angledelta 99.0 deg
step 262 - lrate 0.000001, wchange 0.00000014, angledelta 94.3 deg
step 263 - lrate 0.000001, wchange 0.00000013, angledelta 95.4 deg
step 264 - lrate 0.000001, wchange 0.00000012, angledelta 94.1 deg
step 265 - lrate 0.000001, wchange 0.00000011, angledelta 96.1 deg
step 266 - lrate 0.000001, wchange 0.00000010, angledelta 94.8 deg
step 267 - lrate 0.000001, wchange 0.00000010, angledelta 94.5 deg
step 268 - lrate 0.000001, wchange 0.00000010, angledelta 97.7 deg
step 269 - lrate 0.000001, wchange 0.00000008, angledelta 95.1 deg
Sorting components in descending order of mean projected variance ...
Permuting the activation wave forms ...
>>
>>
```

Practicum finished!
Thank you for your attention.

Supplementary Material
Ten Q & As about ICA

Easy

1. ICA vs. PCA?
2. IC reject then ICA again to clean data? What is rank?
3. What if I have 10,000 'boundary' events in data?

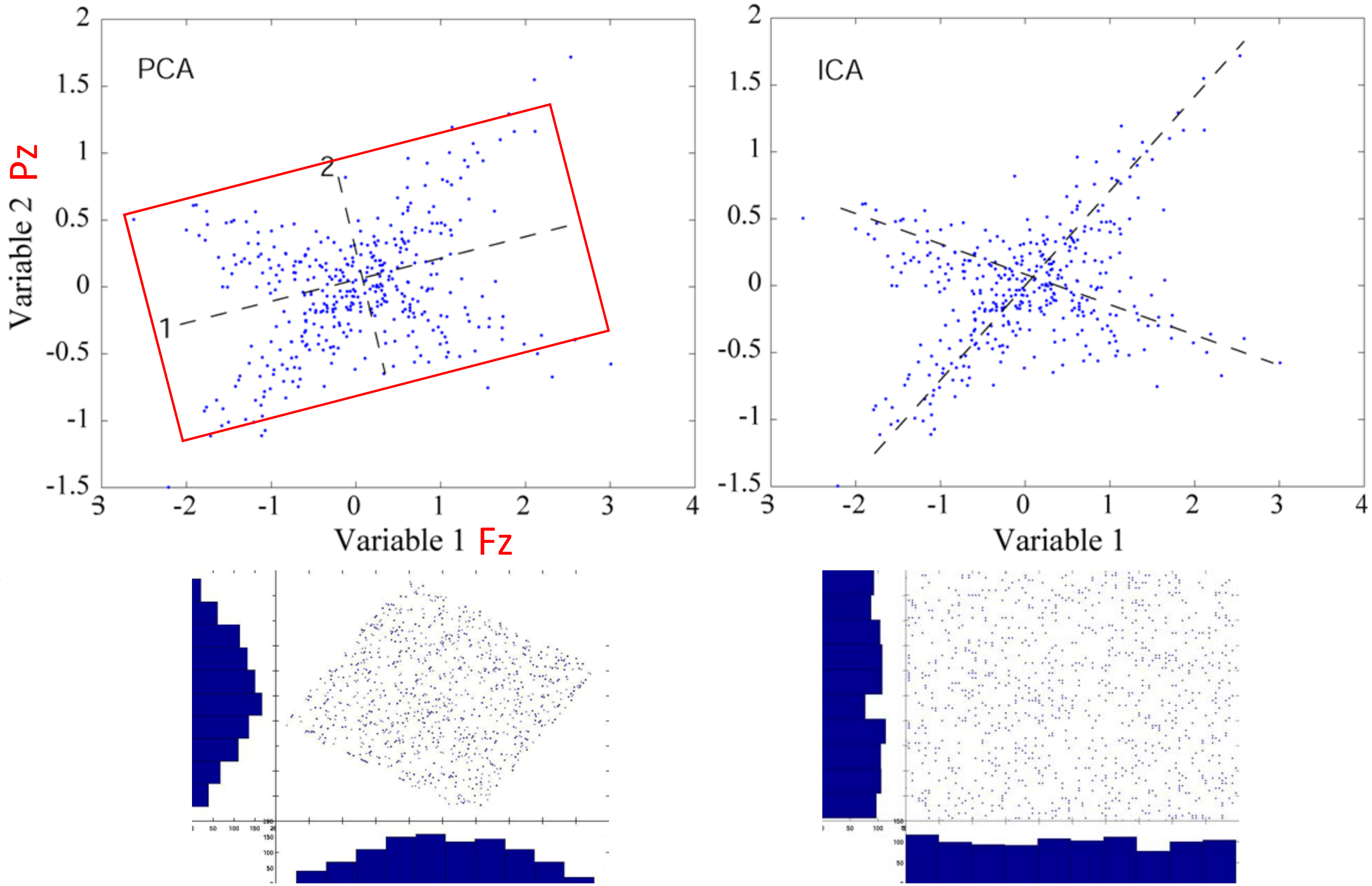
Normal

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. (Why do dipoles tend to be located too deeply?)

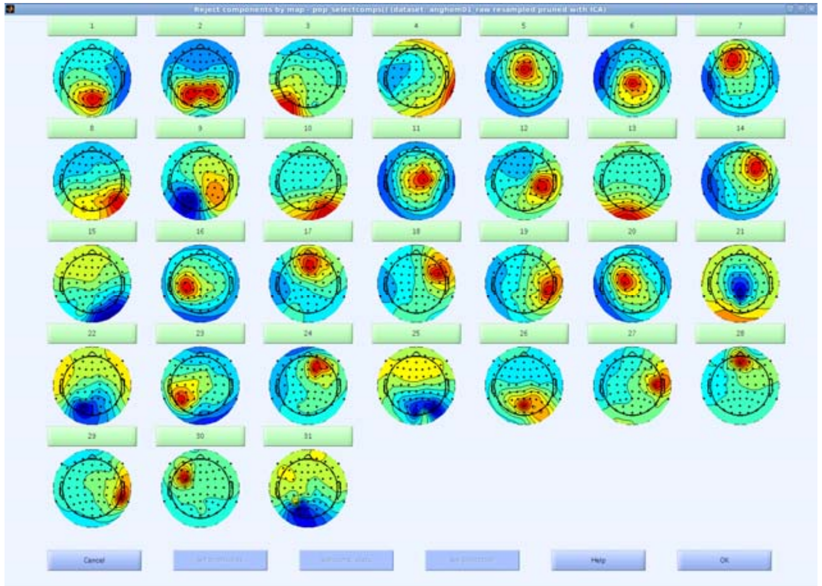
1. ICA vs. PCA?

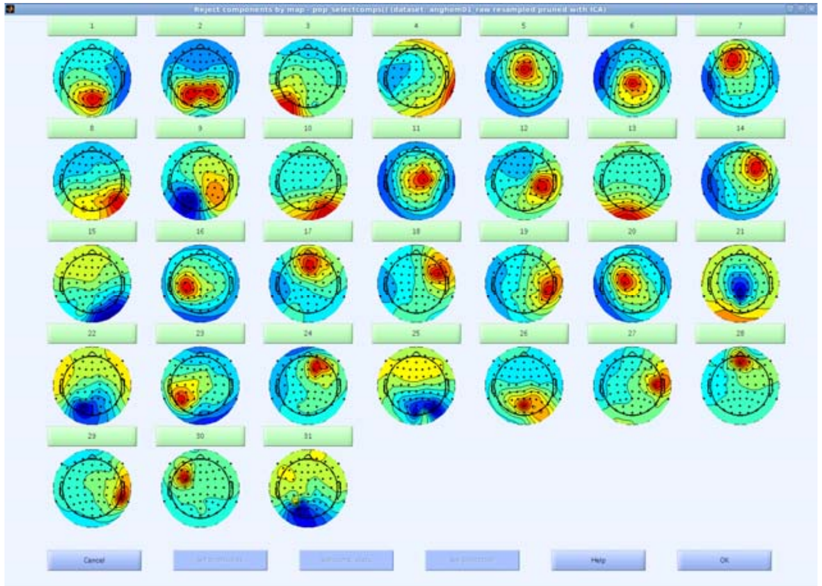


From http://arnauddelorme.com/ica_for_dummies/

See the histogram—make them maximally non-Gaussian and ICA is done!

2. IC reject then ICA again to clean data? What is rank?





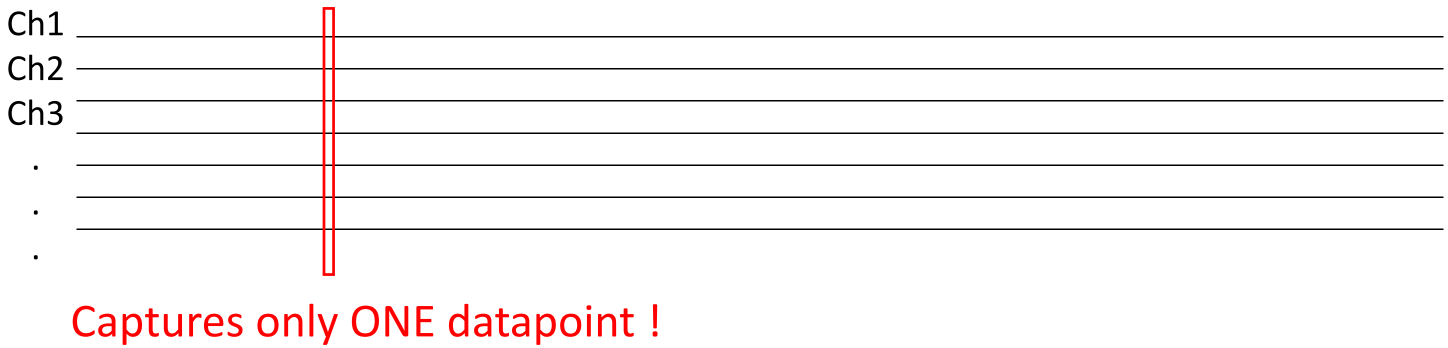
64ch data, 31 ICs selected -> Backproject to channels -> ICA again -> same 31 ICs show up, not the new 64 ICs! This is becaue ICA computes data *rank* to set the number of ICs to calculate. This is a very common misunderstanding, and people expect ICA cleans data in this way.

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

$$2x + 3y - 5 = 0$$
$$3x + 5y + 3 = 0$$
$$4x + 6y - 10 = 0$$

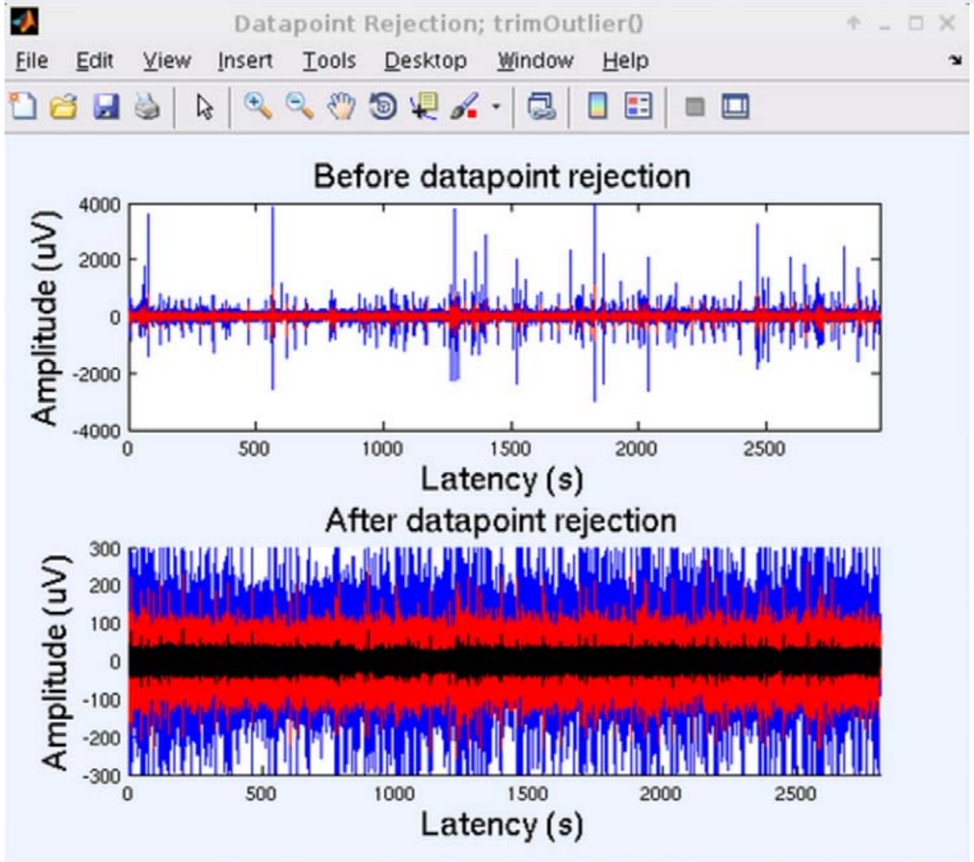
3. What if I have 10,000 ‘boundary’ events in data?

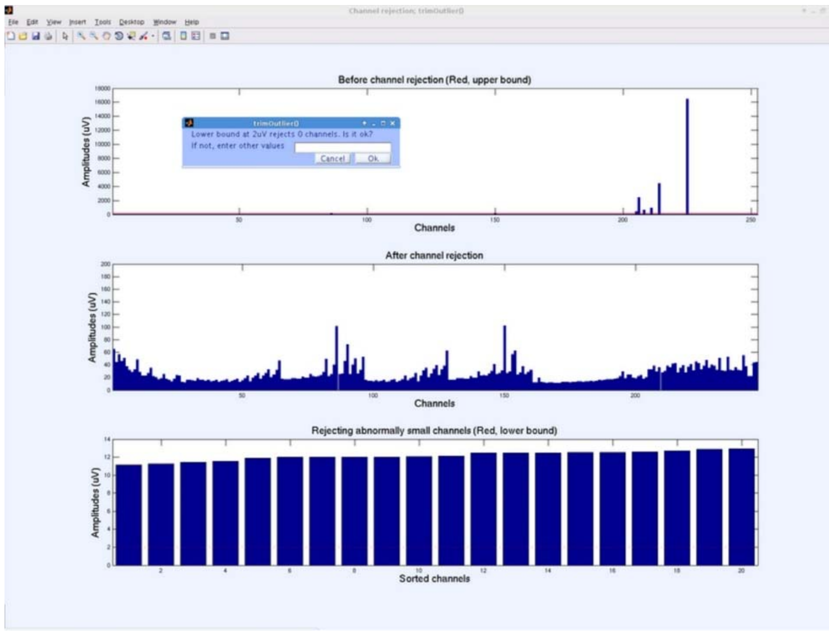
- To ICA, chronologically ordered and randomly shuffled EEG are equivalent. This is becaue ICA processes each time point individually.



- When you start ICA, it DOES randomize 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...

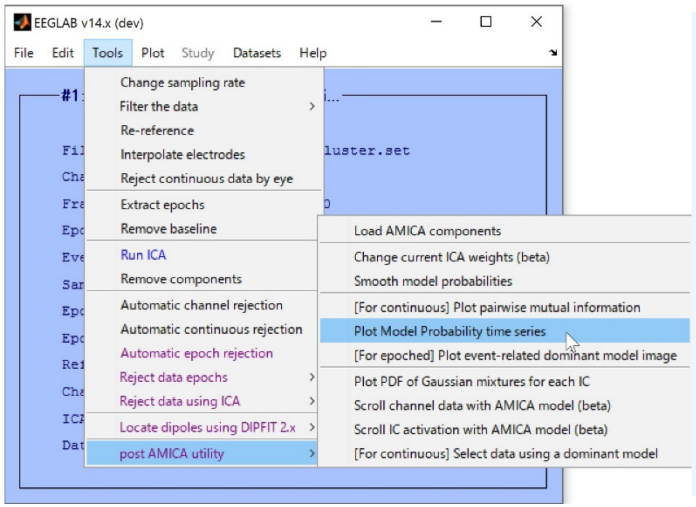


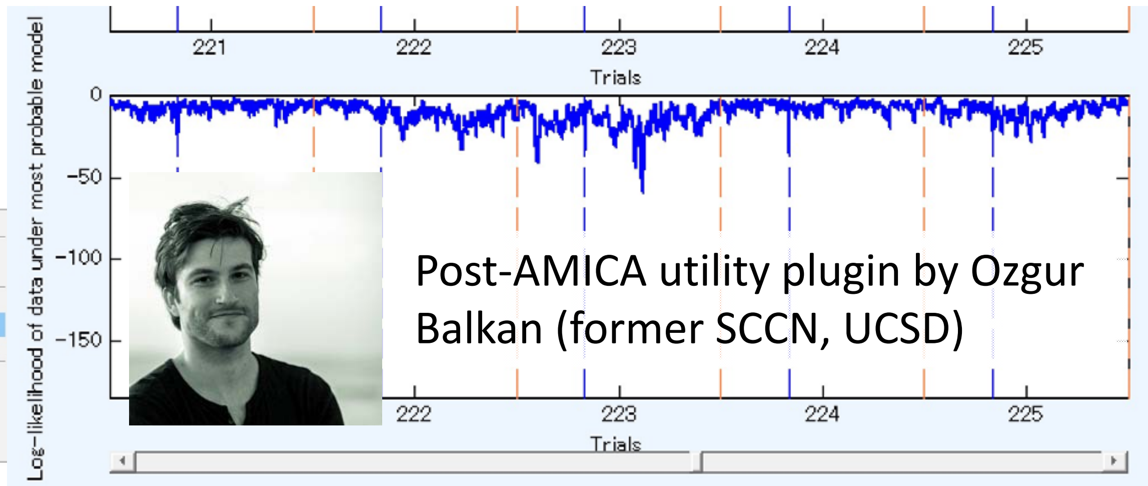


Use *trimOutlier()* to force all-channel EEG to be +/-300 uV, for example, ICA, and apply the weight matrix to midly and decently processed data.

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).





Post-AMICA utility plugin by Ozgur Balkan (former SCCN, UCSD)

Can we measure goodness of fit in infomax?

Infomax ICA [Lee et al. 1999]:

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

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

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

Source Power

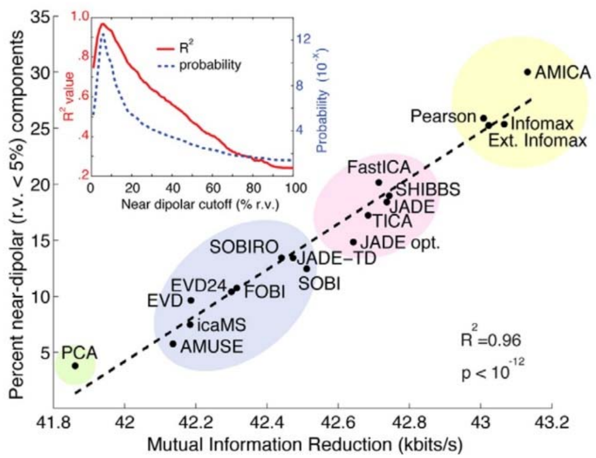
Sheng-Hsiou (Shawn) Hsu
(SCCN, UCSD)

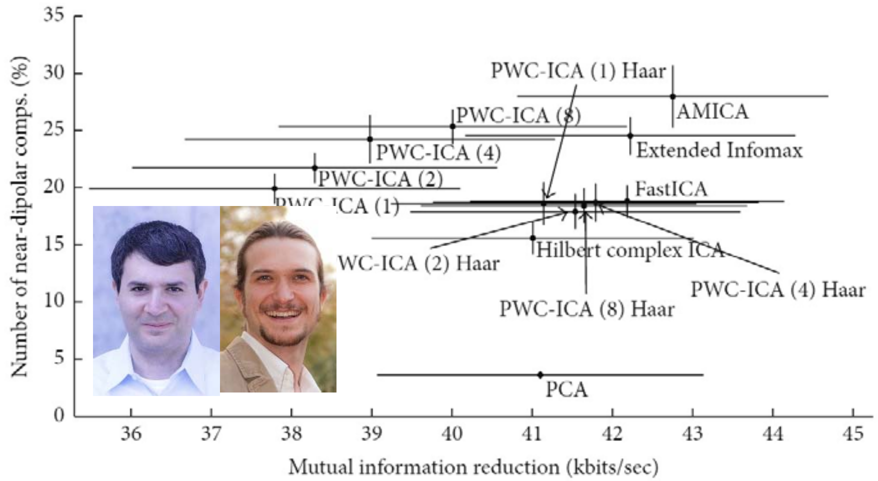
- If it is worthwhile, it can be made into a plugin...?

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

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

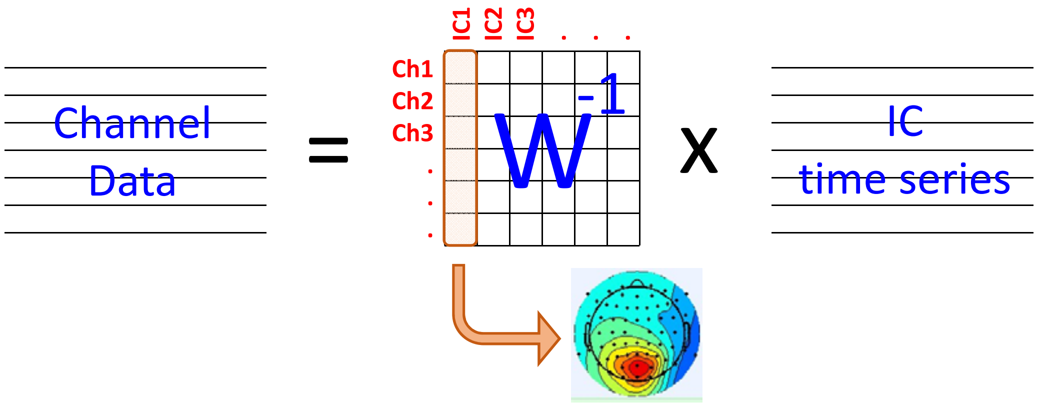
An Amari index of zero for a nonsingular matrix **P** indicates that **P** is a permutation of a diagonal matrix, while an index of one indicates that the entries of matrix **P** are the same constant. Thus, a lower value of the Amari index indicates that **W** is better at isolating the individual sources.

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 exponentially 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

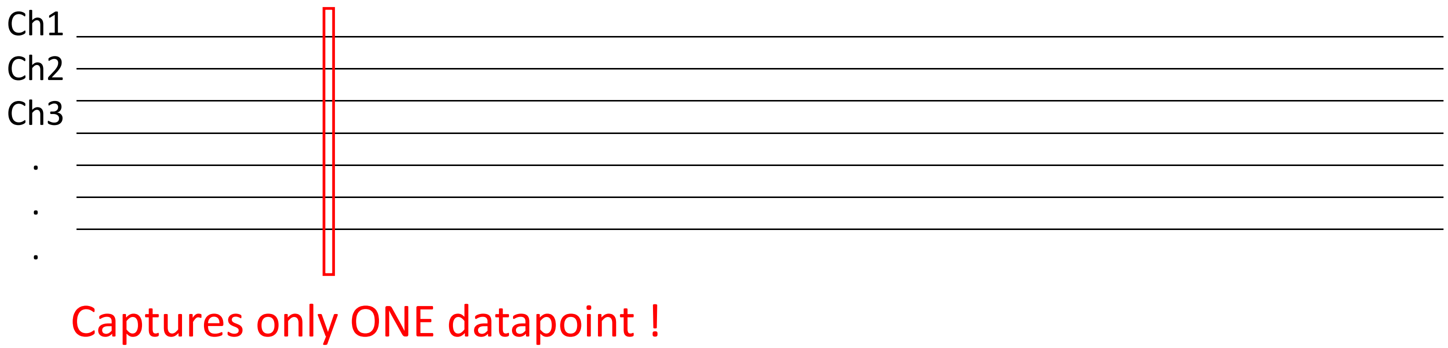
$$\frac{1.5\ \text{cm}}{0.002\ \text{m/ms} \times 100\ \text{cm/m}} / 100\ \text{ms/cycle} \times 360^\circ/\text{cycle} = 27^\circ$$

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.

8. Why do connectivity analyses work?

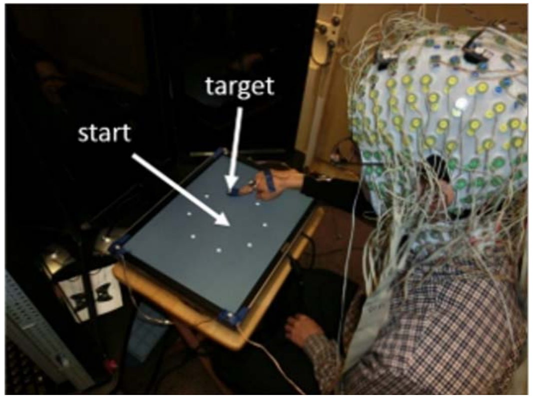
- Again, to ICA, chronologically ordered and randomly shuffled EEG are equivalent. This is becaue ICA processes each time point individually.



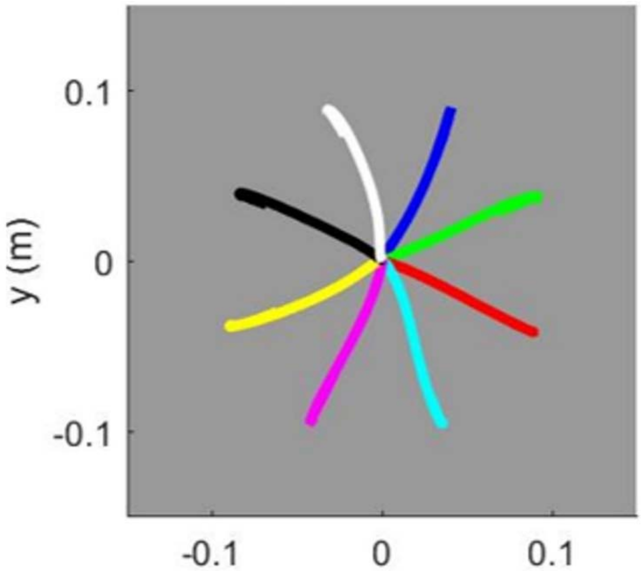
- ICA has no sensitivity to time-delayed dynamics between sources. ICA preserves across-source dynamics, which can be exploited later to study effective connectivity (e.g. Granger Causality).

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

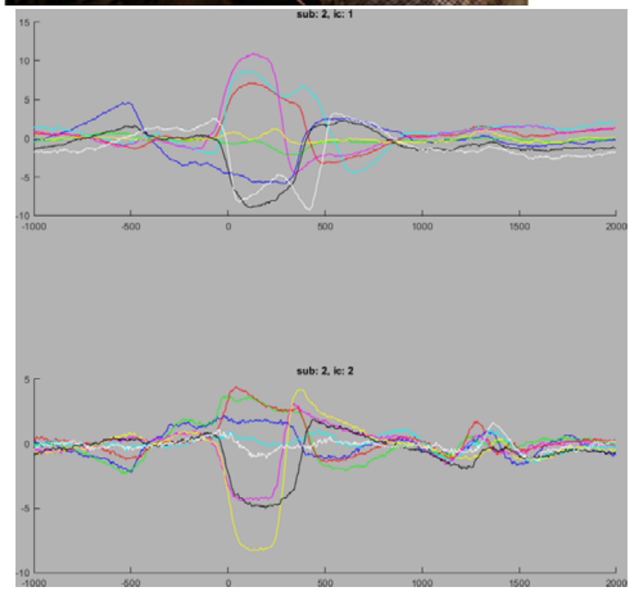
Takana et al. (submitted)
Reaching task using MoBI system at SCCN.



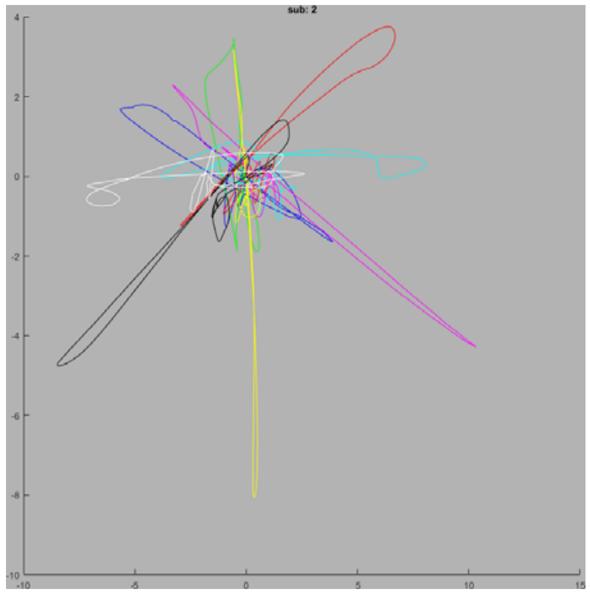
Hirokazu Tanaka (JAIST)
Computational neuroscientist. PhD in theoretical physics.



Trajectory of finger point by motion capture



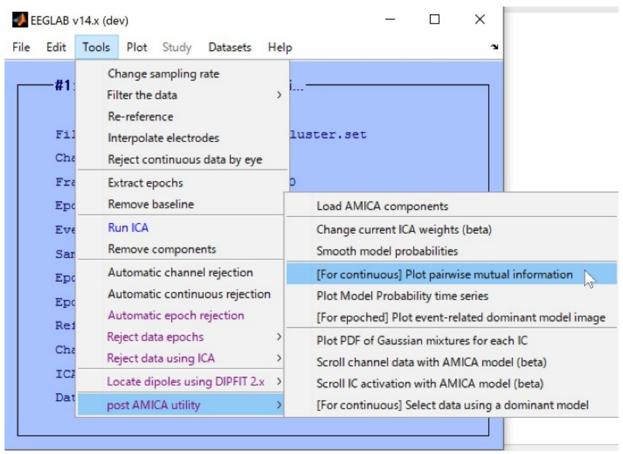
Horizontal and vertical EOGs: IC ERPs



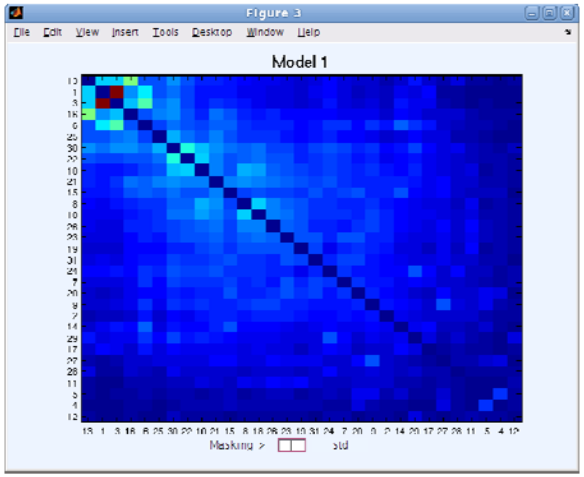
Lissajous figure by the two EOG IC ERPs

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

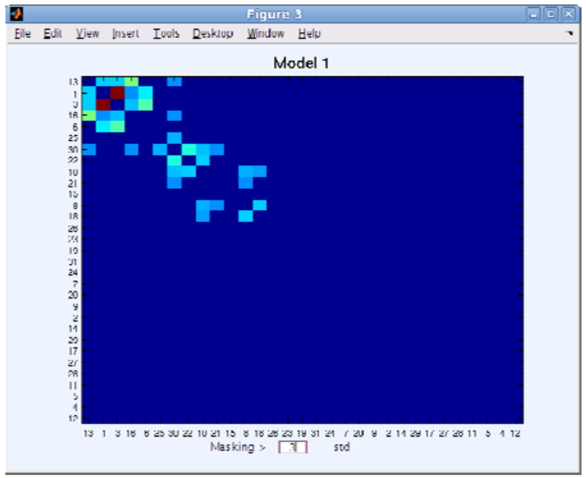
- Independent subspace
- Independent subspace is by definition a group of ICs that are intra-dependent but inter-independent.
- Dependence can be measured by computing pairwise mutual information (PMI) across ICs.



postAmicaUtility() plugin



Pairwise Mutual Information across ICs



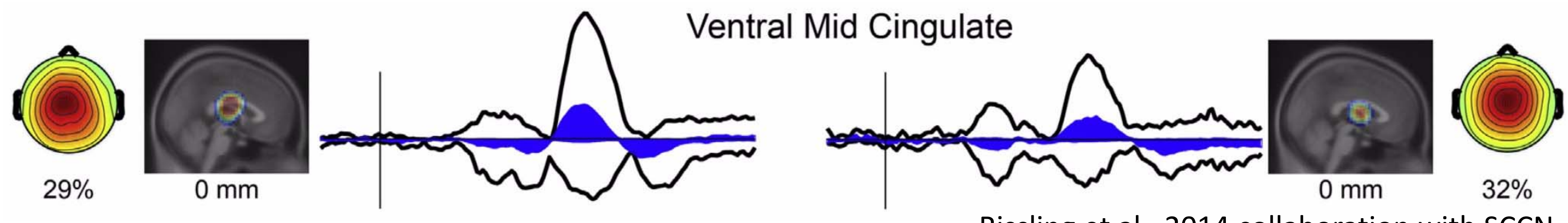
Masked with 3SD. IC1 and IC3 most likely forms subspace.

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.
- On the other hand, we have unpublished simulation study that demonstrated that moving dipole is decomposed into a subspace (Maki, in prep.)
- If subspaces have physiological significance, then eliminating them loses information.
  - Other blind-source separation techniques may fit such non-stationarity more explicitly.
  - Empirically, better signal quality allows ICA to resolve more biologically plausible independent components/subspaces.
- This is an open question.

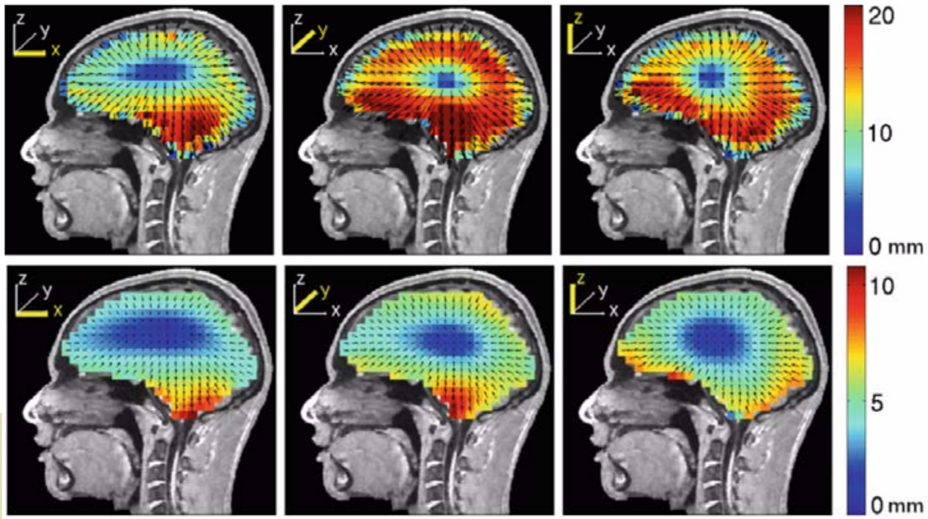
Thank you!

10. Why do dipoles tend to be localized too deeply?



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 when conductivity values used should err towards shallow dipoles.



Zeynep Akalin Acar
Forward-model specialist

- This is an open question.

How to interpret too deep dipoles in practice?

- There seems patterns that ICs near mid-cingulate or inferior occipital regions tend to be localized subcortically.
- If the bias toward depth comes from some systematic structure that work as a stable filter, across study consistency could be still maintained.
- In reporting, I suggest one report them as ‘Upper/Lower basal’ without mention spacific namesm such as ‘caudate’ or ‘amygdala’. But do report their coordinates, becaue coordinates are computational useful and necessary for future studies with solutions for this problem.
- When estimating the cortical source locations, pull them out radially to the cortical surface (i.e. Corpus callosum -> midcingulate cortex.)