Detecting Semantic Category in
Simultaneous EEG/MEG Recordings

Brian Murphy
Centre for Mind/Brain Sciences
University of Trento
corso Bettini 31,
38068 Rovereto, Italy
`brian.murphy@unitn.it`

Massimo Poesio
Centre for Mind/Brain Sciences
University of Trento
corso Bettini 31,
38068 Rovereto, Italy
`massimo.poesio@unitn.it`

Abstract

Electroencephalography (EEG) and magnetoencephalography (MEG) are closely related neuroimaging technologies that both measure summed electrical activity of synchronous sources of neural activity. However they differ in the portions of the brain to which they are more sensitive, in the frequency bands they can detect, and to the amount of noise to which they are subject. Since semantic representations are thought to be widely distributed in the brain, this preliminary study considered if the broader coverage offered by simultaneous EEG/MEG recordings would increase sensitivity to these cognitive states. The results showed that MEG data allowed stimuli in two semantic categories (mammals and tools) to be distinguished more accurately, despite some experimental settings that were optimised for EEG. The addition of EEG data did not prove informative, indicating that it may be redundant relative to MEG, even when using dimensionality reduction techniques to combat overfitting.

1 Introduction

Electroencephalography (EEG) and magnetoencephalography (MEG) are similar methods for recording activity in the brain. Both detect signals that are produced by the mixing of neural sources, where each source represents macro-scale synchronisation between the firing of individual neurons. The sum of these activities induce voltages at the scalp that are recorded with EEG, and magnetic fields that are detected with MEG. But the signals yielded by each technique are not identical for several reasons. EEG signals are heavily attenuated and filtered (both in time in space) by the passage through skull and tissue. As a result, MEG signals are less noisy, have finer spatial resolution, capture a wider range of frequencies, and so have the potential to be more informative. Further, the signal footprint of MEG and EEG signals on the brain is not the same: EEG sensors are more sensitive to currents that are radial to the scalp and so predominantly detect activity in the at the top of gyri and the bottom of sulci (the top and bottom of folds in the surface of the brain); while MEG is more sensitive to currents that are tangential to the scalp, and so detects more activity in the side walls of sulci. The high spatial resolution of MEG means that it cannot see as deeply into the brain as EEG can. Finally, MEG sensors of different types (in this case magnetometers and planar gradiometers) are sensitive to magnetic fields of different orientations (see Figure 1): planar gradiometers are most sensitive to current generators of a particular orientation directly under the sensor position; magnetometers record generators that are tangential and peripheral to the sensor area.

The distribution of sensor coverage may be important for the decoding of semantic categories in particular. Neuroimaging evidence suggests that semantic representations may be widely distributed in the brain. For example there are well-established differences in neural activity in the fusiform gyrus that correspond to higher level categories (natural vs non-natural kinds; people vs places - see e.g. Chao et al., 2002); there is also evidence that the

Figure 1: Schematic from above of selective sensitivity of three co-located MEG sensors
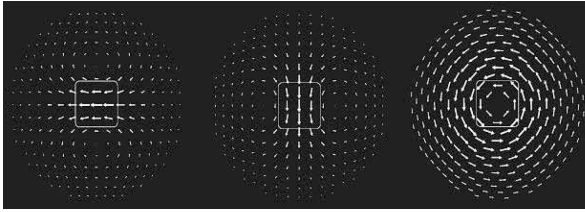
Left and centre panels show perpendicular planar gradiometers; right panel shows magnetometer. A co-located EEG electrode would be most sensitive to currents perpendicular to the scalp. Image courtesy of Elekta AB.

meaning of bodily actions is encoded in the motor-cortex (Pulvermüller, 2005); and concepts associated with eating (e.g. foodstuffs) seem to be represented at least in part by activations in gustatory cortex (Mitchell et al., 2008; Just et al., 2010). Hence a wide coverage of sensors that are sensitive to different but overlapping portions of brain tissue may provide a fuller description of semantic memories.

Given the fact that it has been possible to decode conceptual categories and language semantics from EEG signals (Murphy et al., 2008, 2009), the question is if MEG signals can be shown to be more informative. Similar studies on lower-level tasks typically used in brain-computer interfaces suggests that it may be: Hill et al. (2006) find that there is a modest increase in the decoding accuracy on imagined motor activity with MEG, relative to EEG, and Waldert et al. (2008) have similar findings detecting the direction of hand movements.

A related question is whether the information supplied by EEG and MEG is complementary, and if so how best it should be combined. This depends critically on the number of signals used: raising the number of input signals increases the information supplied to the machine learning methods, but interacts with their tendency to overfit, if the number of descriptive dimensions (recorded signals) is of a similar order of magnitude to the number of training cases (experimental trials in which a stimulus is presented). This is often the case with data from neuroimaging experiments, as there are practical limitations on the number of data points that can be collected: individual stimuli must usually be separated

by several seconds so that neural signals can return to baseline between each, and participants can usually only be expected to perform a task at full attention for 60 minutes or so, in such experimental environments.

To investigate this question, we replicated an existing EEG experiment (Murphy et al., 2010). In that experiment participants had been presented with images of animals and tools, while EEG activity was recorded at 64 standard 10-10 locations, and single trials (stimulus presentations) could be classified as representing the category of animal or tool with an average accuracy of 72% over all seven participants. The classification methods used were an adaptive time/frequency window optimisation (Dalponte et al., 2007), a supervised spatial component signal decomposition (Common Spatial Patterns, Koles et al., 1990) that yielded measures of neural activity based on signal power, and a support-vector machine (Boser et al., 1992).

The replication experiment reported here was carried out with two participants, and used the same task and materials, while simultaneously recording with a 306-channel MEG system (204 gradiometers, 102 magnetometers) and a high-density 124-channel EEG system. This data was then analysed using the same machine learning methods as previously, but varying the number and type of input signals, and using dimensionality reduction to address increased dimensionality.

2 Methods

2.1 Experiment and Materials

Two male native speakers of Italian took part in the study, aged 30 and 47. Both were right-handed with corrected or normal vision. Participants in this study receive compensation of 7 euros per hour. The experiment is conducted under the approval of the ethics committee at the University of Trento, and participants gave informed consent.

The participants were asked to perform a silent naming task on grey-scale images of 30 land-mammals and 30 work tools. Each stimulus was presented between four and six times, in randomised order.[1] The participants sat in a relaxed upright posi-

---

[1] Participant 1 saw 264 stimulus trials (144 mammal and 120 tool trials); participant 2 saw 360 (180 in each class).

tion 1.5m from a projector screen in moderate lighting conditions. Images were presented on a medium grey background and fell within a 6 degree viewing angle. The task duration was split into five blocks and the participants were given the choice to pause between each. The cumulative task time did not exceed 45 minutes.

Each trial began with the presentation of a fixation cross for 0.25s, followed by the stimulus image, a further fixation cross for 0.75s and a blank screen for 1s. Participants were instructed to silently name the object represented in their native tongue (Italian), using the first appropriate label that came to mind, and to press the keyboard space-bar with the left-hand to indicate they had found an appropriate word. If the participant could not think of a suitable label, they were asked not to make a response. The image remained on the screen until the participant responded, or until a time-out of three seconds was reached. The participants were asked to keep still during the task, and to avoid eye-movements and facial muscle activity in particular, except during the blank period.

The materials were chosen to represent well-defined semantic categories and to minimise non-semantic, associative confounds. The set of 30 land mammals were chosen to be both non-domesticated and non-threatening, to avoid emotional valence whether positive (e.g. pets) or negative (e.g. predators). Thirty hardware and garden implements were chosen as genuine work tools. Appropriate photographs were sourced from the internet, and normalised visually: each image file measured 300 pixels square; the image proper was converted to grey-scale, superimposed on a homogeneous light-grey background and had maximal horizontal and vertical dimensions of 250 pixels; image contrast was normalised. The concepts represented are listed below.

Land Mammals ant-eater, armadillo, badger, beaver, bison, boar, camel, chamois, chimpanzee, deer, elephant, fox, giraffe, gorilla, hare, hedgehog, hippopotamus, ibex, kangaroo, koala, llama, mole, monkey, mouse, otter, panda, rhinoceros, skunk, squirrel, zebra (*Italian* formichiere, armadillo, tasso, castoro, bisonte, cinghiale, cammello, camoscio, scimpanz, cervo, elefante, volpe, giraffa, gorilla,

coniglio, riccio, ippopotamo, stambecco, canguro, koala, lama, talpa, scimmia, topo, lontra, panda, rinoceronte, puzzola, scoiattolo, zebra)

Work Tools Allen key, axe, chainsaw, craft-knife, crowbar, file, garden fork, garden trowel, hacksaw, hammer, mallet, nail, paint brush, paint roller, penknife, pick-axe, plaster trowel, pliers, plunger, pneumatic drill, power-drill, rake, saw, scissors, scraper, screw, screwdriver, sickle, spanner, tape-measure (*Italian* brugola, ascia, motosega, taglierino, piede di porco, lima, forcone, paletta, seghetto, martello, mazza, chiodo, pennello, rullo, coltellino svizzero, piccone, cazzuola, pinza, stura lavandini, martello pneumatico, trapano, rastrello, sega, forbici, spatola, vite, cacciavite, falce, chiave inglese, metro)

2.2 Neural Recordings

The experiment was conducted at the LNiF imaging laboratories at the University of Trento, using a 306-sensor Elekta Neuromag system (2 planar gradiometers and 1 magnetometer at each of 102 sensor locations). A dense-coverage 124-electrode EEG cap was used also, using a right mastoid reference and forehead ground. Both sets of signals were recorded simultaneously at 1000Hz in a magnetically shielded room. At the start of the session the relative positions of the MEG and EEG sensors were determined using a Polyhemus 3-D digitisation system.

Data preprocessing was conducted using the MNE, FieldTrip and EEGLAB packages.[2] The data was band-pass filtered at 1-50Hz to remove slow drifts in the signal and high-frequency noise, and then down-sampled to 125Hz. Eye and muscle artefacts were not removed, but these lie outside the range of frequencies that were considered in the analysis described below.

---

[2]Martinos Centre for Biomedical Imaging (http://www.nmr.mgh.harvard.edu/martinos/); Donders Institute for Brain, Cognition and Behaviour (http://www.ru.nl/neuroimaging/fieldtrip); and Schwartz Center for Computational Neuroscience (http://sccn.ucsd.edu/eeglab/) respectively.

2.3 Analysis

The analysis method first applies a time/frequency filter to select an information-rich band and interval for the distinction of interest; a supervised decomposition to extract components of whole-scalp synchronous activity that are sensitive to this class distinction (Common Spatial Patterns, or CSP – see Parra et al., 2005; Model and Zibulevsky, 2006; Philiastides et al., 2006 for examples of other applications to cognitive neuroscience); and a general purpose machine learning algorithm (Support-Vector Machine or SVM) that uses the resulting measures of signal power to predict the semantic class of each trial. Individual trial epochs are arbitrarily allocated to one of k interlaced partitions of equal size in a k-fold training/evaluation procedure.

The time/frequency filter applied here was adopted from the earlier experiment, as it had been found to provide optimal separation between trials of the two classes over the participants of that study. Using this common window (4-18Hz, 95-360ms after image onset) allows direct comparison between the informativity of each type of sensor, or combination of sensor types. However this may disadvantage MEG, since it is more sensitive to higher frequency activity (> 50Hz), which at least one study has found to vary systematically with semantic classes (Tanji et al., 2005).

The decomposition method used, CSP (Koles et al., 1990), extracts spatial components of electrophysiological activity (linear combinations of raw signals) that correspond to synchronous neural subassemblies. It is a supervised technique that yields signals whose level of activity (measured as signal power) is modulated by the binary class distinction of interest – that is signals that show high power when processing mammal concepts, and low power when processing tool concepts, or vice-versa. CSP identifies C components (where C is the number of input channels) that are ranked by their sensitivity to the class-separation of interest, in terms of optimal variance for the two populations of signals (i.e., high variance between classes and low variance within classes). In this case we selected the first and the last rows of this matrix (Ramoser et al., 2000) as the components that are most representative for the classes mammals and tools, respectively.

This procedure can be interpreted as extracting the event-related spectral activity (i.e. the relative event-related synchronisation) of two synchronous neural structures which have been found to have an optimally differential response to the semantic categories of interest.

The final categorisation step is based on a Support-Vector Machine (SVM) classifier (Boser et al., 1992; Vapnik, 1998). The input for each trial consisted of two measures of neural activity extracted from the category-sensitive signal components: the variance of the waveform, which is proportional to signal power. The features were further normalised by taking the log, and scaling to a range of -1 to +1 across all trials. The SVM implementation used was LIBSVM (Chang and Lin, 2001), and default parameters were used (radial basis function kernel, cost parameter of 1, and a gamma value of the inverse of the number of data-points).[3] Test and training data were kept strictly separate at all stages of analysis.

In the results that follow here, these techniques were first applied as before to replicate the previous experiment, but then also with an additional step of dimensionality reduction to address the overfitting we expected given the dramatically larger number of input channels (up to 430 if all EEG and MEG channels were used, compared to 64 channels in the previous experiment). The signal recorded in any individual channel will be comprised of a mix of genuine neural activity (both relevant and irrelevant to our classification task), systematic noise sources (e.g. 50Hz electrical line noise, eye-movement artefacts, heart-beat artefacts), and additional random noise. And as EEG and MEG channels record activity from partially overlapping portions of brain tissue, there is considerable redundancy between neighbouring channels. Principle Components Analysis (PCA) is a dimensionality reduction technique that addresses both these issues, grouping redundant activity into the first (strongest)

---

[3]No optimisation of SVM parameters was attempted, as extensive parameter testing in the earlier experiment did not yield any improvements in classification performance. We believe that this is because CSP is in itself a powerful data-mining technique, that here typically yields two simple clusters of data corresponding to each semantic category. We expect a simple linear classifier would have similar performance on this task.

components, and relegating random noise to the last components. Where PCA was used, it was applied directly before the CSP-based extraction of category-specific sources.

3 Results

In the previous EEG experiment, the classification accuracy averaged 72%, but varied substantially from one participant to the next, ranging from 56% to 80%. First we wanted to establish how representative these two new simultaneous MEG/EEG sessions had been, by replicating the EEG-based analysis. To do this, an arbitrary subset of the 60 EEG channels were selected (taking roughly every second channel among the total of 124), the standard time/frequency filter window was applied, and the resulting data was classified using a 5-fold test-training procedure.[4] The first participant's data was typical of the previous cohort, classifying with accuracy of 70% (in this session, accuracy over 61% is significant at $p < 0.05$, using a one-sided binomial test, $n = 264$, $p = 0.54$), while the second participant's data only achieved 52% accuracy (accuracy over 56% significant at $p < 0.05$, $n = 360$, $p = 0.5$).

To get a first impression of the relative informativity of each signal type, the same procedure was performed with subsets of 60 MEG channels: magnetometers alone yielded markedly higher results (78% and 61% for participants 1 and 2 respectively), while planar gradiometers alone gave marginally lower results (67% and 48% respectively).

Next, to examine the effect of increasing the amount of input data, we performed these analyses using all available channels of each type. In one case (participant 1, magnetometers) there was a drop in 5% points, and another (participant 2, magnetometers) an increase of 3% points, but generally this had little effect on results, indicating that in most cases any increase in available information was offset by overfitting.

These results are summarised in the first two columns of in Tables 1 and 2. The tables also show

---

[4]In each test/training partition, the labelled training data alone was used to derive two category specific scalp-maps. These scalp-maps were used to extract signal components and resulting signal power measures for all trials. The data was then partitioned again along the same folds for SVM training and prediction.

Table 1: Classification accuracy, participant 1

Type (available signals)	60 ch.	all ch.	60 cp.
EEG (124)	70%	69%	76%
Magnetometers (102)	78%	73%	78%
Gradiometers (204)	67%	66%	71%
Mag.+Grad. (306)	72%	63%	77%
EEG+Mag. (224)	68%	67%	77%
EEG+Grad. (328)	69%	54%	73%
EEG+Mag.+Grad. (430)	72%	55%	77%

ch: raw channel input; cp: PCA component input

significance: 61% at $p < 0.05$; 65% at $p < 0.001$

Table 2: Classification accuracy, participant 2

Type (available signals)	60 ch.	all ch.	60 cp.
EEG (124)	52%	50%	52%
Magnetometers (102)	61%	64%	68%
Gradiometers (204)	48%	51%	60%
Mag.+Grad. (306)	63%	50%	56%
EEG+Mag. (224)	56%	53%	58%
EEG+Grad. (328)	52%	53%	62%
EEG+Mag.+Grad. (430)	58%	51%	55%

ch: raw channel input; cp: PCA component input

significance: 56% at $p < 0.05$; 59% at $p < 0.001$

the results for all possible combinations of the three signal types, and it is apparent that the effect of overfitting is more pronounced for these larger signal sets. And though the base level of classification accuracy is very different for these two participants, both show a similar pattern with respect to signal type and dimensionality: magnetometers are most informative for these semantic distinctions, and all signal types are vulnerable to overfitting effects.

To combat overfitting, we repeated these analyses with dimensionality reduction. Since PCA is an unsupervised technique, the components were derived and extracted in one step over the whole data set. The first (strongest) 60 components were then taken as input to the same analysis procedure as before (CSP-derived signal power estimates fed to the SVM), to give a global description of whole scalp neural activity, presumably with reduced redundancy and noise. As can be seen in the final columns of Tables 1 and 2, this resulted in optimal classification accuracy in almost all cases, both relative to the full collections of signals, and the 60

channel subsets.

A serious limitation of these results however is the arbitrary selection of signal subsets. While much of the information recorded between signals is likely redundant, it could be that the random inclusion or exclusion of one channel or component could dramatically affect accuracy, if that signal was particularly informative, or particularly subject to spurious noise. So to have a more comprehensive view, we conducted an exhaustive parameter search through possible subsets of each combination of signal type, increasing set size in steps of five, and calculating average classification accuracy with a moving window of nine points. The results are illustrated in Figures 2 (using the raw signals as input) and 3 (using PCA components of each signal set), and show the average prediction performance across both experimental participants.

Several things stand out when considering the difference between the classification performance using raw signals directly, and dimensionality reduced sets. In the PCA case, the classification accuracy levels start higher, rise faster, and peak earlier in almost all cases. In absolute terms optimum performance is little changed for magnetometer and EEG signals alone (peaking just above 70% and 60% respectively), while gradiometers seem to benefit somewhat (by about 3% points). But the PCA lines are also smoother, reflecting more stability in classification, and so more independence from particular parameter settings.

Common to both plots is that magnetometers are the most informative type, followed consecutively by gradiometers and EEG channels. In terms of mutual redundancy, the information encoded in EEG channels seems to largely be a subset of that encoded by gradiometers (gradiometer performance is not improved by the addition of EEG channels). The interaction of magnetometer data and these signal types is more complex – magnetometer performance is *reduced* by the addition of either or both EEG and gradiometer channels.

4   Conclusion

This paper reports only two sessions of simultaneous MEG/EEG recording, and there were some clear differences in the results for each participant, so the conclusions must be considered tentative. Nevertheless they suggest that EEG data are to a large extent redundant with respect to MEG signals. MEG magnetometers in particular can lead to substantially higher classification accuracy, with smaller numbers of channels, than EEG alone. In the case of the second participant, prediction with EEG signals did not approach significance, while MEG signals allowed highly significant ($p \ll 0.001$) performance. We believe that this advantage is due to the lack of attenuation and higher spatial resolution inherent in MEG, allowing it to pick out individual neural sources with more precision.

Regardless of the signal types chosen, the high dimensionality of the data posed challenges. Any arbitrary subset of channels may leave informative aspects of brain-activity undetected and this led to fluctuating results; but including large numbers of channels invariably leads to overfitting, and consequent falls in classification accuracy. In light of this, a reduction in dimensions that kept most of the global signal intact (in this case a principle components analysis) proved very effective in preventing overfitting, giving reliably superior performance with lower numbers of channels.

While MEG signals proved more informative, there was not always a dramatic difference in performance (peak performance in participant 1 was similar for MEG or EEG; for participant 2 there was a ca. 15% point difference). However this study used a time interval and frequency band in the signal that had been optimised for EEG, so it may be that considering a wider range of frequencies, higher in the spectrum, could allow MEG to achieve better results. Also, though steps were taken to avoid it, slight movements by the participants relative to the MEG apparatus will have compromised the reliability of its signals (EEG does not suffer from the same problem as electrodes are placed directly on the scalp). This could be addressed in future studies with continuous head tracking and correction.

Finally, several variations could be tried to improve the overall classification performance of the system. The spatial decomposition used (CSP) is particularly prone to overfitting (Parra et al., 2005), and could be replaced with less aggressive techniques like Linear Discriminant Analysis. Principle component analysis is a rather brittle technique

Figure 2: Classification accuracy taking subsets of raw signals from sensors of different types, 9-point smoothed
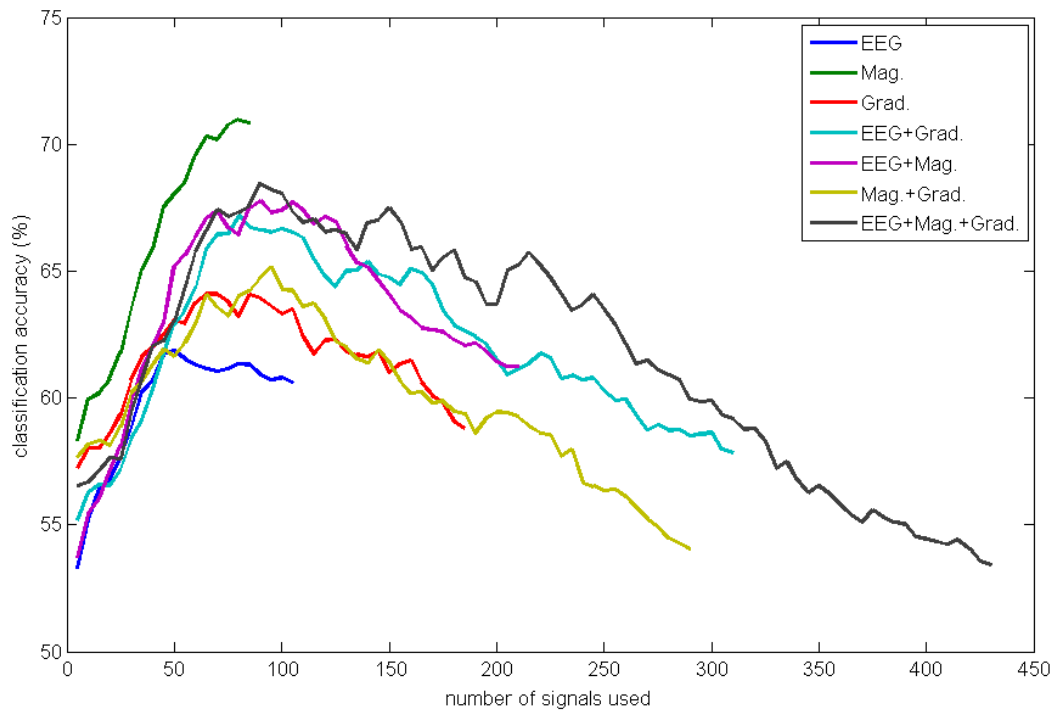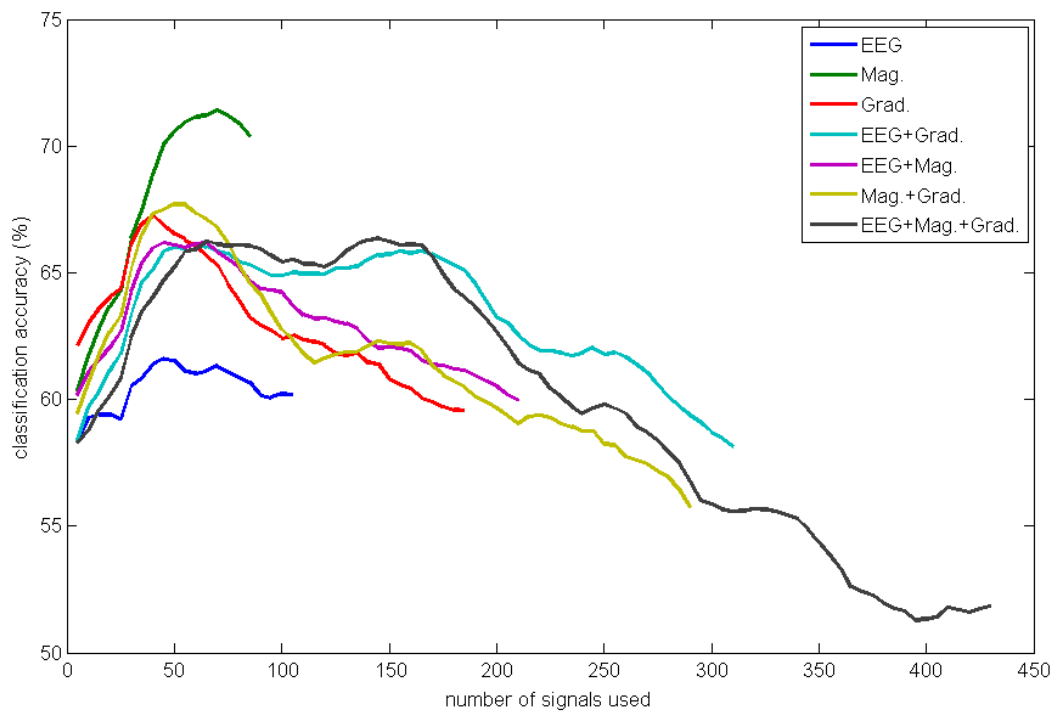

Figure 3: Classification accuracy taking subsets of PCA components derived from raw signals from sensors of different types, 9-point smoothed

which is heavily biased towards the few strongest sources in a system, and so independent component analysis (ICA) may be a more effective choice for dimensionality reduction (Makeig et al., 1996). And data from the various sensor types could be combined in other ways, using an ensemble of classifiers, each based on different subsets of signals, or by taking more than one class-sensitive component per category.

Acknowledgements

References

Boser, B. E.; I. M. Guyon; and V. N. Vapnik (1992): A training algorithm for optimal margin classifiers. In: *5th Annual ACM Workshop on COLT*, ed. D. Haussler. ACM Press, Pittsburgh, pp. 144–152.

Chang, Chih-Chung and Chih-Jen Lin (2001): *LIB-SVM: a library for support vector machines.*

Chao, Linda L.; Jill Weisberg; and Alex Martin (2002): Experience-dependent modulation of category related cortical activity. *Cerebral Cortex*, 12:545–551.

Dalponte, Michele; Francesca Bovolo; and Lorenzo Bruzzone (2007): Automatic selection of frequency and time intervals for classification of EEG signals. *Electronics Letters*, 43:1406–1408.

Hill, N.J.; T.N. Lal; M. Schroder; T. Hinterberger; G. Widman; C.E. Elger; B. Scholkopf; and N. Birbaumer (2006): Classifying event-related desynchronization in EEG, ECoG and MEG signals. *Lecture Notes in Computer Science*, 4174:404.

Just, M.A.; V.L. Cherkassky; S. Aryal; and T.M. Mitchell (2010): A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PLoS ONE*, 5.

Koles, Zoltan J.; Michael S. Lazar; and Steven Z. Zhou (1990): Spatial patterns underlying population differences in the background EEG. *Brain Topography*, 2(4):275–284.

Makeig, Scott; Anthony J. Bell; Tzyy-ping Jung; and Terrence J. Sejnowski (1996): Independent Component Analysis of Electroencephalographic Data. In: *Advances in Neural Information Processing Systems*. MIT Press, vol. 8, pp. 145–151.

Mitchell, Tom M.; Svetlana V. Shinkareva; Andrew Carlson; Kai-Min Chang; Vicente L. Malave; Robert A. Mason; and Marcel Adam Just (2008): Predicting Human Brain Activity Associated with the Meanings of Nouns. *Science*, 320:1191–1195.

Model, Dmitri and Michael Zibulevsky (2006): Learning Subject-Specific Spatial and Temporal Filters for Single-Trial EEG Classification. *NeuroImage*, 32(4):1631–1641.

Murphy, Brian; Marco Baroni; and Massimo Poesio (2009): EEG responds to conceptual stimuli and corpus semantics. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. The Association for Computational Linguistics, pp. 619–627.

Murphy, Brian; Michele Dalponte; Massimo Poesio; and Lorenzo Bruzzone (2008): Distinguishing Concept Categories from Single-Trial Electrophysiological Activity. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Murphy, Brian; Massimo Poesio; Francesca Bovolo; Michele Dalponte; Lorenzo Bruzzone; and Heba Lakany (2010): EEG decoding of semantic category reveals distributed representations for single concepts. *Brain and Language, under review*.

Parra, Lucas C.; Clay D. Spence; Adam D. Gerson; and Paul Sajda (2005): Recipes for the linear analysis of EEG. *NeuroImage*, 28:326–341.

Philiastides, M.G.; R. Ratcliff; and P. Sajda (2006): Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. *Journal of Neuroscience*, 26(35):8965.

Pulvermüller, Friedemann (2005): Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6:576–582.

Ramoser, H.; J. M. Gerking; and Gert Pfurtscheller (2000): Optimal spatial filtering of single

trial EEG during imagined hand movement. *IEEE Transactions on Rehabilitation Engineering*, 8(4):441–446.

Tanji, Kazuyo; Kyoko Suzuki; Arnaud Delorme; and Nobukazu Shamoto, Hiroshiand Nakasato (2005): High-Frequency gamma-Band Activity in the Basal Temporal Cortex during Picture-Naming and Lexical-Decision Tasks. *Journal of Neuroscience*, 25(13):3287–3293.

Vapnik, V. N. (1998): *Statistical Learning Theory*. Wiley.

Waldert, S.; H. Preissl; E. Demandt; C. Braun; N. Birbaumer; A. Aertsen; and C. Mehring (2008): Hand movement direction decoded from MEG and EEG. *Journal of Neuroscience*, 28(4):1000.