Comparative evaluation of Independent Components Analysis algorithms for isolating target-relevant information in brain-signal classification

N. Jeremy Hill,[1] Michael Schröder,[2] Thomas Navin Lal,[1] and Bernhard Schölkopf[1]

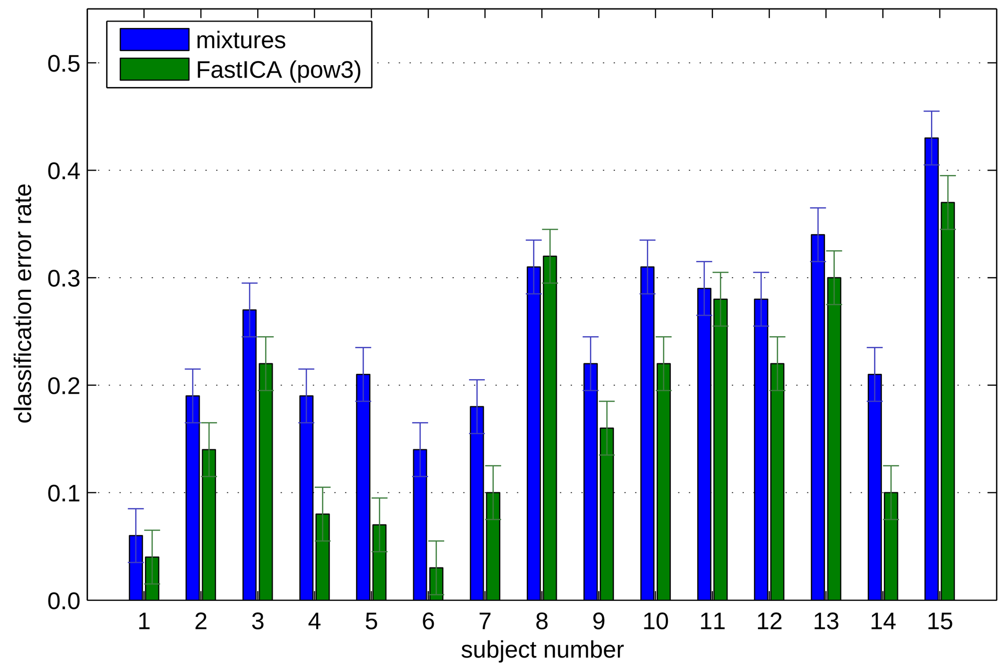
[1] Max Planck Institute for Biological Cybernetics, Tübingen
[2] Eberhard Karls University, Tübingen

MAX-PLANCK-GESELLSCHAFT

BIOLOGISCHE KYBERNETIK

1 Introduction

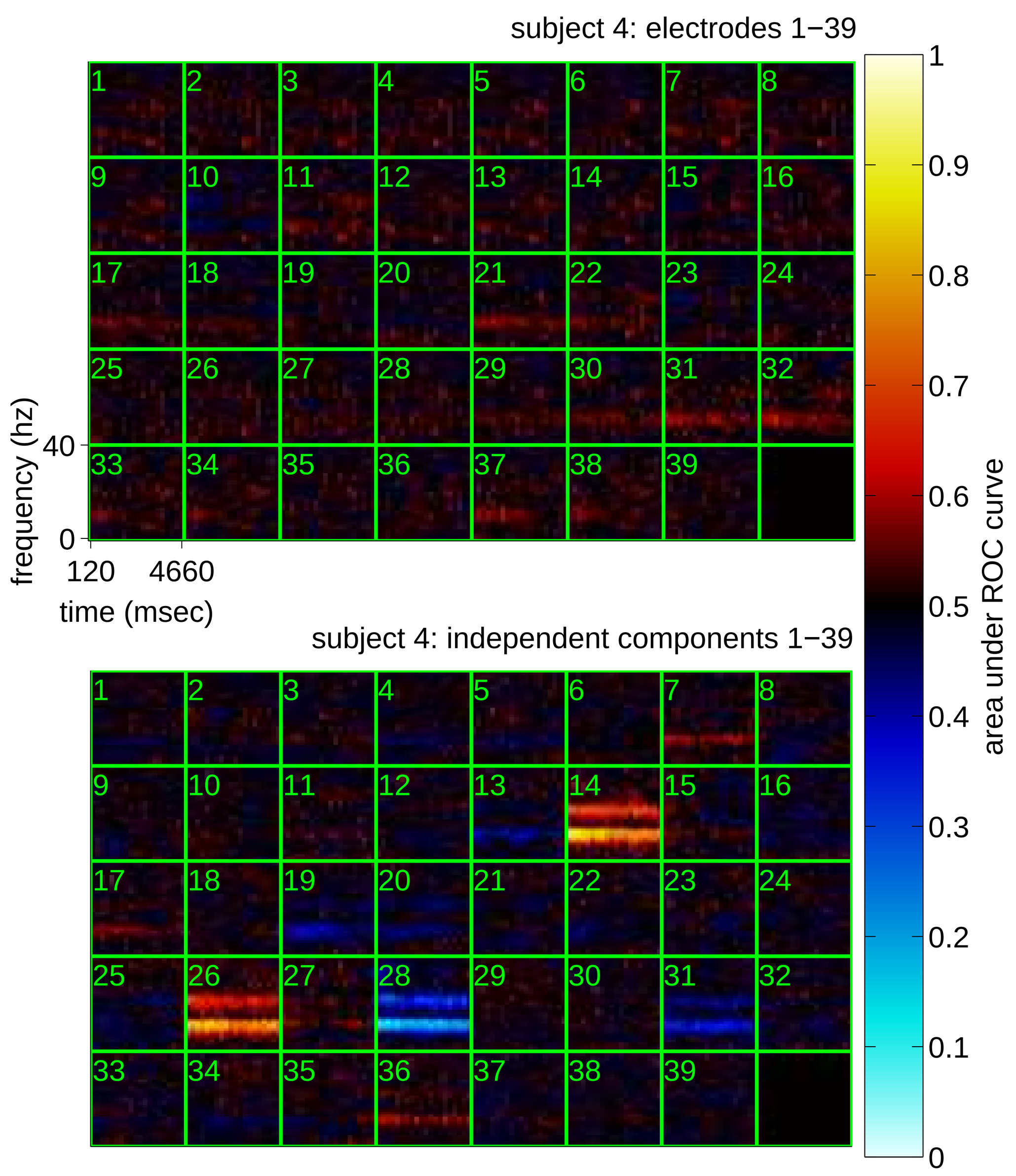
Like many other researchers, we have previously found that blind source separation using Independent Components Analysis (ICA) can significantly improve classification performance in single-trial brain signal classification. The following figure illustrates the data presented by Hill et al. (2005) [Advances in Neural Information Processing Systems 17, 569–576], in which auditory ERPs were classified in a binary (left-vs-right) attention-based BCI.



Note that, for 14 out of 15 subjects, the error rate drops significantly and considerably (by 5–10%).

Here we present further quantitative results and a comparison of different ICA variants by re-examining the left-vs-right-hand motor-imagery data of Lal et al. (2004) [IEEE Trans. Biomed. Eng. 51, 1003–1010]. The previous authors' methodology of auto-regressive (AR) models, Support Vector Machine classifiers (SVMs) and Recursive Feature Elimination (RFE) is retained, but here we illustrate the improvements in classification error rate and feature elimination that can be achieved by combining this with ICA.

The following spectrograms give a feel for what ICA can do in a motor-imagery paradigm. Note that the μ bands have been isolated into a smaller number of channels, and considerably denoised. In particular the higher-frequency band (around 20 Hz) is now visible, whereas it could not be seen as target-correlated at all in the individual electrode traces.



2 EEG Input

- Data from Lal et al. (2004): Binary motor imagery paradigm (left vs. right hand): 400 trials x 39 EEG channels x 5 seconds.
- Analog band-pass filter 0.1–40 Hz, digitize at 256 Hz
- Digitally low-pass filter and downsample to 100 Hz

3 Analysis

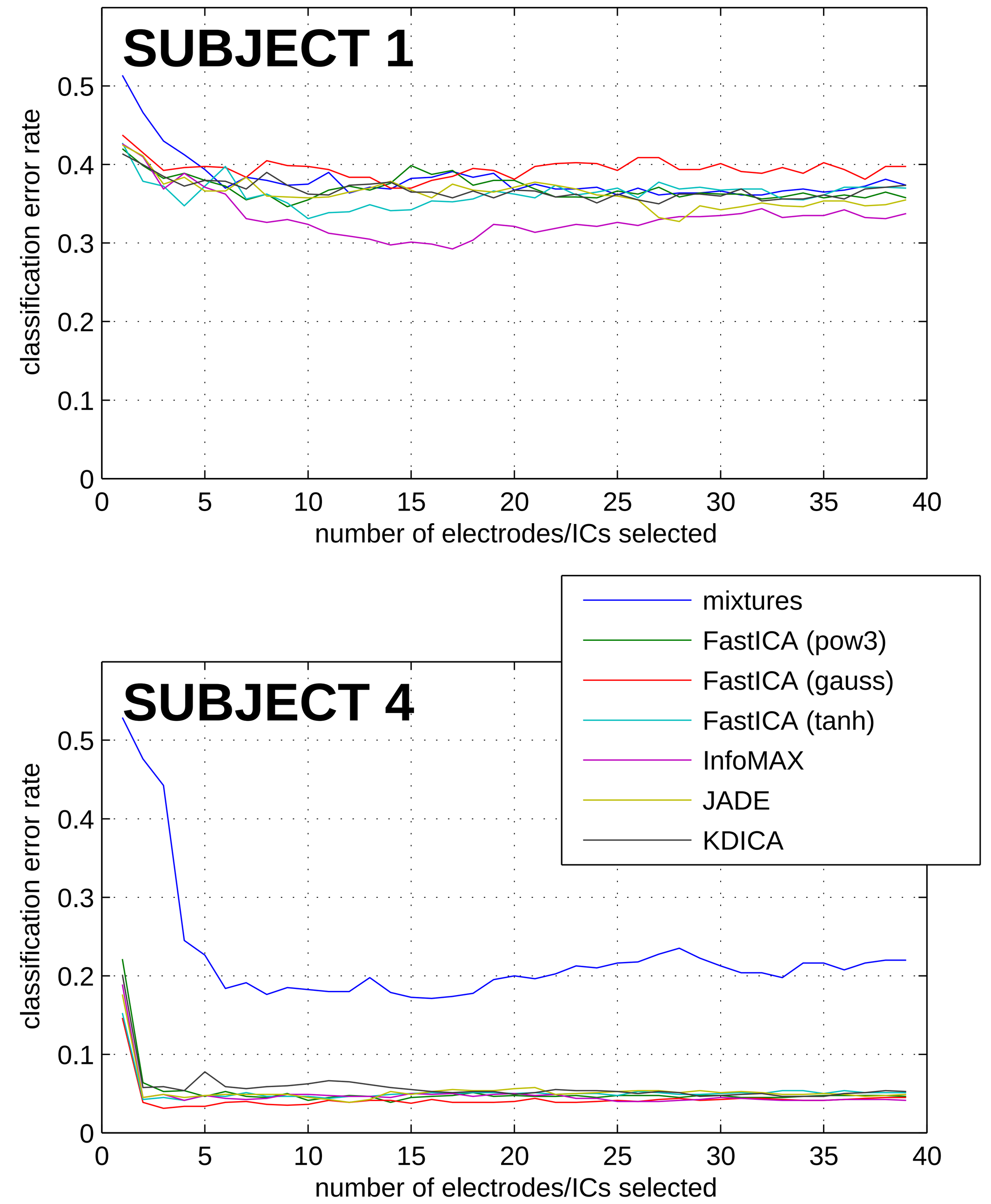
```
repeat twice with different random seeds:
 1 randomly split trials into 10 non-overlapping test folds
   for each fold:
     with training fold:
       2 Concatenate trials to form 39 long time series. Compute ICA demixing matrix using every 10th time sample. Demix channels into independent components (ICs). Cut the time series up into trials again.
       3 Fit least-square forward-backward linear AR model of order 4 to each IC.
       4 Train linear ν-SVM on the 39x4 coefficients, optimizing ν in {0, 0.1, ..., 0.9} by cross-validation with 10 sub-folds.
       5 Perform Recursive Independent Component Elimination (i.e. Recursive Feature Elimination with one whole IC eliminated at each step) on all training data to obtain a rank order of ICs.
       6 Cross-validate RICE with 10 sub-folds to obtain estimated error rates for each number of ICs.
       7 Find the lowest number of ICs at which the CV error is within 2 standard errors of the minimum.
       8 Reduce number of ICs to this number, taking them in the rank order obtained in step 5.
       9 Re-optimize ν to produce final trained classifier for this fold.
     end with
     with test fold:
      10 Apply ICA separating matrix computed from training fold.
      11 Retain only the ICs selected by RICE in step 8.
      12 Compute AR coefficients.
      13 Test trained classifier on test fold and record test error (final error).
      14 Also record test error estimates during step 5 (elimination error trace).
     end with
   end for
end repeat
```

4 Output

- Final error score from step 13 (averaged across 2x10 test folds): an estimate of the potential online performance of the algorithm.
- Elimination error traces from step 14 (also averaged across 20 folds): the minimum of this curve gives the error that might be obtained online if the algorithm knew exactly how many ICs to select (which it does not, but the shape of the curve is instructive).

We gratefully acknowledge the financial support of the Max-Planck-Gesellschaft, and of the Deutsche Forschungsgemeinschaft under grant #RO1030/12.

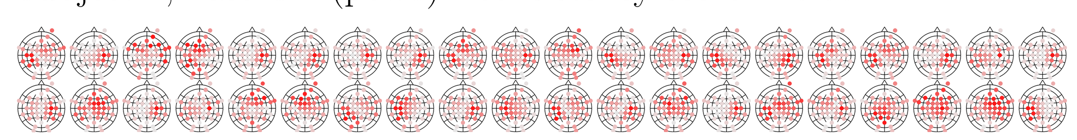
5 Example results: elimination error traces



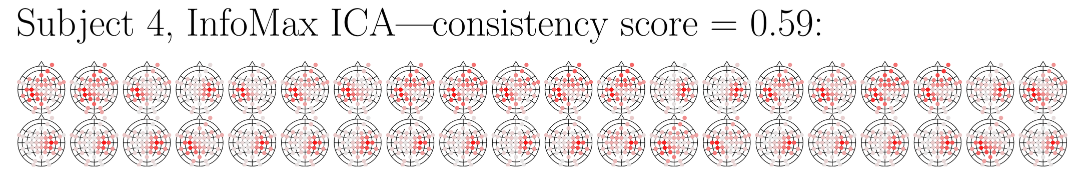
6 Example results: consistency of ICA separation

The following contrast illustrates how, depending on the subject, some ICA algorithms are more consistent than others. In each of the 20 different folds (left to right), the ICA algorithm sees a slightly different subset of the input. For the best subjects (3 and 4), it is remarkable how much more consistent the mixture weights are across folds when InfoMax ICA is used, as opposed to any of the other ICA variants tested. The figures below show the 2 top-ranked ICs from each fold (hence, two rows in each).

Subject 4, FastICA (pow3)—consistency score = 0.33:

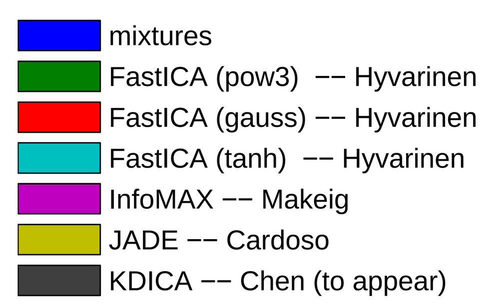


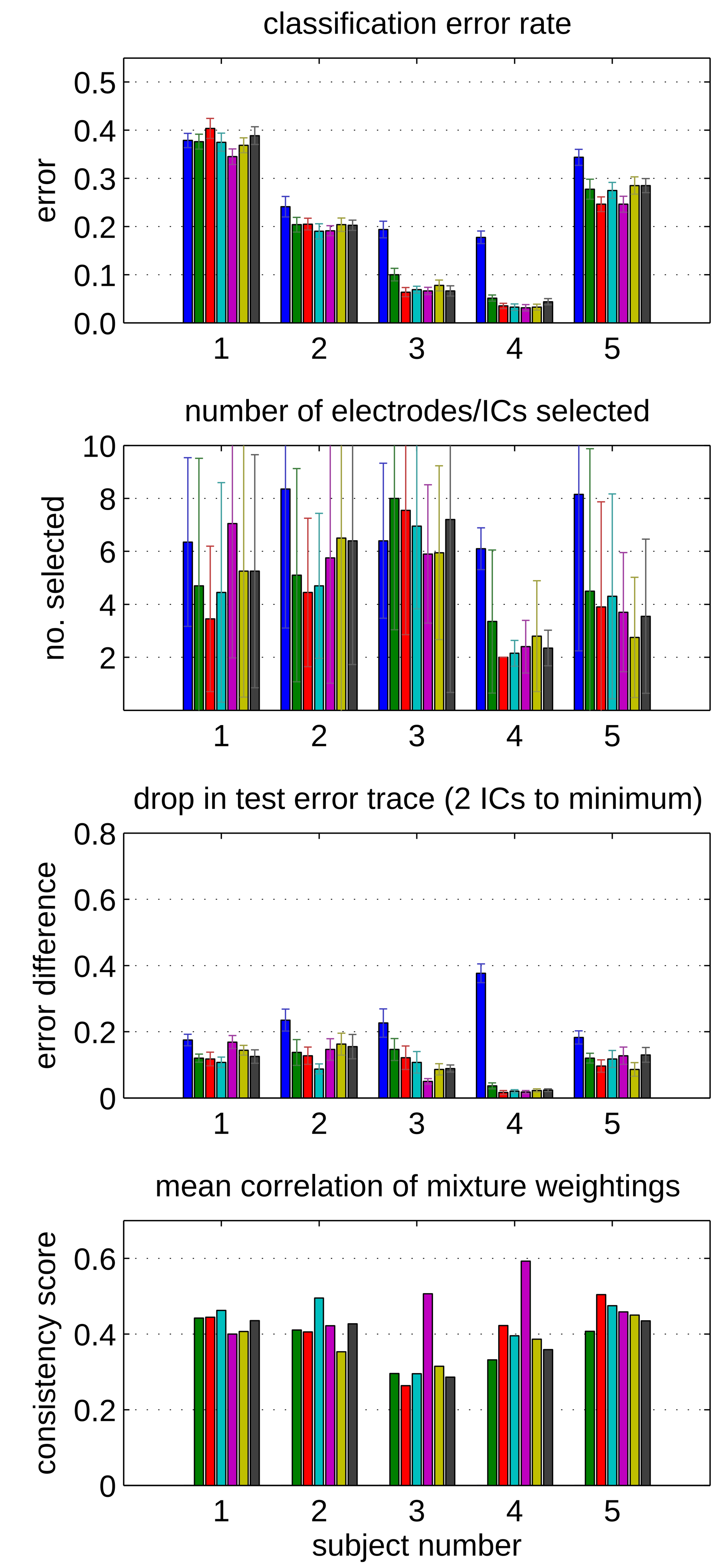
Subject 4, InfoMax ICA—consistency score = 0.59:



7 Overview of results

The charts to the right compare the performance of different ICA algorithms, which differ according to the *contrast function* they use to define independence. The algorithms are colour-coded as shown.





- For most subjects and all ICA variants, ICA improves performance considerably (up to 97% correct). Our claim in the submitted abstract, that InfoMax is significantly better than the others in this respect, is erroneous and was based on a preliminary subset of the data. Over all performance measures, however, InfoMax compares very favourably with the others.
- Depending on subject and ICA variant, ICA interacts more or less consistently with Recursive Feature Elimination. For example, for subject 4, FastICA (gauss) results in exactly 2 ICs being selected on every fold, whereas for JADE or FastICA (pow3), the number selected is very variable.
- Again depending on subject, ICA variants can be differentially sensitive to variation in the signal subsample. InfoMax returns remarkably consistent mixture weightings, at least for the better subjects, and therefore yields potentially the most interpretable results.
- Independence between time-shifted signals was not considered. Preliminary results with a variant of SOBI (not shown) have so far been very poor, but comparison with established good performers such as TDSEP will be interesting.