Meta-analysis and databasing of neuroimaging studies

Finn Årup Nielsen

Lundbeck Foundation Center for Integrated Molecular Brain Imaging
at
Informatics and Mathematical Modelling
Technical University of Denmark
and
Neurobiology Research Unit,
Copenhagen University Hospital Rigshospitalet

October 13, 2009

When you have published a study you haven't published the study!

Publishing a study means:

Writing a ‘paper’ in a text processing environment, submitting it to a journal and let the journal publish the paper.

Publishing a study means:

Writing a ‘paper’ in a text processing environment, submitting it to a journal and let the journal publish the paper.

What is wrong with that?

Publishing a study means:

Writing a 'paper' in a text processing environment, submitting it to a journal and let the journal publish the paper.

What is wrong with that?

The results is typically a neuroimage volume, but the paper cannot display volume. ☹

Publishing a study means:

Writing a ‘paper’ in a text processing environment, submitting it to a journal and let the journal publish the paper.

What is wrong with that?

The results is typically a neuroimage volume, but the paper cannot display volume. ☹

Even without the volume the data and the meta-data of the study in the paper are not in a form where it is readable for a computer. ☹

Publishing a study means:

Writing a 'paper' in a text processing environment, submitting it to a journal and let the journal publish the paper.

What is wrong with that?

The results is typically a neuroimage volume, but the paper cannot display volume. ☹

Even without the volume the data and the meta-data of the study in the paper are not in a form where it is readable for a computer. ☹

The paper is not published for computers to read its specialized data. ☹

Information increase

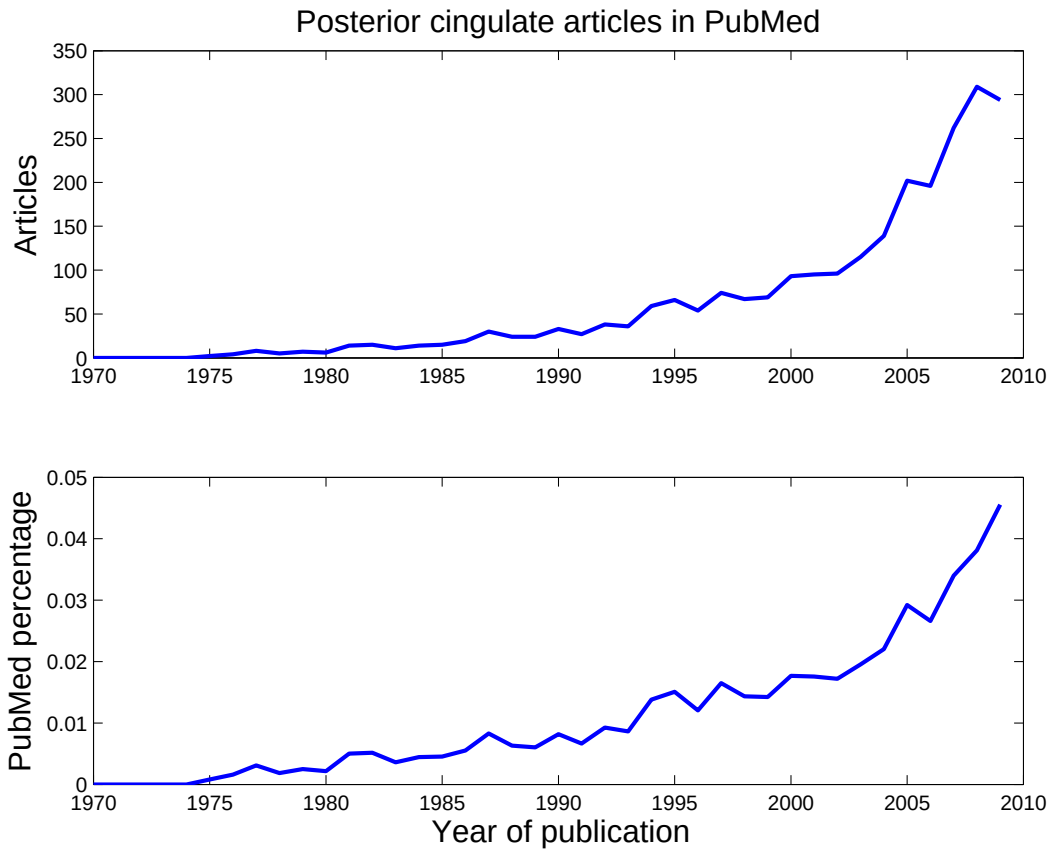


Figure 1: Increase in the number of articles in PubMed which are returned after searching on posterior cingulate and related brain areas.

There are too much data for one person to grasp

The results across experiments are too conflicting

Need for tools that collect data across studies, bring order to data, make search easy and automate analyses to bring out consensus results: **meta-analytic databases**

Classical: PubMed, OMIM, Google Scholar, The Cochrane Collaboration, . . .

When you have published your study you need to publish you data in neuroinformatics databases.

Content

Neuroinformatics databases for MRI & Co. results

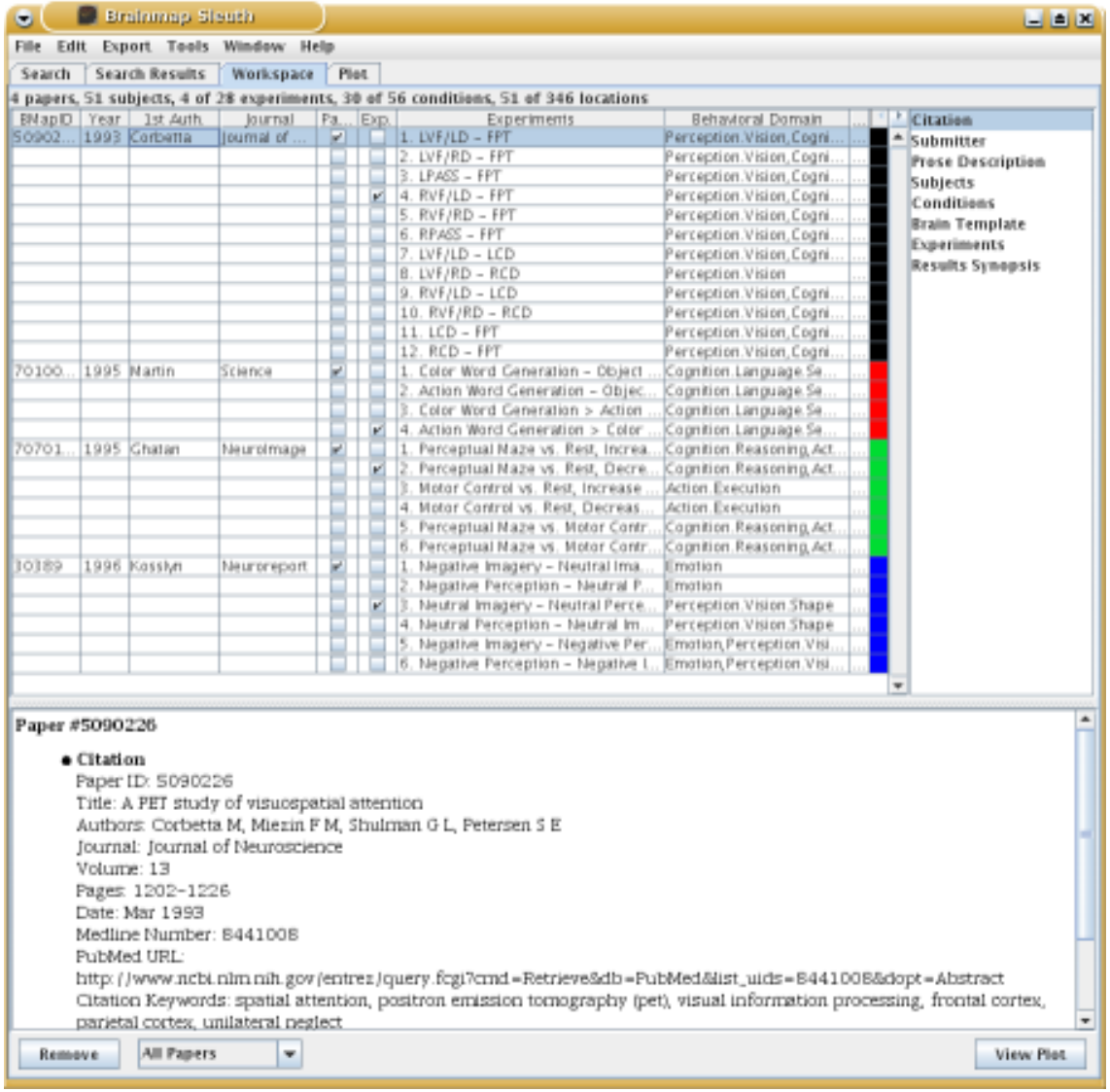
Searching in databases.

Meta-analysis of coordinates: Supervized with one set of coordinates. Supervized with two sets of coordinates. Unsupervized.

Text mining

Combining text mining and coordinate-based meta-analysis.

BrainMap



One of the first and most comprehensive databases (Fox et al., 1994; Fox and Lancaster, 2002)

Presently 69210 locations from 1831 papers (2009 October)

Graphical Internet-based interface in Java, *sleuth*, with search facilities, e.g., on author, 3D coordinate, an others

BrainMap

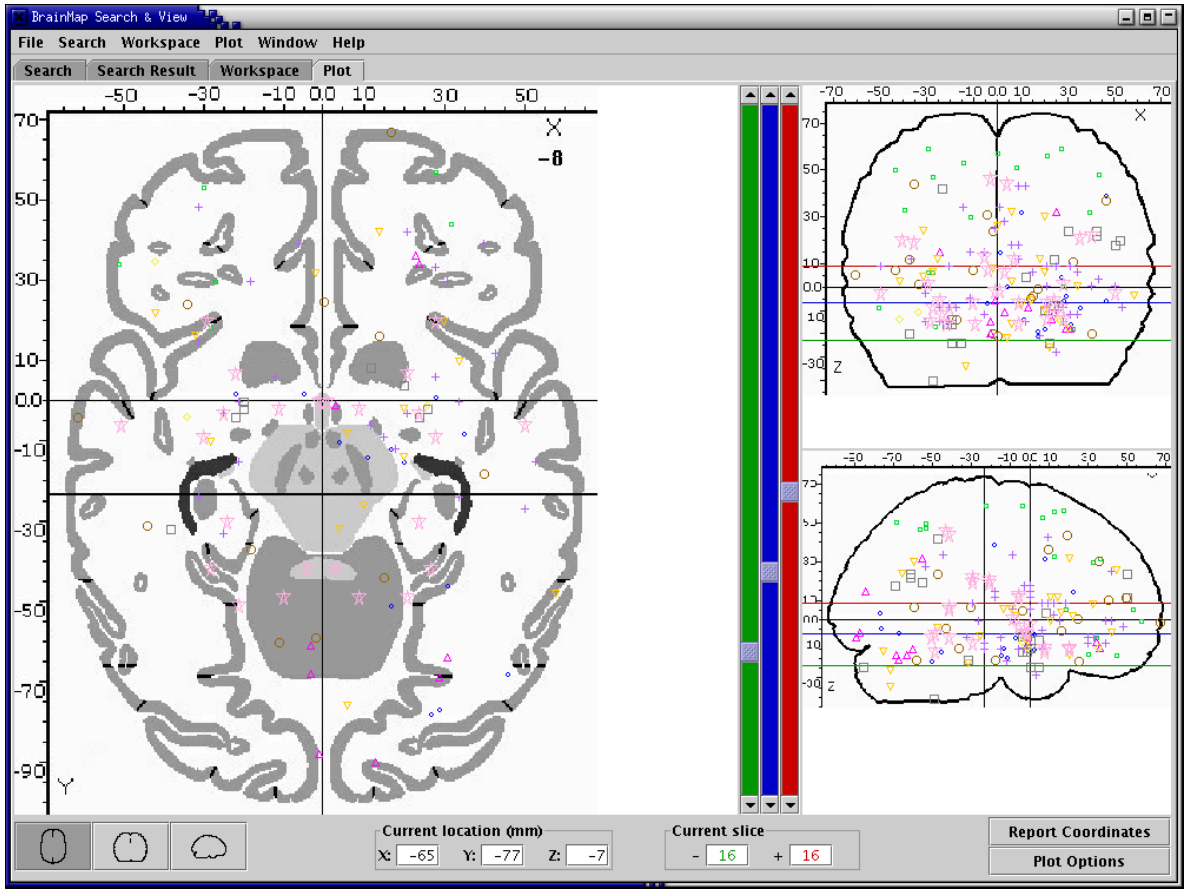


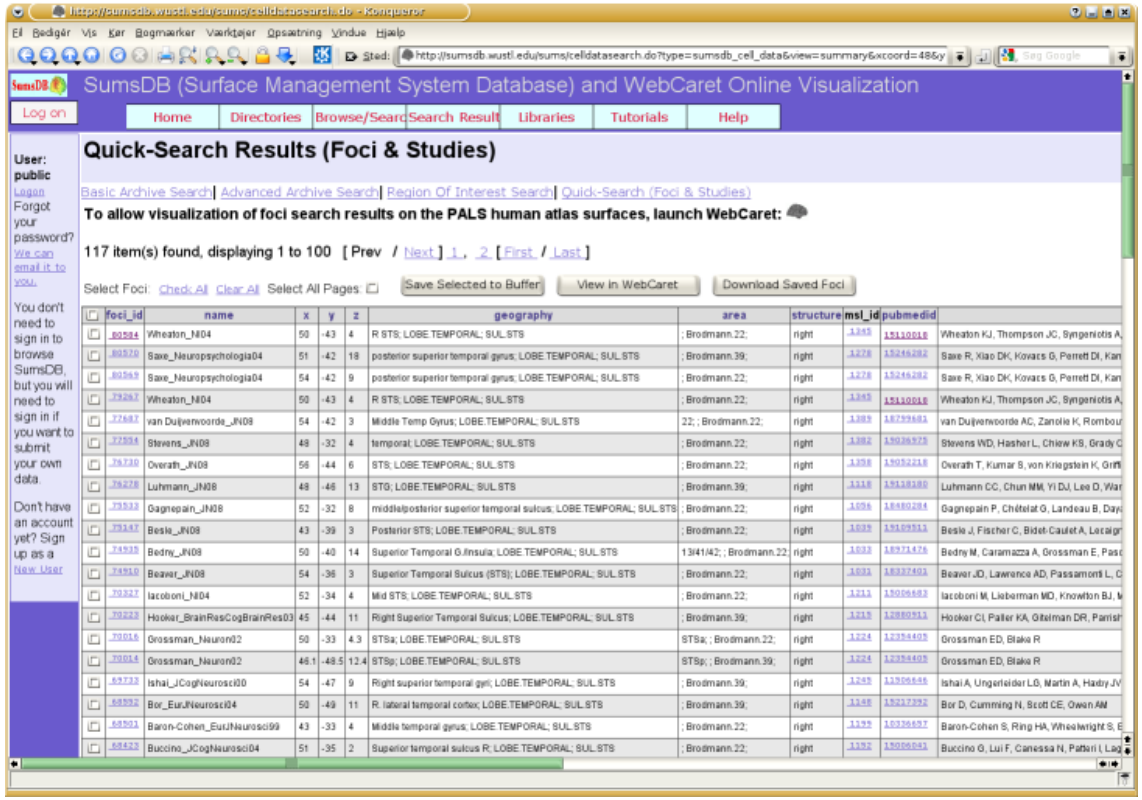
Figure 2: Screen shot of a graphical user interface to the BrainMap database with Talairach coordinates plotted after a search for experiments on olfaction.

The Java program, *sleuth*, is able to show retrieved coordinates in 2D interactive plots.

Possible to enter data with the *Scribe* Java program.

http://brainmap.org

SumsDB



SumsDB (Van Essen, 2009)

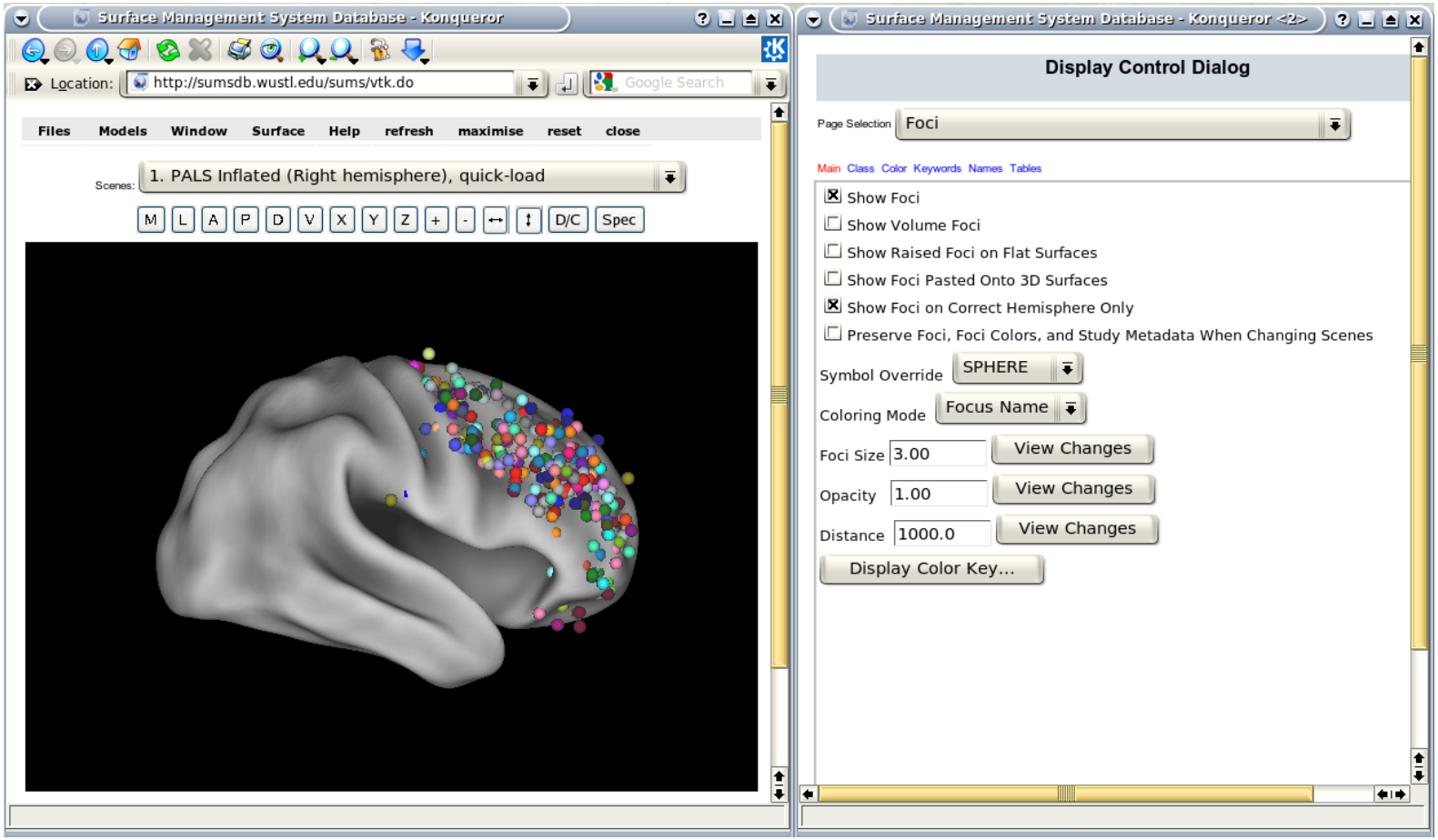
http://sumsdb.wustl.edu/sums/

93919 foci(?)

Less annotated, younger and more(?) coordinates than BrainMap.

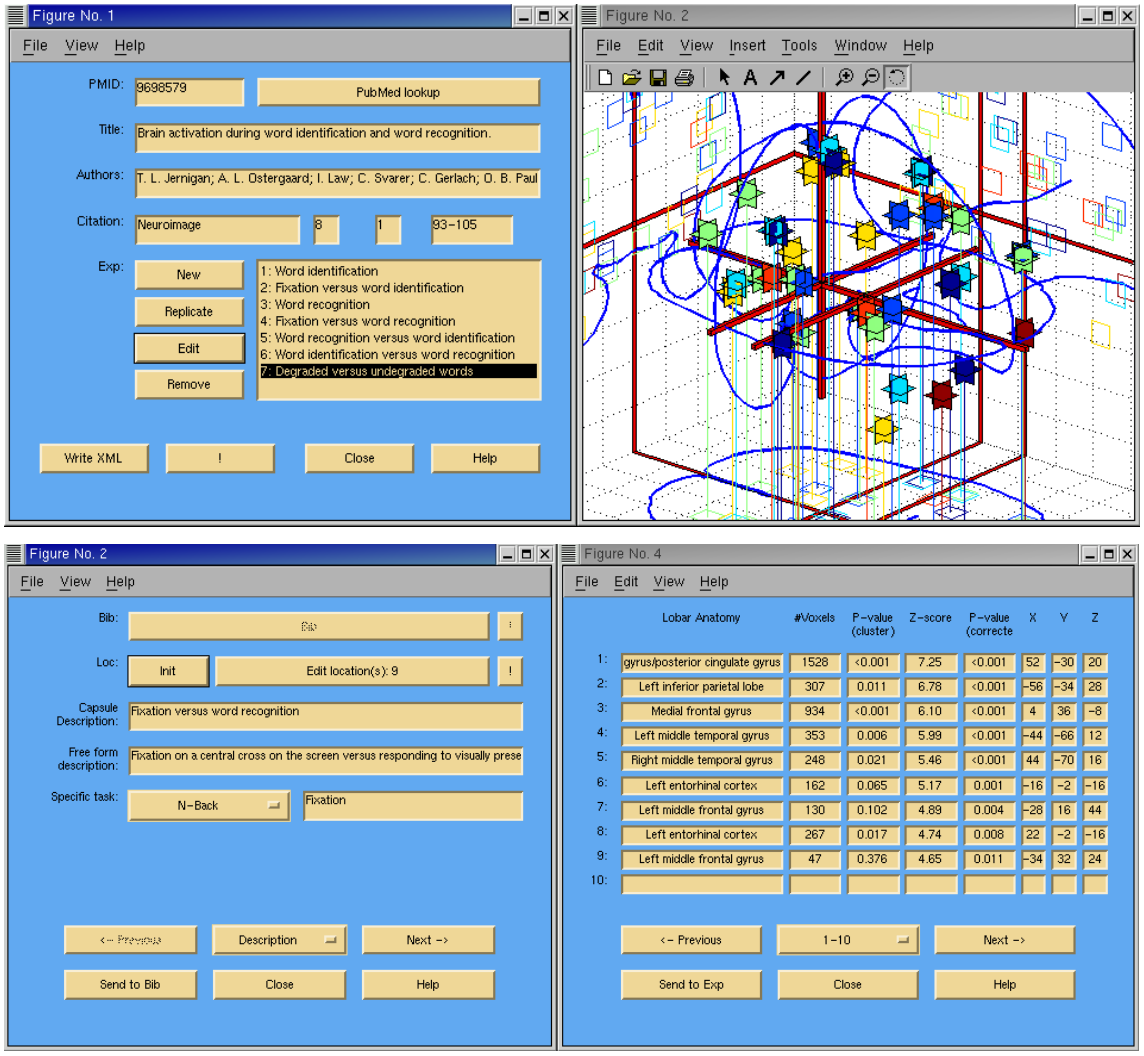
Possible to upload other data, e.g., surfaces.

SumsDB



WebCaret server-side display of returned coordinates from the Surface Management System Database (SumsDB) with a query on 'middle frontal gyrus'

The Brede Database



A database with results from published neuroimaging studies as well as ontologies for, e.g., brain regions and brain function (Nielsen, 2003).

Data stored in XML available on the Web

Data entered in graphical user interface programmed in Matlab: The "Brede Toolbox".

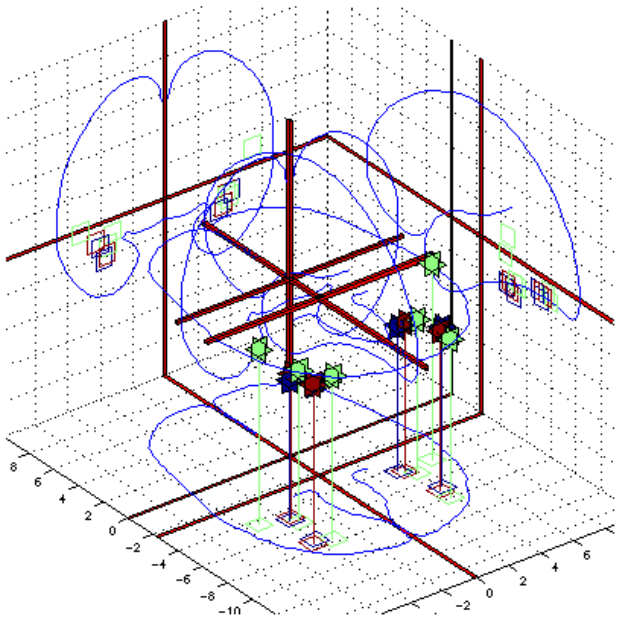
The Brede Database on the Web

WOEXT: 23. **Face recognition.**
Processing of face images.

Parents	Siblings	Children
Visual object recognition		

Experiments:

1. **Face visual object.** *Visual objects: Faces versus building.* WOEXP: 11.
   I Levy; U Hasson; G Avidan; T Hendler; R Malach. *Center–periphery organization of human object areas.*. *Nat Neurosci* **4**(5):533–9, 2001.
   PMID: 11319563. WOBIB: 5.
2. **Photographs of faces versus houses and chairs.** *Conjunction between passive viewing and delayed match–to sample of gray–scale photographs versus scrambled pictures and faces versus houses and chairs, with matching choice indicated by pressing a button with the right of left thumb.* WOEXP: 91.
   A. Ishai; L. G. Ungerleider; A. Martin; J. V. Haxby. *The representation of objects in the human occipital and temporal cortex.*. *J Cogn Neurosci* **12 Suppl 2**:35–51, 2000.
   PMID: 11506646. FMRIDCID: 2–2000–1113D.
   WOBIB: 28.
3. **Front–face.** *Line drawings of front face versus line drawings of tumblers.* WOEXP: 123.
   U. Hasson; T. Hendler; D. Ben Bashat; R. Malach. *Vase or face? A neural correlate of shape–selective grouping processes in the human brain.*. *J Cogn Neurosci* **13**(6):744–53, 2001.
   PMID: 11564319. FMRIDCID: 2–2001–111P8.
   WOBIB: 36.

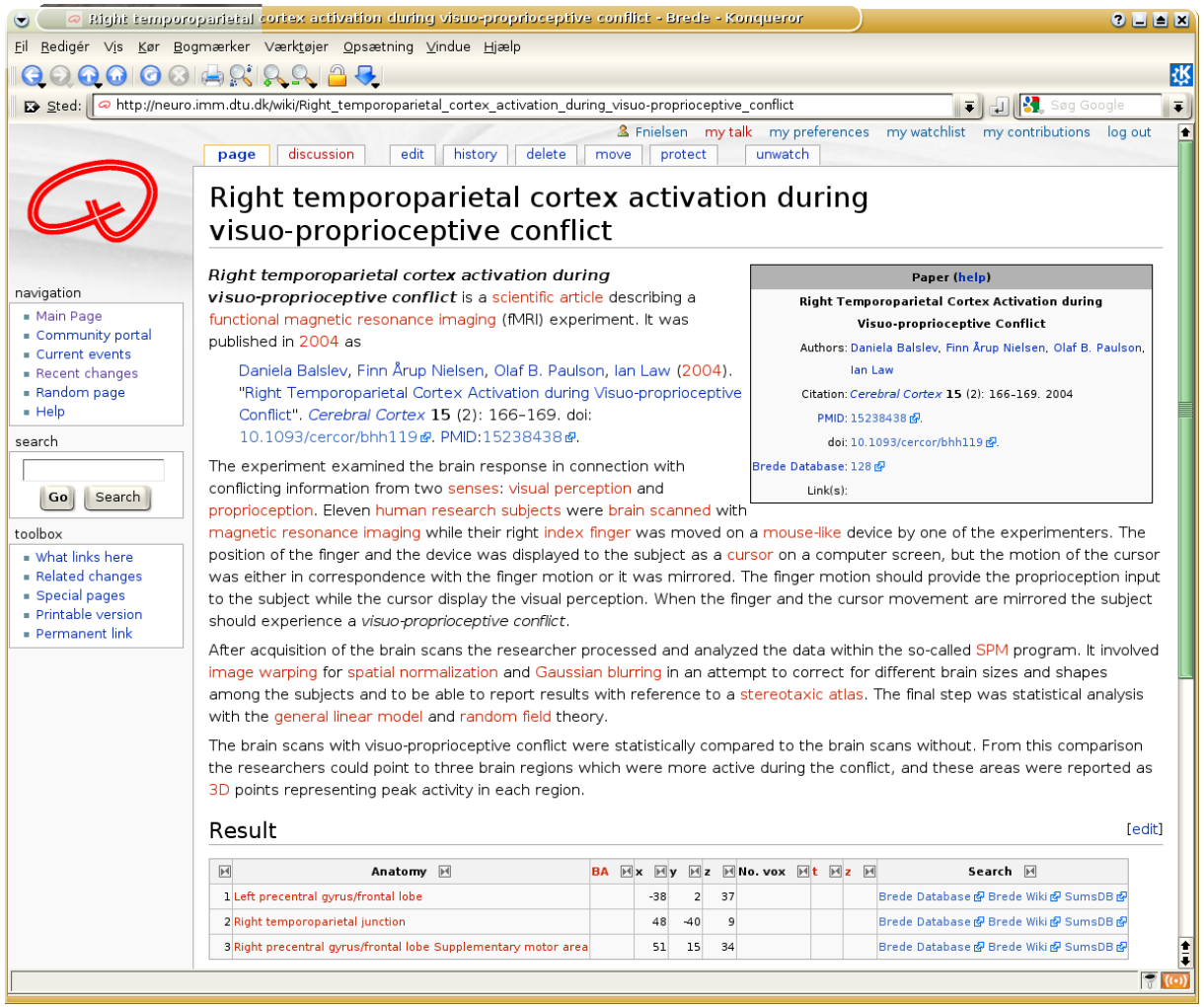


Presentation on the Web via Matlab batch scripts from the Brede Toolbox.

Off-line meta-analysis and generation of indices and visualization in static HTML.

Interactive search on co-ordinates from Web page or within a image analysis program (Wilkowski et al., 2009).

Brede Wiki



Wiki with structured data ☺

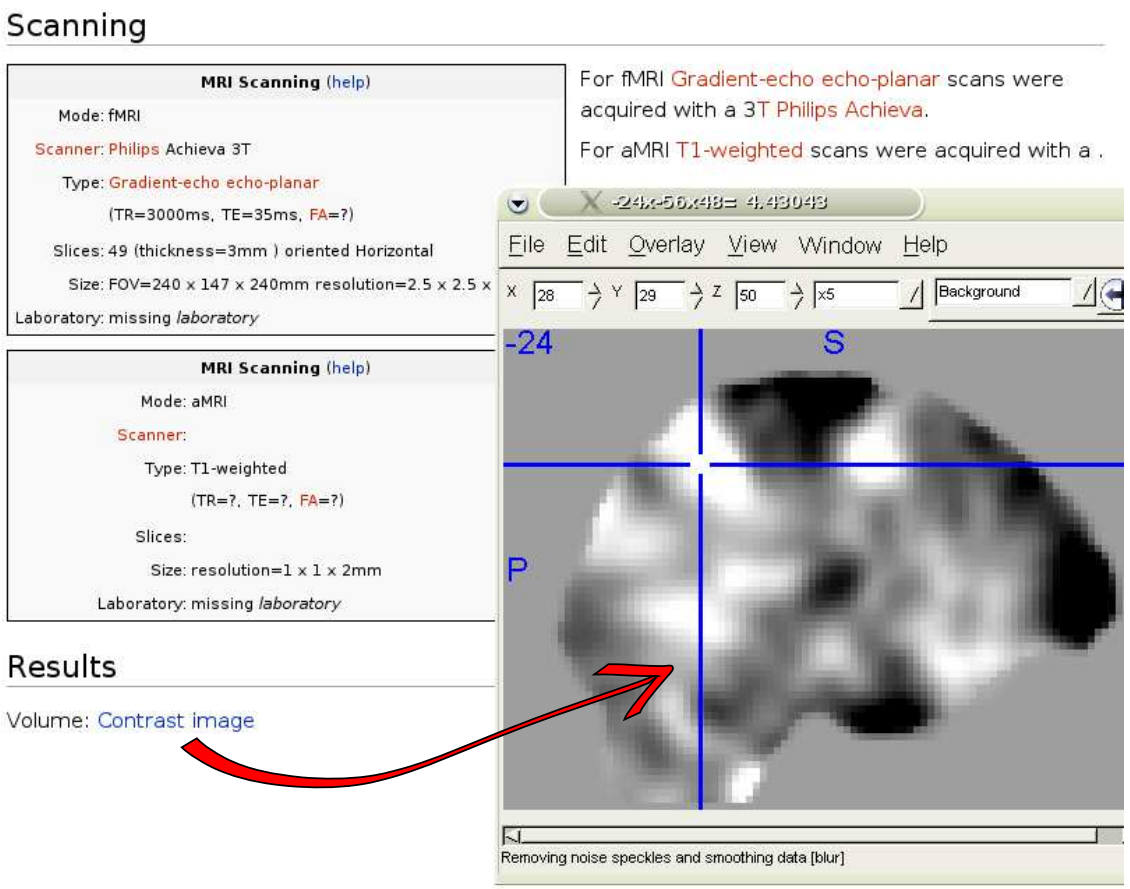
Quick to add new information ☺

Incremental edit possible ☺

Brede Wiki = MediaWiki templates + Extraction + SQL

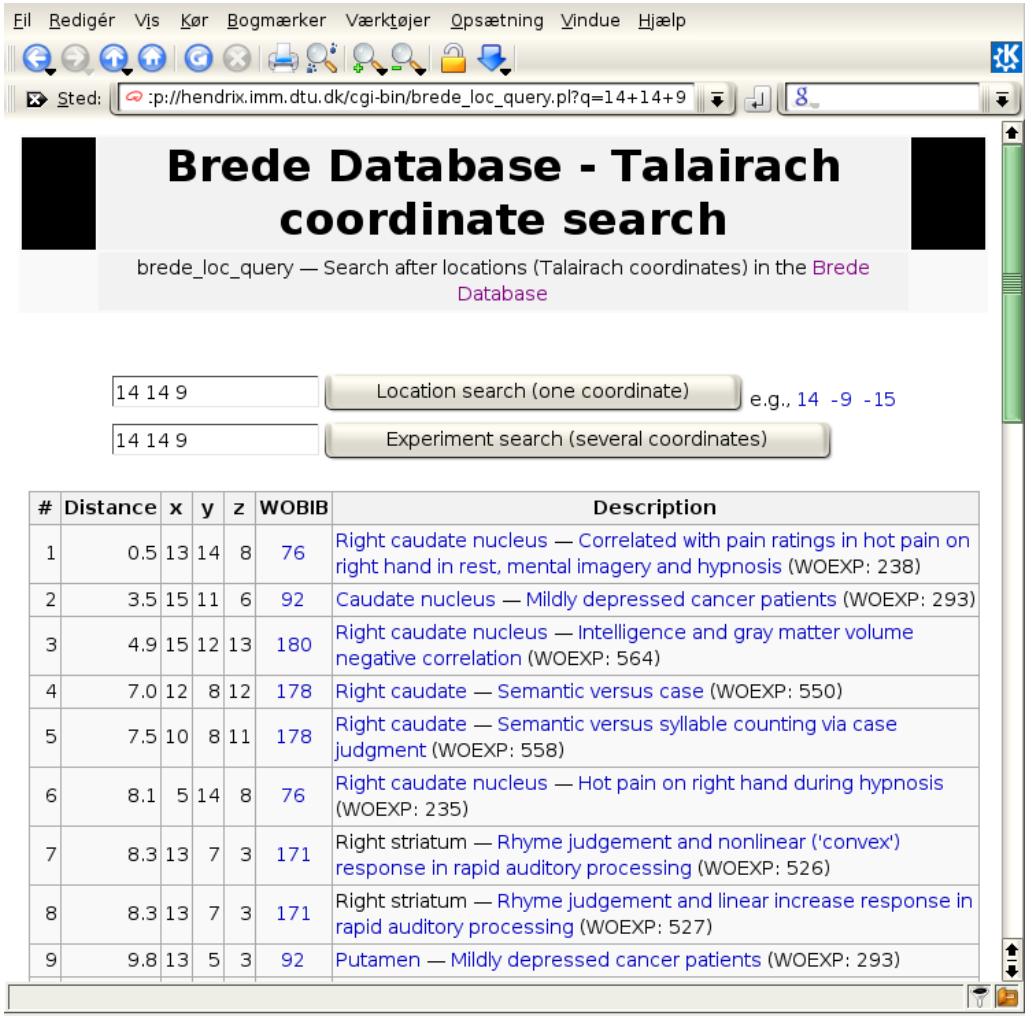
Possible to search outside the wiki

Brede Wiki



Possible to add volume to the database.

Searching on Talairach coordinate



#	Distance	x	y	z	WOBIB	Description
1	0.5	13	14	8	76	Right caudate nucleus — Correlated with pain ratings in hot pain on right hand in rest, mental imagery and hypnosis (WOEXP: 238)
2	3.5	15	11	6	92	Caudate nucleus — Mildly depressed cancer patients (WOEXP: 293)
3	4.9	15	12	13	180	Right caudate nucleus — Intelligence and gray matter volume negative correlation (WOEXP: 564)
4	7.0	12	8	12	178	Right caudate — Semantic versus case (WOEXP: 550)
5	7.5	10	8	11	178	Right caudate — Semantic versus syllable counting via case judgment (WOEXP: 558)
6	8.1	5	14	8	76	Right caudate nucleus — Hot pain on right hand during hypnosis (WOEXP: 235)
7	8.3	13	7	3	171	Right striatum — Rhyme judgement and nonlinear ('convex') response in rapid auditory processing (WOEXP: 526)
8	8.3	13	7	3	171	Right striatum — Rhyme judgement and linear increase response in rapid auditory processing (WOEXP: 527)
9	9.8	13	5	3	92	Putamen — Mildly depressed cancer patients (WOEXP: 293)

Result after search for nearest coordinates to (14, 14, 9) with the Brede Database.

Translation of the data from XML to SQL (Szewczyk, 2008)

Perl + SQLite web-script

Similar searches possible in Antonia Hamilton's AMAT programs, BrainMap, SumsDB and Brede Wiki.

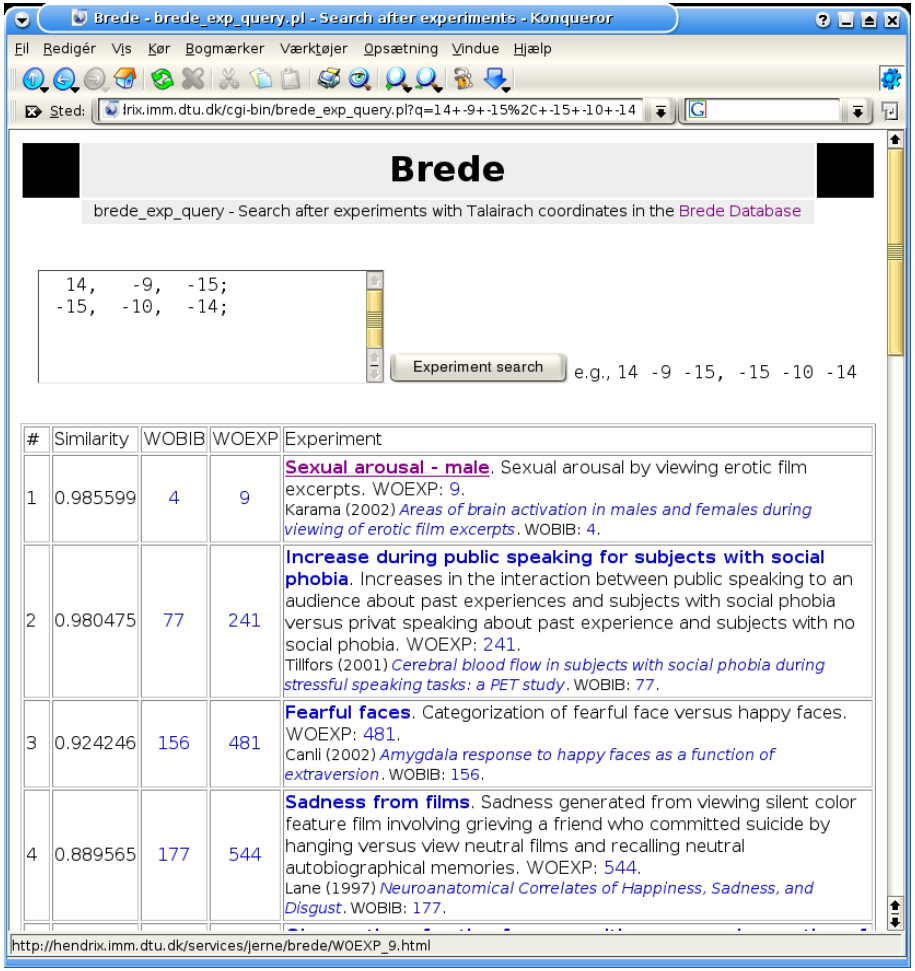
Online experiment search (multiple coordinates)

Online search on two coordinates in left and right amygdala in the experiments recorded in the Brede Database.

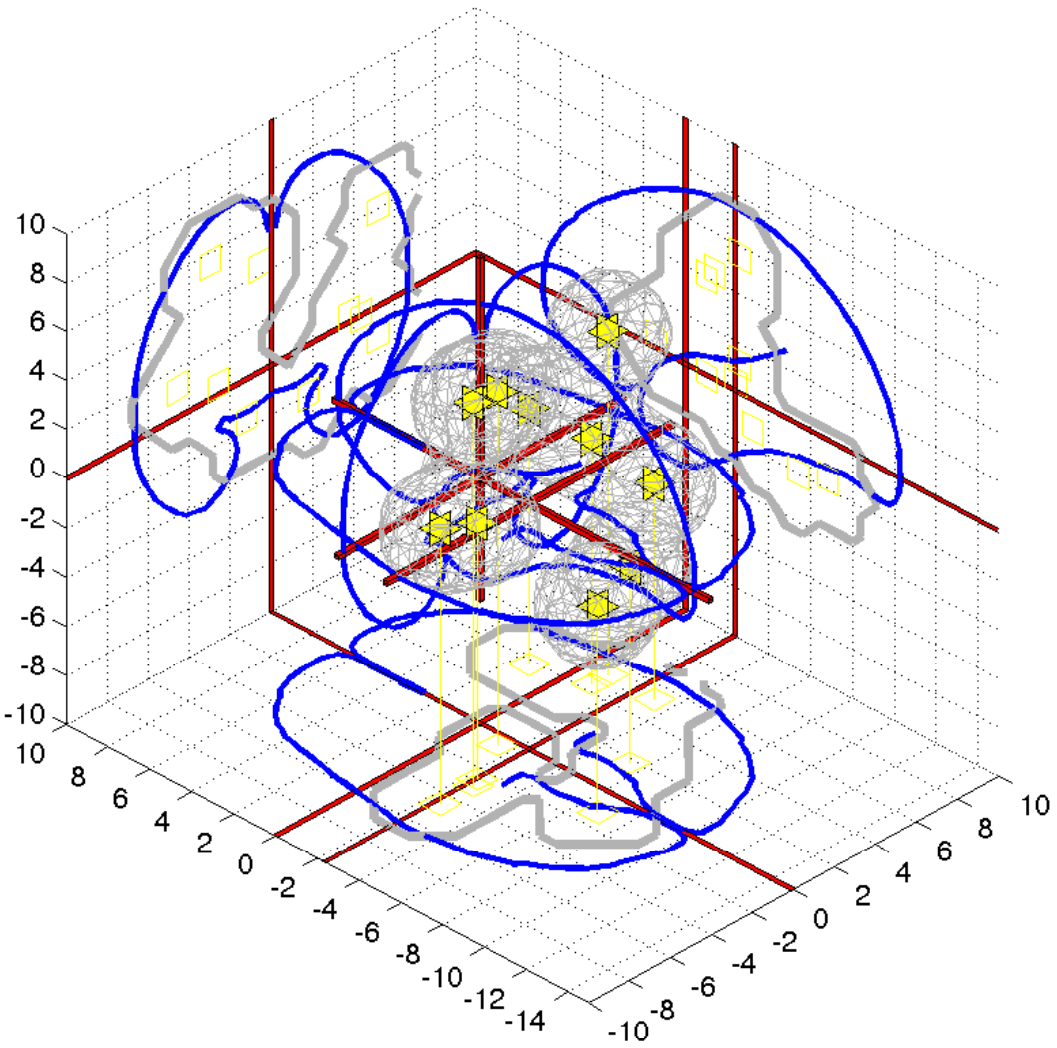
"Related volume" also available from the "original" BrainMap database (Nielsen and Hansen, 2004):

http://neuro.imm.dtu.dk/services/jerne/ninf/

Search available to the Brede Database from SPM plugin (Wilkowski et al., 2009).



Coordinates-to-volume transformation



Coordinates in an article converted to volume-data by filtering each point (kernel density estimation) (Nielsen and Hansen, 2002; Turkeltaub et al., 2002)

One volume for each article or one volume for a set of coordinates in multiple articles.

Yellow coordinates from a study by (Blinkenberg et al., 1996), with grey wireframe indicating the isosurface in the generated volume

Kernel density estimators for coordinates

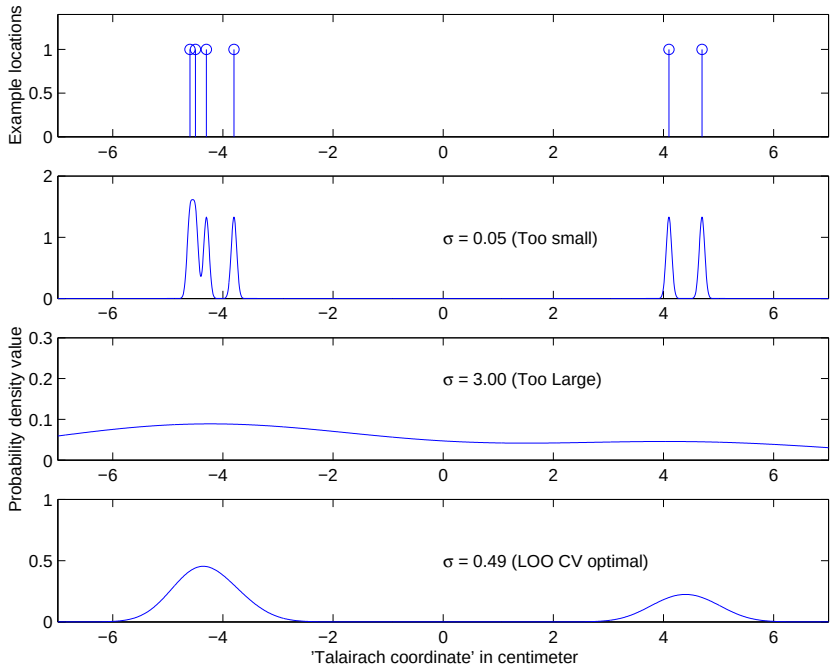


Figure 3: Example in one dimension with six coordinates and their kernel density estimate.

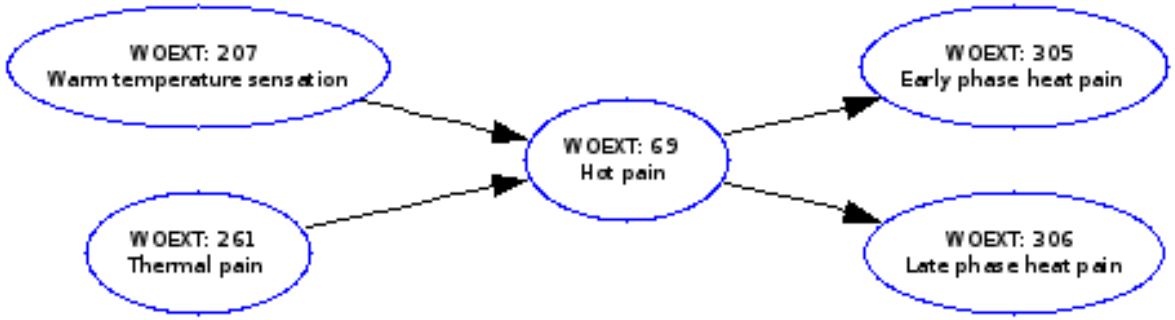
Regard the coordinates as being generated from a distribution $p(\mathbf{x})$, where $\mathbf{x}$ is in 3D Talairach space (Fox et al., 1997).

Kernel methods (N kernels centered on each location: $\boldsymbol{\mu}_n$) with homogeneous Gaussian kernel in 3D Talairach space $\mathbf{x}$

$$\widehat{p}(\mathbf{x}) = \frac{(2\pi\sigma^2)^{-3/2}}{N} \sum_n^N e^{-\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu}_n)^2}$$

σ^2 fixed ($\sigma = 1$cm) or optimized with leave-one-out cross-validation (Nielsen and Hansen, 2002).

Taxonomy for cognitive components, . . .



Brede Database: Memory, episodic memory, episodic memory retrieval, empathy, disgust, 5-HT2A receptor, . . .

Organized in a hierarchy — a directed acyclic graph.

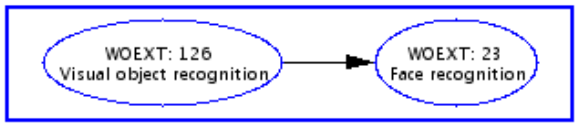
Mass meta-analysis possible with the graph (Nielsen, 2005)

Others: BrainMap taxonomy. Brede Wiki "Topics", MeSH. Under development: Cognitive Atlas (Poldrack), Cognitive Paradigm Ontology (Laird, Turner)

Supervised labeling

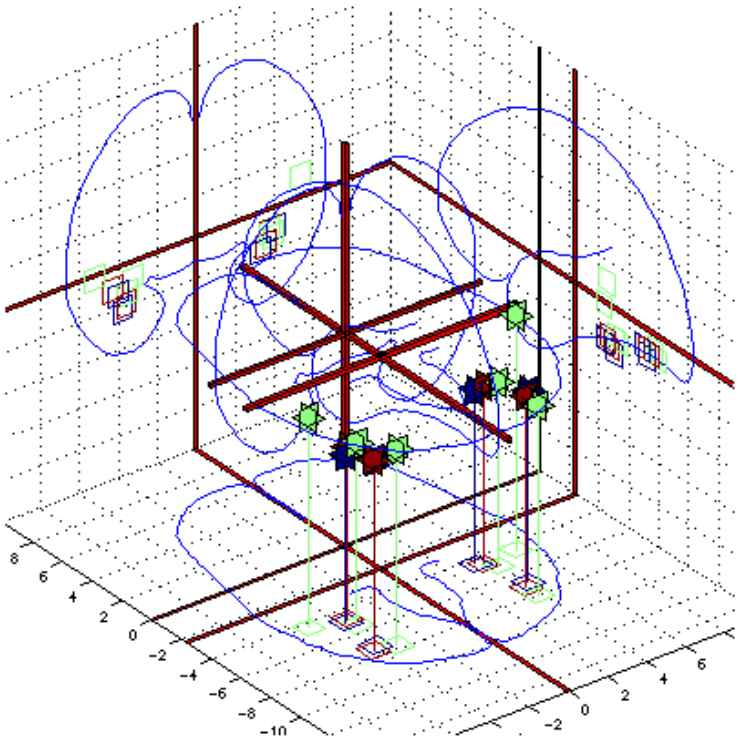
WOEXT: 23. **Face recognition**.
Processing of face images.

Parents	Siblings	Children
Visual object recognition		



Experiments:

1. **Face visual object**. *Visual objects: Faces versus building*. WOEXP: 11.
   I Levy; U Hasson; G Avidan; T Hendler; R Malach. *Center–periphery organization of human object areas.*. *Nat Neurosci* **4**(5):533–9, 2001.
   PMID: 11319563. WOBIB: 5.
2. **Photographs of faces versus houses and chairs**. *Conjunction between passive viewing and delayed match–to sample of gray–scale photographs versus scrambled pictures and faces versus houses and chairs, with matching choice indicated by pressing a button with the right of left thumb*. WOEXP: 91.
   A. Ishai; L. G. Ungerleider; A. Martin; J. V. Haxby. *The representation of objects in the human occipital and temporal cortex.*. *J Cogn Neurosci* **12 Suppl 2**:35–51, 2000.
   PMID: 11506646. FMRIDCID: 2–2000–1113D.
   WOBIB: 28.
3. **Front–face**. *Line drawings of front face versus line drawings of tumblers*. WOEXP: 123.
   U. Hasson; T. Hendler; D. Ben Bashat; R. Malach. *Vase or face? A neural correlate of shape–selective grouping processes in the human brain.*. *J Cogn Neurosci* **13**(6):744–53, 2001.
   PMID: 11564319. FMRIDCID: 2–2001–111P8.
   WOBIB: 36.

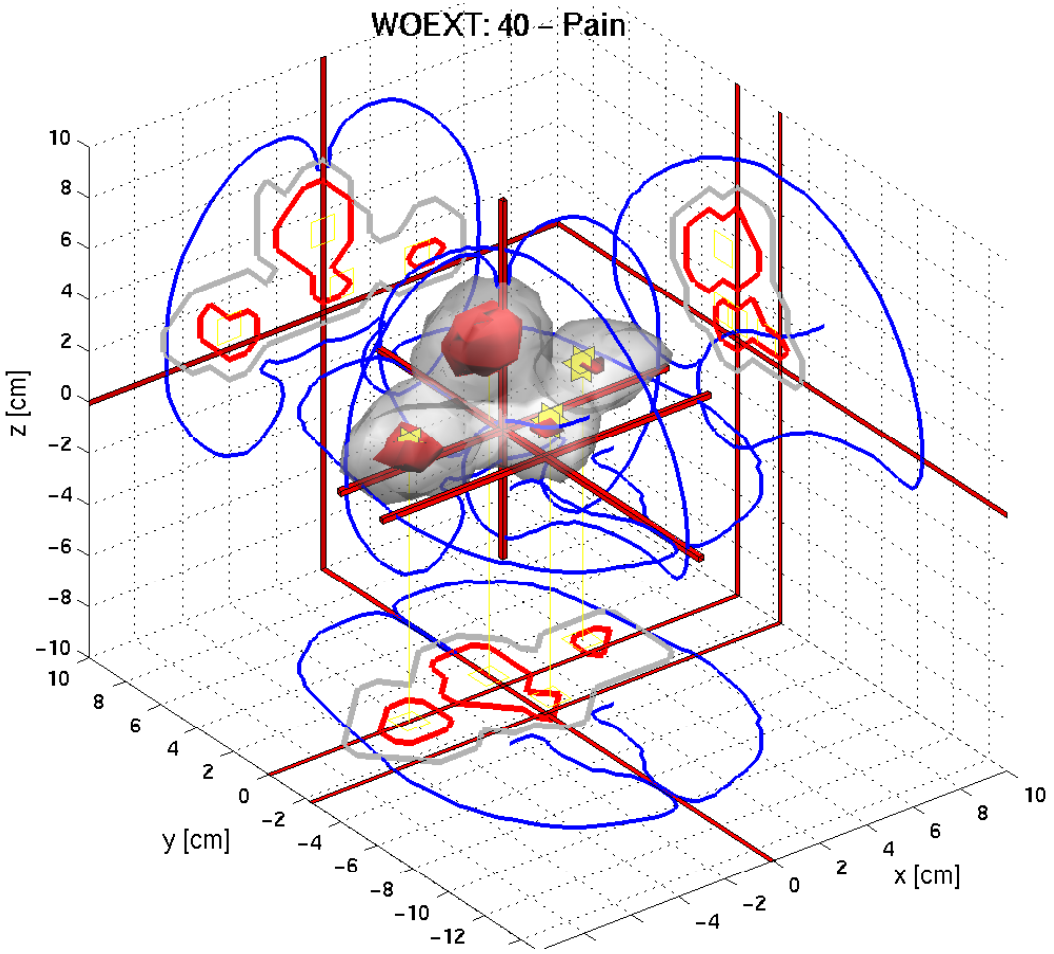


Example with "Face recognition" studies in a "corner cube" visualization.

The "expert" label added during database entry can provide the grouping structure.

Statistical tests can be constructed to measure whether the spatial distribution is "clustered" (Turkeltaub et al., 2002; Nielsen, 2005).

Supervised data mining



Volume for a specific taxonomic component: "Pain" Volume threshold at statistical values determined by resampling statistics (Nielsen, 2005). Red areas are the most significant areas: Anterior cingulate, anterior insula, thalamus. In agreement with "human" reviewer (Ingvar, 1999).

Implementations of supervized datamining in the Brede Toolbox and in GingerALE.

Two sets of coordinates: Compare these!

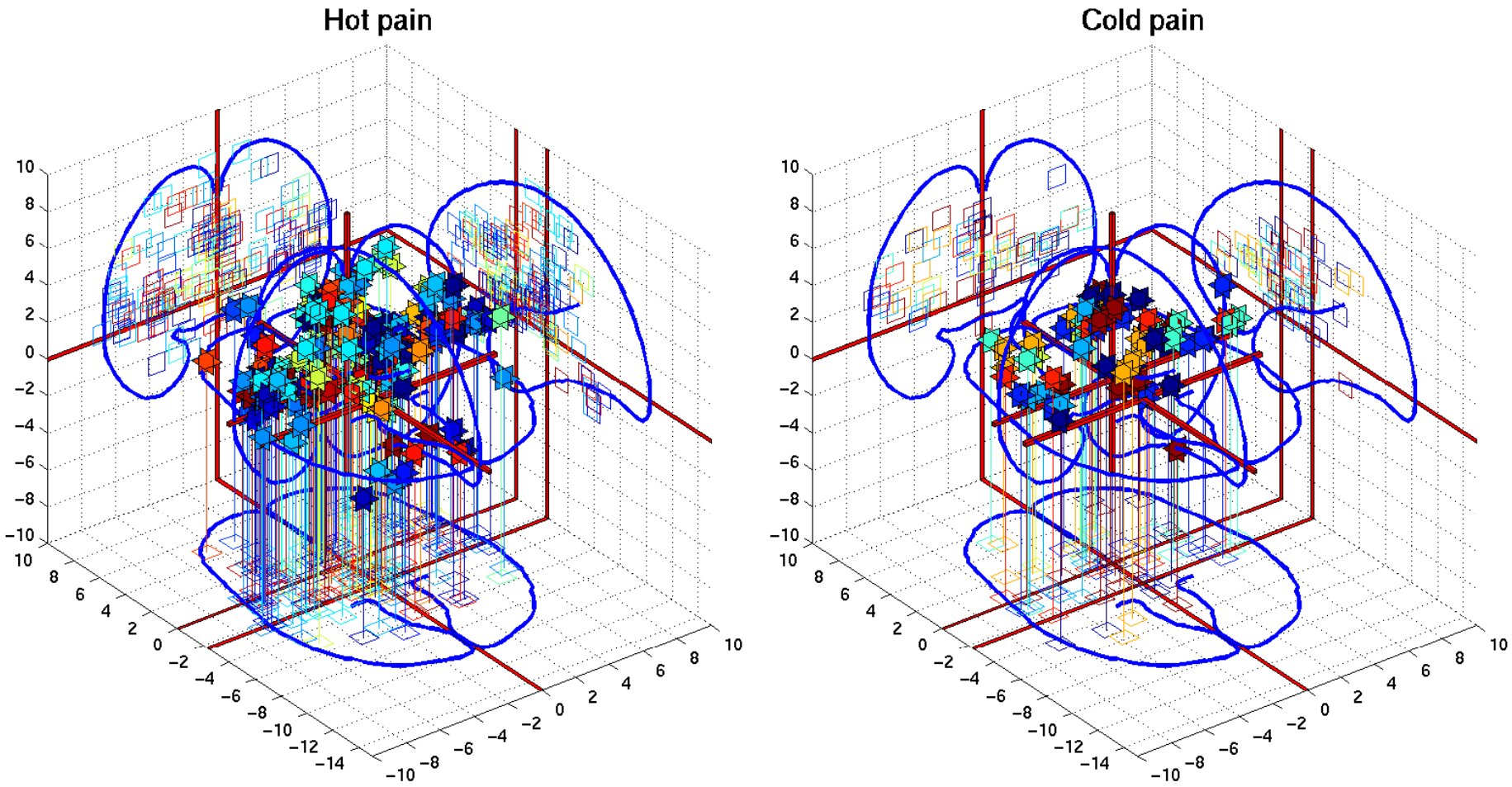


Figure 4: Visualization of the Talairach coordinates from hot pain and cold pain studies

Testing with resampling distribution

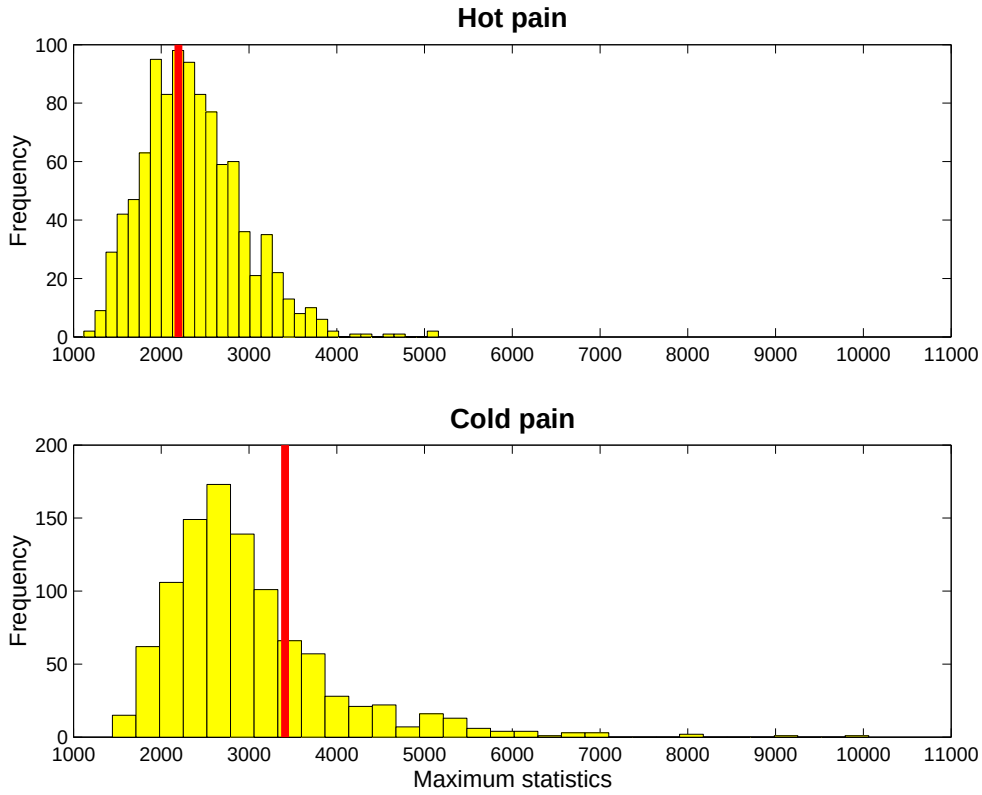


Figure 5: Empirical histograms of the maximum statistics t^* after 1000 permutations. The thick red lines indicate the maxima for the hot and cold pain statistics t_{hot} and t_{cold}.

Two groups are compared by looking at the subtraction volume image

$$\mathbf{t} = \mathbf{v}_1 - \mathbf{v}_2.$$

Histogram of resampled maximum statistics with 1000 resamplings:

$$\begin{aligned} t_{\text{hot}} &= \max\left(\mathbf{v}_{\text{hot}} - \mathbf{v}_{\text{cold}}\right) \\ t_{\text{cold}} &= \max\left(\mathbf{v}_{\text{cold}} - \mathbf{v}_{\text{hot}}\right). \end{aligned}$$

(Nielsen et al., 2004a)

Testing between pain and object vision

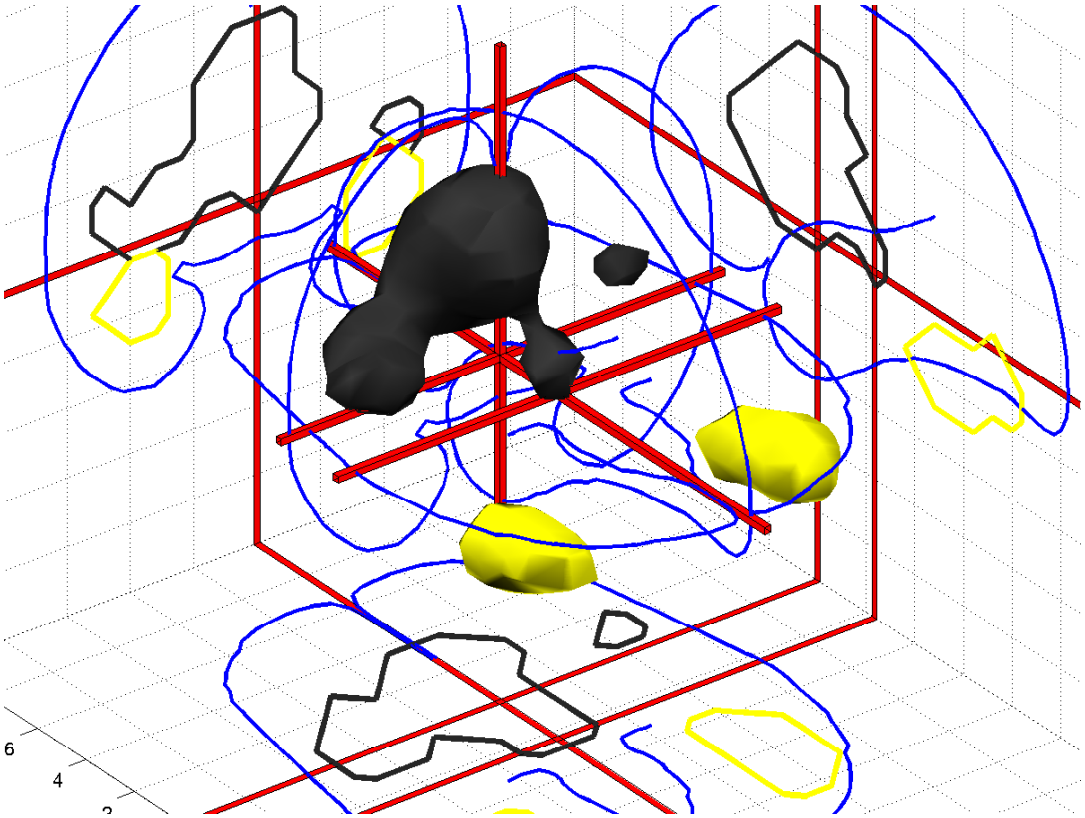


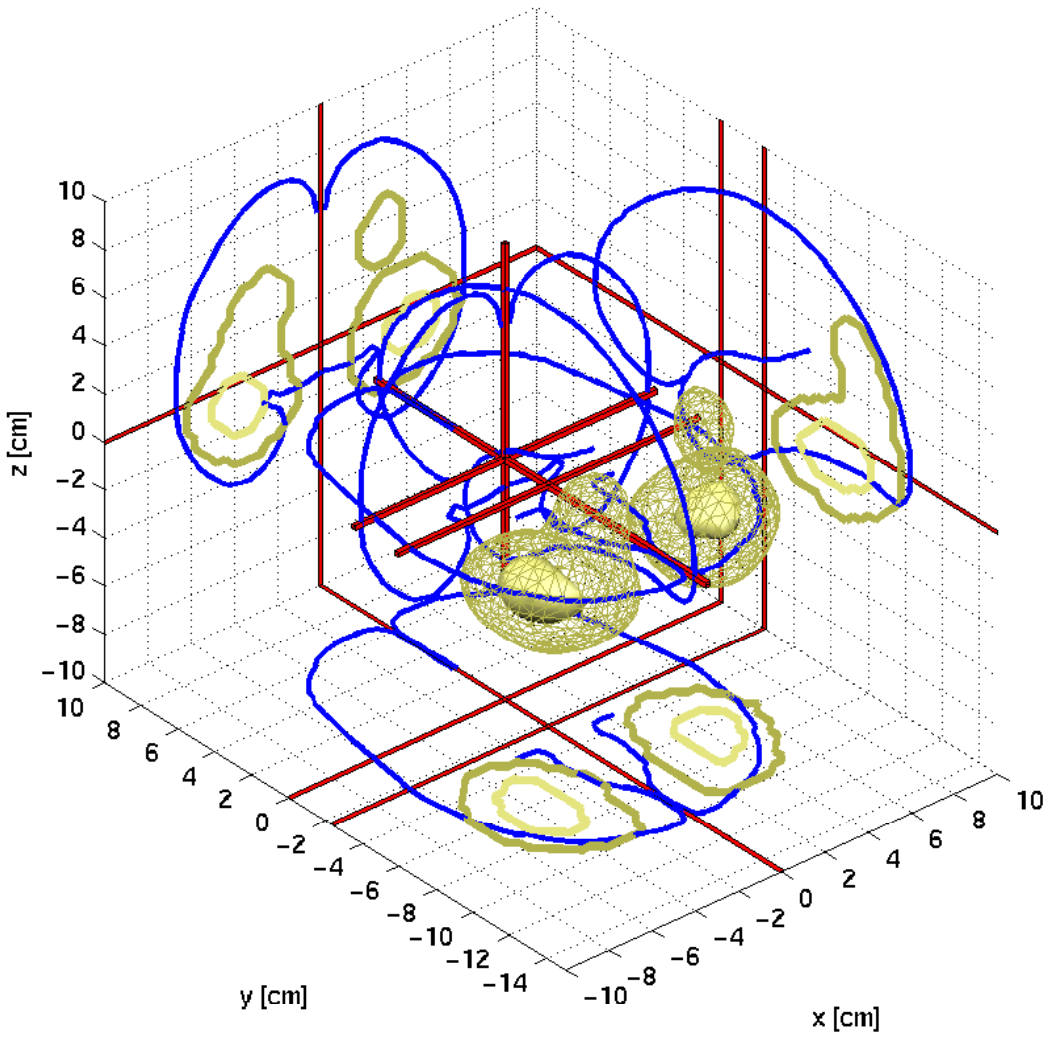
Figure 6: Statistical image. Black is thermal pain and yellow is visual object recognition.

Isosurfaces at thresholds in t_{pain} and t_{object}.

Thresholds are at the usual 0.05-level.

Expected areas appear above threshold. For pain: Anterior cingulate, insula, thalamus. For visual object recognition: fusiform gyrus.

Unsupervised data mining



Construction of a matrix $\mathbf{X}$(experiments $\times$ voxels)

Decomposition of this matrix by multivariate analysis PCA, ICA, NMF, clustering (Nielsen and Hansen, 2004; Nielsen et al., 2004b).

Other technique: Replicator dynamics (Neumann et al., 2005).

Comparison of components with resting-state (Smith et al., 2009)

Issues with meta-analysis

Variable number of subjects between studies.

Varying brain structures examined and reported: Field of view for the scanner, scanner sequence regional sensitivities. ☹

Varying statistical levels used. ☹

Small volume correction is bad for whole brain meta-analysis. ☹

Varying strength between individual coordinates. ☹

In summary: Be careful in interpreting the result of a neuroimaging meta-analysis.

Image-based meta-analysis is coming ☺ Neurogenerator, Brede Wiki upload, SumsDB, (Salimi-Khorshidi et al., 2009b; Salimi-Khorshidi et al., 2009a)

Text representation: a "bag-of-words"

	'memory'	'visual'	'motor'	'time'	'retrieval'	...
Fujii	6	0	1	0	4	...
Maddock	5	0	0	0	0	...
Tsukiura	0	0	4	0	0	...
Belin	0	0	0	0	0	...
Ellerman	0	0	0	5	0	...
⋮	⋮	⋮	⋮	⋮	⋮	⋱

Representation of the abstract of the articles in "bag-of-word". Table counts how often a word occurs

Exclusion of "stop words": common words (the, a, of, ...), words for brain anatomy, and a large number of common words that appear in abstracts. Mostly words for brain function are left. More advanced extraction: Match to ontologies

Grouping of words from articles

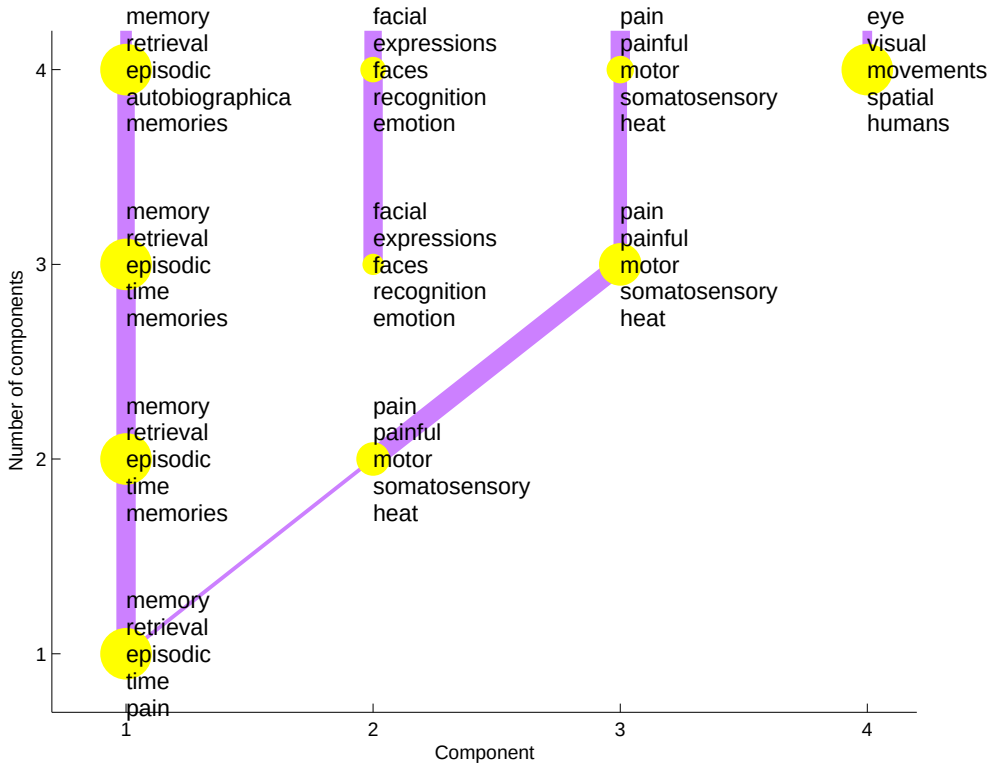


Figure 7: Grouped words.

Multivariate analysis (NMF) of the text in *posterior cingulate* articles to find "themes", which can be represented with weights over words and articles (Nielsen et al., 2005).

Most dominating words: memory, retrieval, episodic

pain, painful, motor, somatosensory

facial, expressions, faces,

eye, visual, movements

Text and volume: Functional atlas

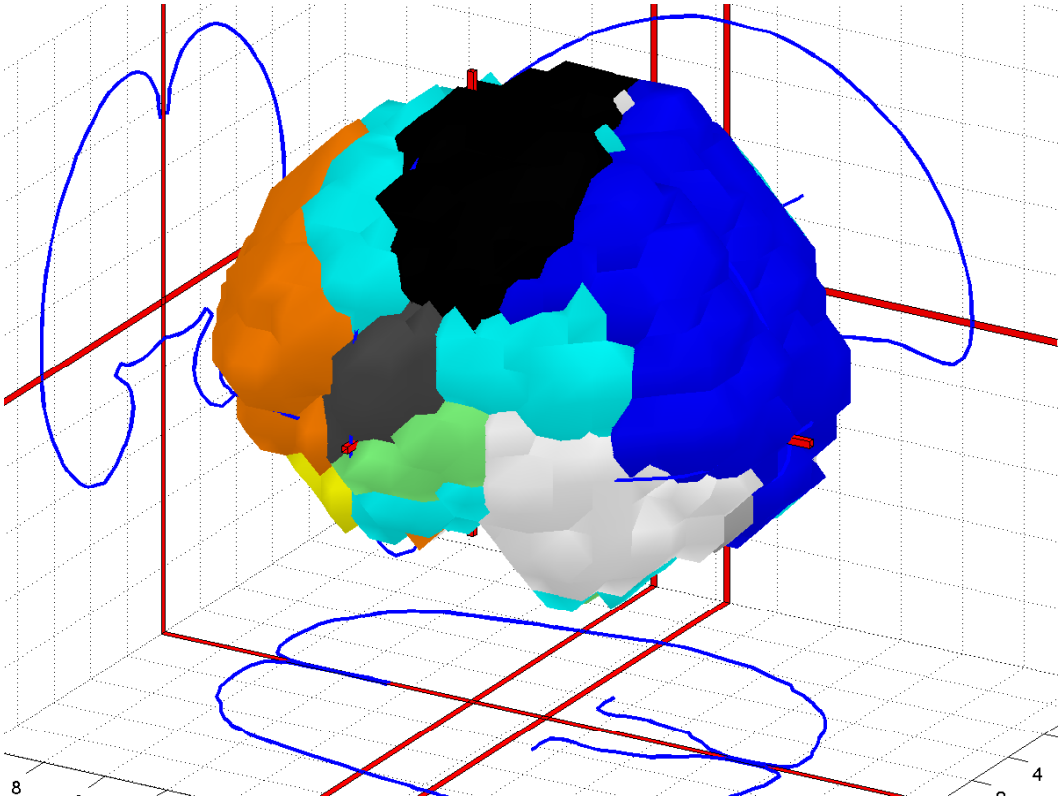


Figure 8: Functional atlas in 3D visualization.

Automatic construction of functional atlas, where words for function become associated with brain areas

Two matrices: Bag-of-words matrix, matrix from voxelization of coordinates. NMF on the product matrix.

Example components: Blue area: visual, eye, time. Black: motor, movements, hand. White: faces, perceptual, face.

Functional atlas — medial view

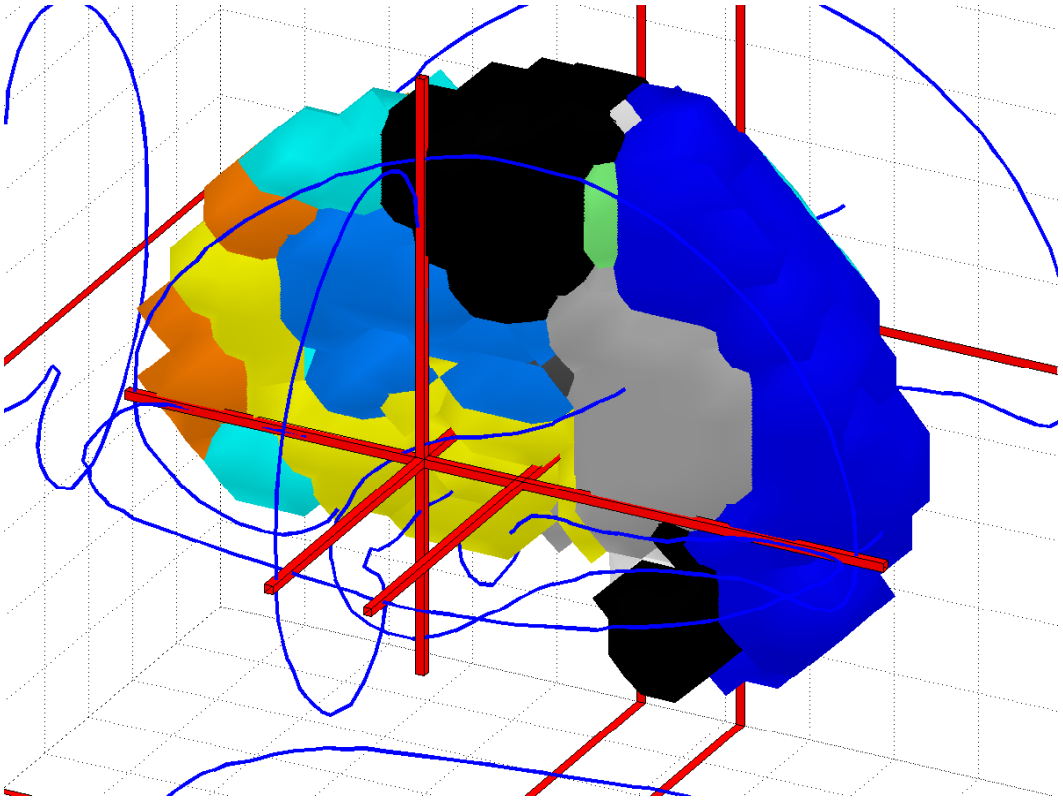


Figure 9: Visualization of the medial area.

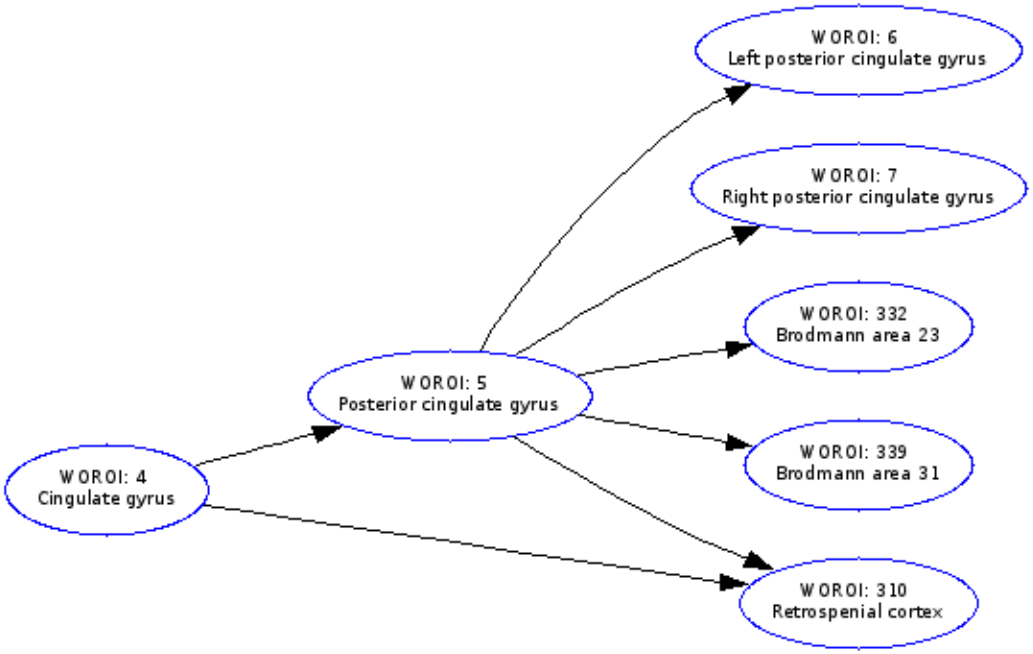
Grey area: retrieval, neutral, words, encoding.

Yellow: emotion, emotions, disgust, sadness, happiness

Light blue: pain, noxious, verbal, unpleasantness, hot

See also *PubBrain* Web service which queries the PubMed database and count occurences of brain regions in abstracts.

Brede brain region taxonomy

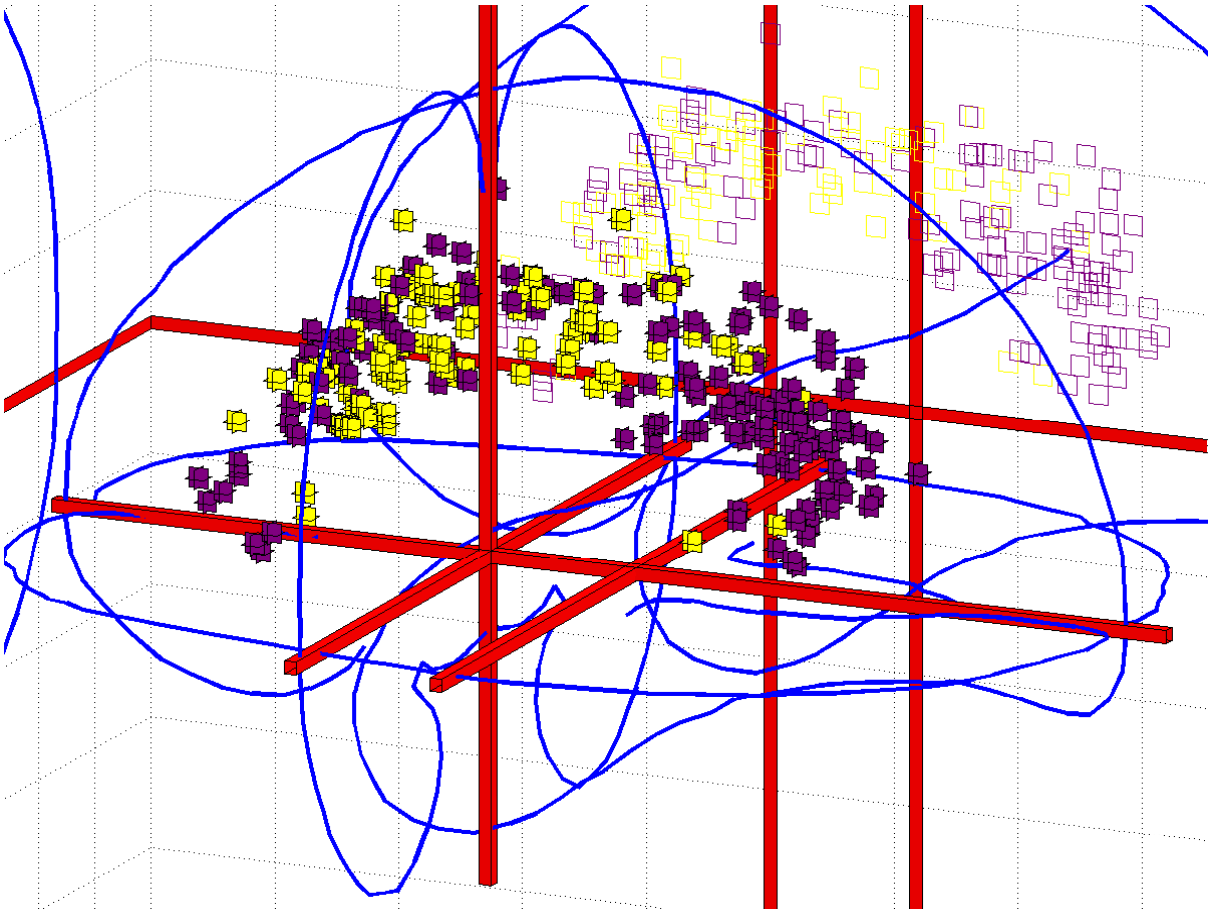


Taxonomy of neuroanatomical areas with items linked in a hierarchy with "Brain" in the top root and smaller areas in the leafs.

Based on another neuroanatomical database "BrainInfo/NeuroNames" (Bowden and Martin, 1995; Bowden and Dubach, 2003) and atlases, e.g. "Mai atlas" (Mai et al., 1997).

Searching for all "cingulate" coordinates

Combining text analysis and coordinates

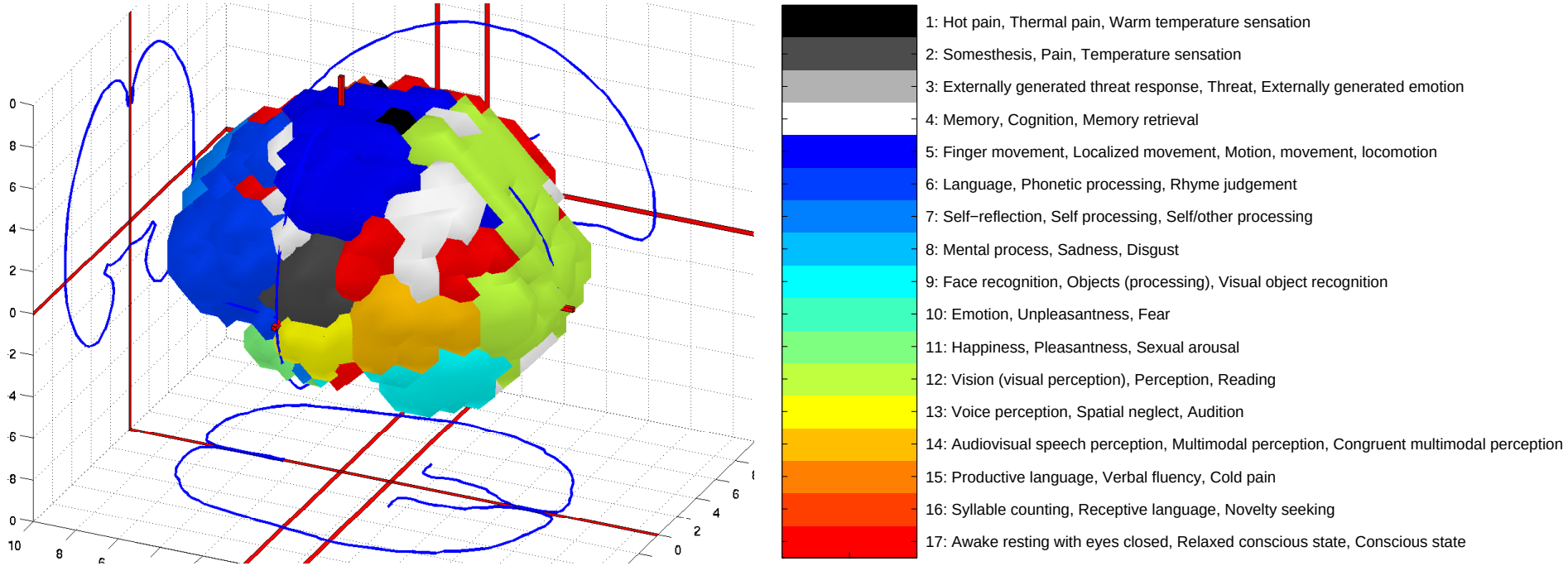


Is there a different between how brain functions distribute in the cingulate gyrus?

Possible to find the corresponding articles for the coordinates — and text mine these articles for clustering and label the coordinate according to cluster.

Sagittal plot of memory (magenta) and pain (yellow).

Combining ontologies and coordinates



Conversion of the Brede Database function taxonomy to a matrix and using that together with matrix from voxelization of the coordinates in the experiments and non-negative matrix factorization.

More information

Articles about neuroinformatics (Nielsen et al., 2006; Nielsen, 2009)

Brede Database Brede Wiki Brede Toolbox

Bibliography on Neuroinformatics:

http://www.imm.dtu.dk/~fn/bib/Nielsen2001Bib/

You should submit you data to a neuroinformatics database to get published.

The End

References

Blinkenberg, M., Bonde, C., Holm, S., Svarer, C., Andersen, J., Paulson, O. B., and Law, I. (1996). Rate dependence of regional cerebral activation during performance of a repetitive motor task: a PET study. *Journal of Cerebral Blood Flow and Metabolism*, 16(5):794–803. PMID: 878424. WOBIB: 166.

Bowden, D. M. and Dubach, M. F. (2003). NeuroNames 2002. *Neuroinformatics*, 1(1):43–59. ISSN 1539-2791.

Bowden, D. M. and Martin, R. F. (1995). NeuroNames brain hierarchy. *NeuroImage*, 2(1):63–84. PMID: 9410576. ISSN 1053-8119.

Fox, P. T. and Lancaster, J. L. (2002). Mapping context and content: the BrainMap model. *Nature Reviews Neuroscience*, 3(4):319–321. http://www.brainmapdbj.org/Fox01context.pdf. Describes the philosophy behind the (new) BrainMap functional brain imaging database with "BrainMap Experiment Coding Scheme" and tables of activation foci. Furthermore discusses financial issues and quality control of data.

Fox, P. T., Lancaster, J. L., Parsons, L. M., Xiong, J.-H., and Zamarripa, F. (1997). Functional volumes modeling: Theory and preliminary assessment. *Human Brain Mapping*, 5(4):306–311. http://www3.interscience.wiley.com/cgi-bin/abstract/56435/START.

Fox, P. T., Mikiten, S., Davis, G., and Lancaster, J. L. (1994). BrainMap: A database of human function brain mapping. In Thatcher, R. W., Hallett, M., Zeffiro, T., John, E. R., and Huerta, M., editors, *Functional Neuroimaging: Technical Foundations*, chapter 9, pages 95–105. Academic Press, San Diego, California. ISBN 0126858454.

Ingvar, M. (1999). Pain and functional imaging. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 354(1387):1347–1358. PMID: 10466155.

Mai, J. K., Assheuer, J., and Paxinos, G. (1997). *Atlas of the Human Brain*. Academic Press, San Diego, California. ISBN 0124653618.

Neumann, J., Lohmann, G., Derrfuss, J., and von Cramon, D. Y. (2005). Meta-analysis of functional imaging data using replicator dynamics. *Human Brain Mapping*, 25(1):165–173. http://www3.interscience.wiley.com/cgi-bin/abstract/110474181/. ISSN 1065-9471.

Nielsen, F. Å. (2003). The Brede database: a small database for functional neuroimaging. *NeuroImage*, 19(2). http://208.164.121.55/hbm2003/abstract/abstract906.htm. Presented at the 9th International Conference on Functional Mapping of the Human Brain, June 19–22, 2003, New York, NY. Available on CD-Rom.

Nielsen, F. Å. (2005). Mass meta-analysis in Talairach space. In Saul, L. K., Weiss, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems 17*, pages 985–992, Cambridge, MA. MIT Press. http://books.nips.cc/papers/files/nips17/NIPS2004_0511.pdf.

Nielsen, F. Å. (2009). Visualizing data mining results with the Brede tools. *Frontiers in Neuroinformatics*, 3:26. DOI: 10.3389/neuro.11.026.2009.

Nielsen, F. Å., Balslev, D., and Hansen, L. K. (2005). Mining the posterior cingulate: Segregation between memory and pain component. *NeuroImage*, 27(3):520–532. DOI: 10.1016/j.neuroimage.2005.04.034. Text mining of PubMed abstracts for detection of topics in neuroimaging studies mentioning posterior cingulate. Subsequent analysis of the spatial distribution of the Talairach coordinates in the clustered papers.

Nielsen, F. Å., Chen, A. C. N., and Hansen, L. K. (2004a). Testing for difference between two groups of functional neuroimaging experiments. In Olsen, S. I., editor, *Proceedings fra den 13. Danske Konference i Mønstergenkendelse og Billedanalyse*, number 2004/10 in DIKU Technical Reports, pages 121–129, Copenhagen, Denmark. Dansk Selskab for Automatisk Genkendelse af Mønstre, Datalogisk Institut, University of Copenhagen. http://www.diku.dk/dsagm04/proceedings.dsagm04.pdf. ISSN 0107-8283.

Nielsen, F. Å., Christensen, M. S., Madsen, K. H., Lund, T. E., and Hansen, L. K. (2006). fMRI neuroinformatics. *IEEE Engineering in Medicine and Biology Magazine*, 25(2):112–119. PMID: 16568943. http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=3516. An overview of some of the tools for and issues in fMRI neuroinformatics with description of, e.g., the SPM, AFNI and FSL programs and the BrainMap, fMRIDC and Brede databases.

Nielsen, F. Å. and Hansen, L. K. (2002). Modeling of activation data in the BrainMap^{TM} database: Detection of outliers. *Human Brain Mapping*, 15(3):146–156. DOI: 10.1002/hbm.10012. http://www3.interscience.wiley.com/cgi-bin/abstract/89013001/. CiteSeer: http://citeseer.ist.psu.edu/nielsen02modeling.html.

Nielsen, F. Å. and Hansen, L. K. (2004). Finding related functional neuroimaging volumes. *Artificial Intelligence in Medicine*, 30(2):141–151. PMID: 14992762. http://www.imm.dtu.dk/~fn/Nielsen2002Finding/.

Nielsen, F. Å., Hansen, L. K., and Balslev, D. (2004b). Mining for associations between text and brain activation in a functional neuroimaging database. *Neuroinformatics*, 2(4):369–380. http://www2.imm.dtu.dk/~fn/ps/Nielsen2004Mining_submitted.pdf.

Salimi-Khorshidi, G., Smith, S. M., Keltner, J. R., Wager, T. D., and Nichols, T. E. (2009a). Meta-analysis of neuroimaging data: A comparison of image-based and coordinate-based pooling of studies. *NeuroImage*, 45:810–823. DOI: 10.1016/j.neuroimage.2008.12.039.

Salimi-Khorshidi, G., Smith, S. M., and Nichols, T. E. (2009b). Bias and heterogeneity in neuroimaging meta-analysis. 15th Annual Meeting of the Organization for Human Brain Mapping Abstracts Online. 406 SA-PM. http://www.meetingassistant3.com/OHBM2009/planner/abstract_popup.php?abstractno=507.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackey, C. E., Filippini, N., Watkins, K. E., Toro, R., and Beckmann, A. R. L. C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31):13040–13045.

Szewczyk, M. M. (2008). Databases for neuroscience. Master's thesis, Technical University of Denmark, Kongens Lyngby, Denmark. http://orbit.dtu.dk/getResource?recordId=223565&objectId=1&versionId=1. IMM-MSC-2008-92.

Turkeltaub, P. E., Eden, G. F., Jones, K. M., and Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *NeuroImage*, 16(3 part 1):765–780.

PMID: 12169260. DOI: 10.1006/nimg.2002.1131. http://www.sciencedirect.com/science/article/-B6WNP-46HDMPV-N/2/xb87ce95b60732a8f0c917e288efe59004.

Van Essen, D. C. (2009). Lost in localization—but found with foci?! *NeuroImage*, 48(1):14–17. DOI: 10.1016/j.neuroimage.2009.05.050.

Wilkowski, B., Szewczyk, M., Rasmussen, P. M., Hansen, L. K., and Nielsen, F. Å. (2009). Coordinate-based meta-analytic search for the SPM neuroimaging pipeline. In *Proceedings of the Second International Conference on Health Informatics*, pages 11–17. INSTICC Press.