

Combining text mining and coordinate-based meta-analysis

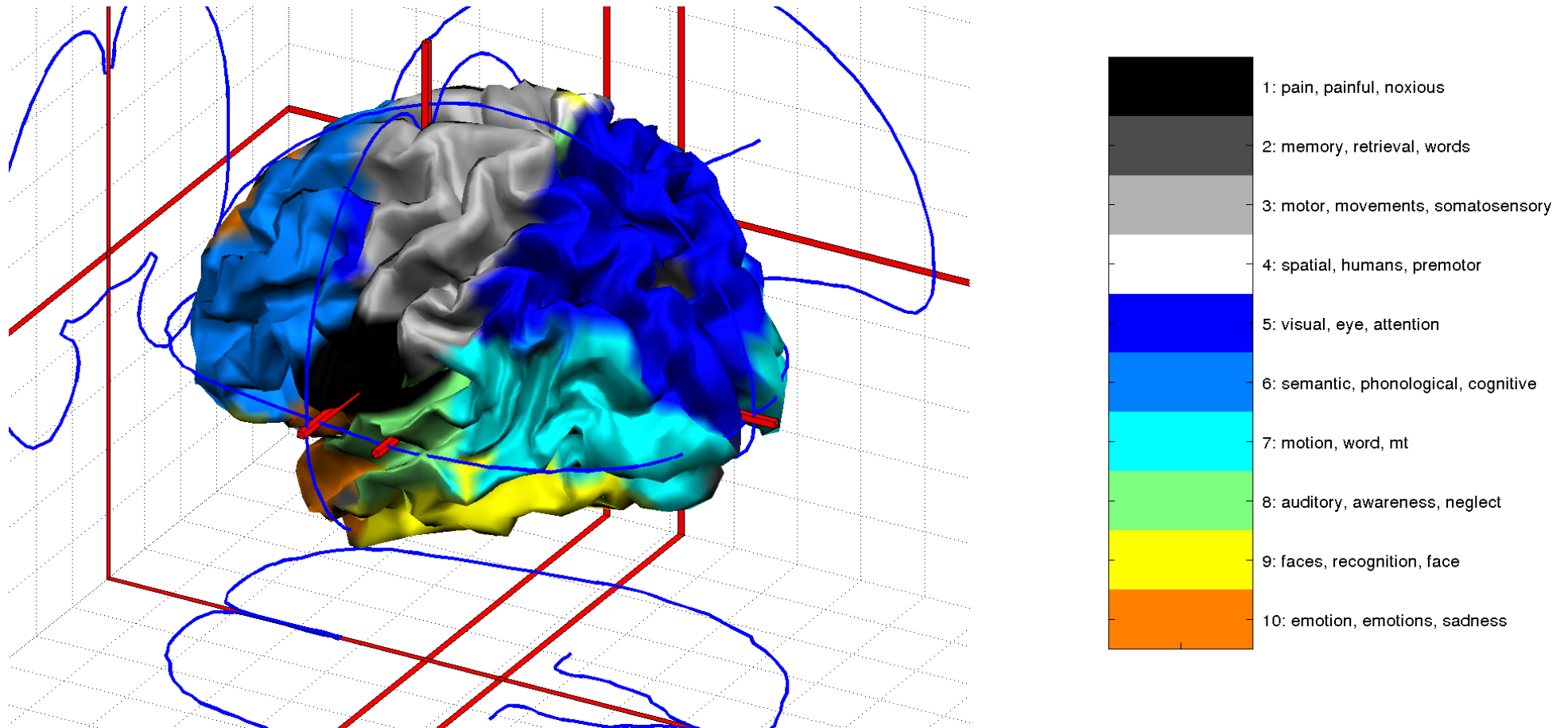
Finn Årup Nielsen

DTU Compute
Technical University of Denmark

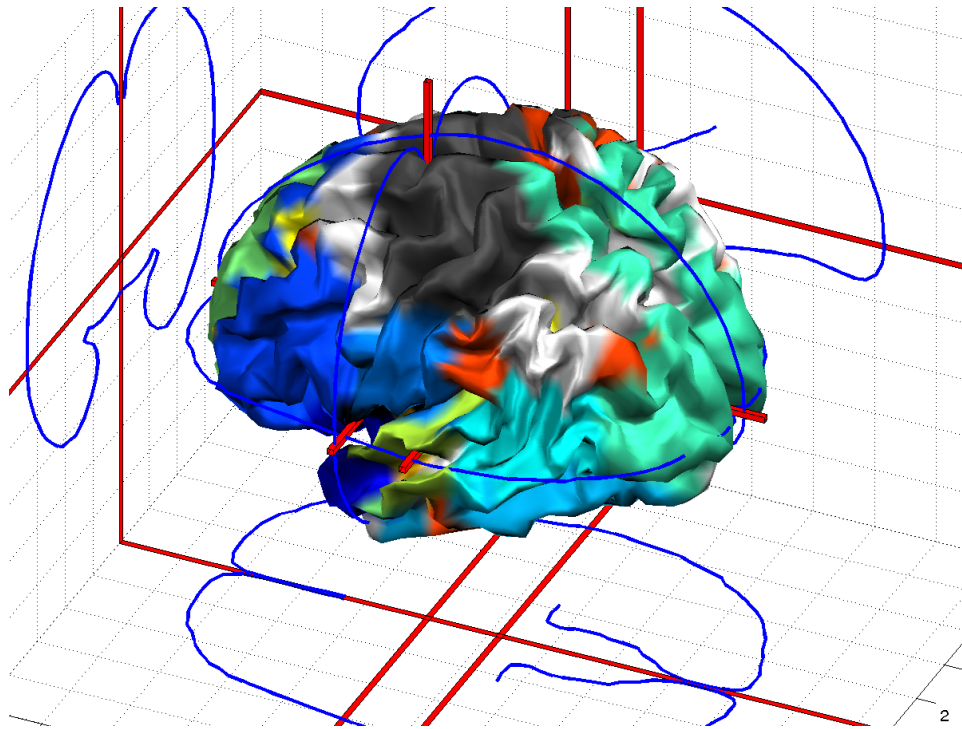
Workshop: Neuroimaging meta-analysis methods

April 17, 2015

Brain atlas constructed from abstract



Brain atlas constructed from experiment labels



1	Hot pain, Thermal pain, Warm temperature sensation
2	Finger movement, Localized movement, Motion, movement, locomotion
3	Externally generated threat response, Externally generated emotion, Threat
4	Memory, Cognition, Memory retrieval
5	Emotion, Mental process, Unpleasantness
6	Language, Rhyme judgement, Phonetic processing
7	Somesthesia, Perception, Cold pain
8	Face recognition, Objects (processing), Visual object recognition
9	Audiovisual speech perception, Multimodal perception, Congruent multimodal perception
10	Vision (visual perception), Reading, Saccadic eye movements
11	Self-reflection, Self processing, Self/other processing
12	Voice perception, Audition, Spatial neglect
13	Verbal fluency, Productive language, Silent word generation
14	Fear, Anger, S allele of promoter region of serotonin transporter gene
15	Syllable counting, Receptive language, Novelty seeking
16	Awake resting with eyes closed, Relaxed conscious state, Conscious state
17	Visuospatial attention, Visuospatial expectancy, Visuospatial processing

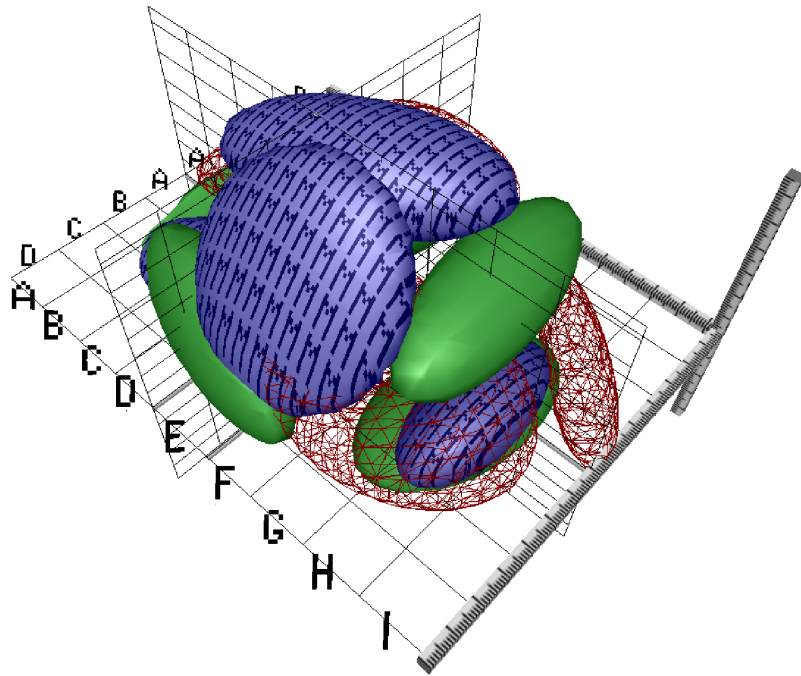
How is this done?

Coordinate-based database

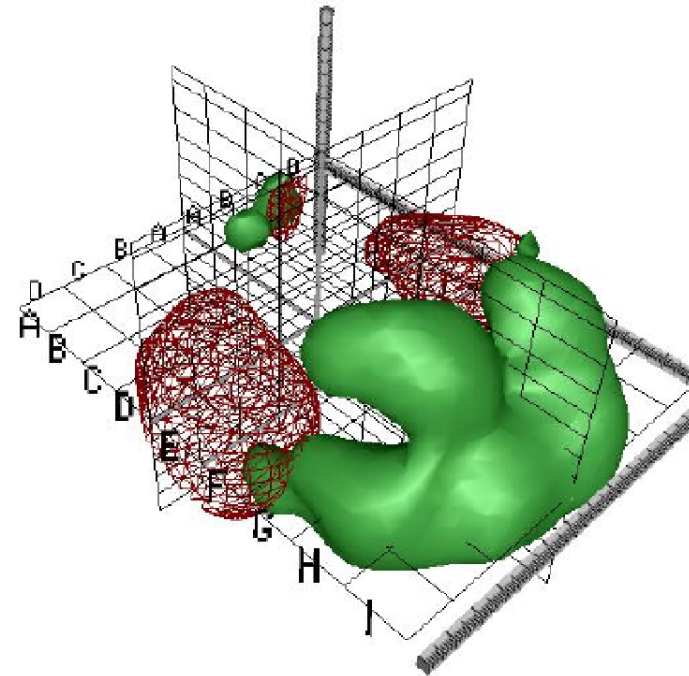
Name	'Old' BrainMap	Brede	Neurosynth
Entry	Manual	Manual	Automatic
Collection	Web scraping	XML	TSV
Papers	224	186	≈ 10,000
Experiments	771	586	> 18,000
Locations	7,263	3,912	>150,000
Abstracts	✓	✓	✓
Loc. labels	✓	✓	
Exp. labels	✓	✓	
Reference	(Fox and Lancaster, 1994)	(Nielsen, 2003)	(Yarkoni et al., 2011)

Other databases: AMAT (Hamilton, 2009), SumsDB (Van Essen, 2009). Part of Brede Database is in AMAT. 'New' BrainMap.

Modeling of data in BrainMap



Gaussian mixture model for perception, cognition, motion (Nielsen and Hansen, 1999)



Kernel density estimation for audition, vision (Nielsen and Hansen, 2000)

$p(\mathbf{x}|l)$: \mathbf{x} is Talairach space, l is a BrainMap label ('behavioral domain').

Modeling anatomical labels in BrainMap

Different perceptual tasks performed with the same visual stimulus attribute activate different regions of the human brain: A positron emission tomography study

- Paper-Id: 72
- First author: Dupont P
- Citation: Proceedings of the National Academy of Sciences, 90, p 10927-10931
- Experiment Id: 1
- Capsule Description: Discriminate
- Modality: PET
- Tracer: O-15 water
- Measured Variables: CBF
- Omnibus significance: 05 (chi-square)
- Task:
 - Behavioral Domain(s):
 - Cognition, Attention
 - Perception, Vision
 - Task states:

#	Stimulus	Resp	Notes
1	Visual, Abstracts - square-wave gratings	Limb-Motor	orientation
2	Visual, Abstracts - square-wave gratings	Limb-Motor	Press key for same orientation - ordered pairs

Location #1

- Magnitude: (-)
- p-Value: (-)
- Coordinates in Talairach, 1988 space:
 - X=3.6
 - Y=-7.6
 - Z=1.2
- Coordinates as reported in experiment:
 - X=3.6
 - Y=-7.6
 - Z=1.2
- Corresponding Brodmann's area: -
- Functional area: occipital gyrus
- Lobar outline: occipital gyrus
- Lobar outline: occipital gyrus
- Left/right-type: 1
- Report/estimate: -
- Point-Type: -

“We downloaded 7,263 location web-pages and 3,935 of these locations had an associated anatomical label” (Nielsen and Hansen, 2002b).

$p(x|l)$: x is Talairach space, l is the anatomical label (word or phrase) associated with each coordinate, e.g., “occipital gyrus”, “gyrus”, “occipital”.

1,231 word/phrases.

Volume available from: neuro.compute.dtu.dk.
as Analyze files: “Meta-analytic region of interest atlas” (Nielsen and Hansen, 2002a).

Modeling anatomical labels in BrainMap

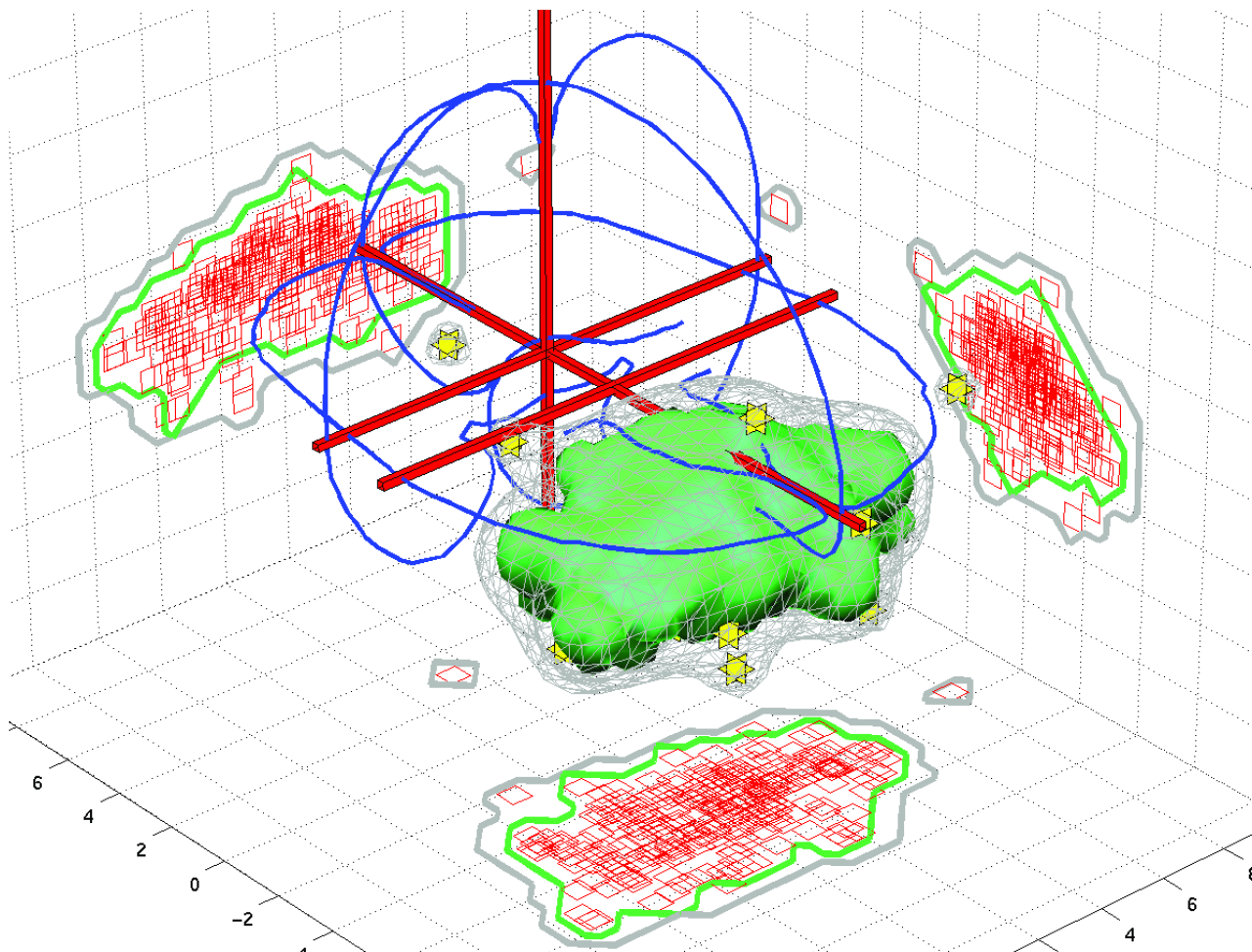
Kernel density estimation with leave-one-out cross-validation for determining the width of the kernel (Nielsen and Hansen, 2002b).

$$E(\sigma^2, l) = - \sum_{n=1}^{N_l} \log p_{-n}(\mathbf{x}_n | \sigma^2, l)$$

$$p_{-n}(\mathbf{x} | \sigma^2, l) = \frac{1}{N_l - 1} \sum_{n' \neq n}^{N_l} (2\pi\sigma^2)^{-3/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{x} - \mathbf{x}_{n'})^2\right)$$

Lars Kai Hansen's algorithm with Newton optimization for optimizing the kernel width. Available in [Brede Toolbox](#) (Matlab).

Modeling anatomical labels in BrainMap



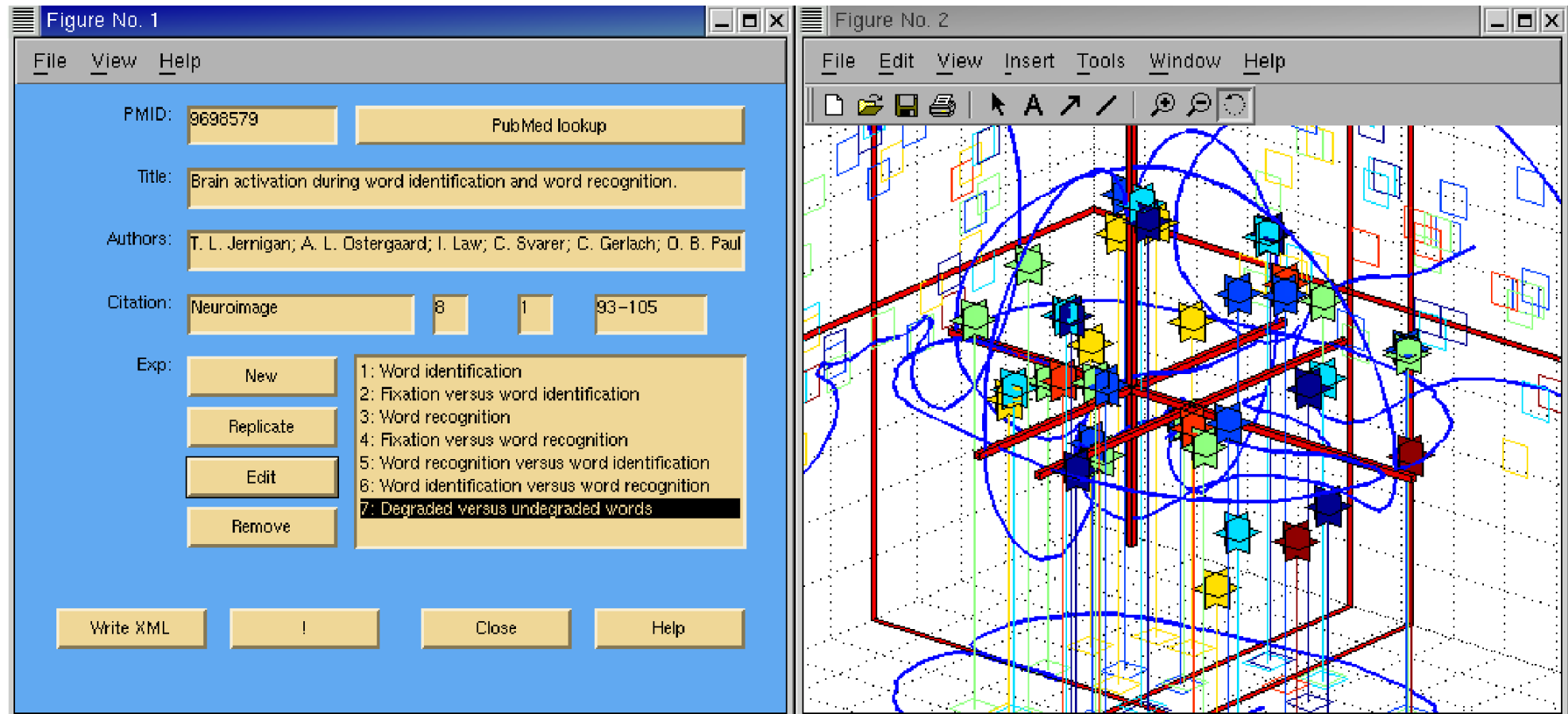
Finding outliers in BrainMap by looking at coordinates located in areas with low probability density.

Here $p(\mathbf{x}|l = \text{cerebellum})$ with a coordinate in the anterior part of the brain

Typos: centimeter/millimeter, left/right, ...

Algorithm also run for the [Brede Database](#).

Brede Database



Matlab graphical user interfaction program for data entry and visualization of studies with Talairach coordinates.

Data entry of brain coordinates

Neuroimaging article

Table 1. Brain activation during anticipation period.

Brain Region (Hemisphere)	MNI coordinate			Cluster Size (Voxels)	T	P
	X	Y	Z			
Insula (L)	40	5	12	148	7.42	0.008
Laudum Node (R)	36	-3	18	77	6.98	0.015
Superior Temporal Gyrus (R)	50	-46	10	76	6.77	0.015
Premotor Cortex (L)	-46	-8	14	76	6.62	0.016
Anterior Insula (L)	-34	14	19	67	6.61	0.016
Laudum Node (L)	-34	-12	2	62	6.52	0.016
Inferior Parietal Lobule (L)	-54	-38	30	31	5.96	0.079
Caudate Gyrus (L)	-4	14	30	32	5.51	0.109

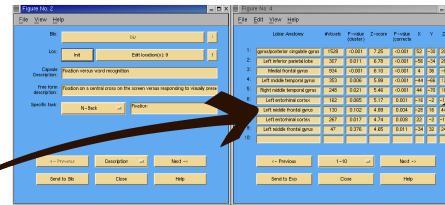
The activation effect could not be replicated with ungrouped family-wise error correction (FWE) and most cases of 30 for post-hoc family-wise error correction.

Studies did not report in details the subjects' behavioral data in regard to the preference to find-debit another.

A few studies out of over one hundred using the similar four-task format and addressing the chosen frequency had consistently demonstrated that the high-frequency group (A and B) were preferentially selected from A and C in the control group (5.8% vs 15.1%, 22.15-43.4-4%). The "premotor task B phenomenon" indicated that subjects were overwhelmed by the high-frequency gain of tasks B, C, and also C (as opposed to A, C, but more specifically to not value another, a real one, this

Brain activation for anticipation and experience

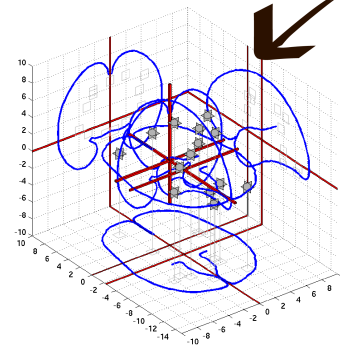
In the current study, the 30 and 30 were strongly associated with events of uncertain situations. The observation does not completely support the basic assumption of the ICI or SMH neural loop (14) which did not point that ICI participants in the anticipation of uncertainty (Figure 3). The consistent activation of these two structures might reflect that positive and negative emotions were simultaneously involved (8) by this uncertain monetary gains, which led to a higher arousal state (34-36). The arousal might indicate the subject into a self-activated situation where better neuroanatomical conditions can be used in



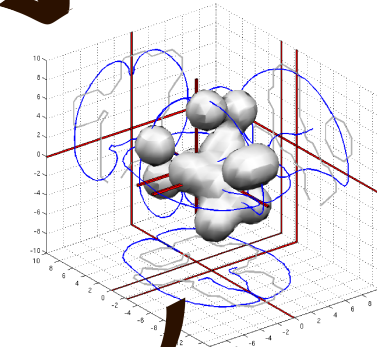
XML storage

```
<AnalysisSoftware>MIR</AnalysisSoftware>
<AnalysisSoftware>MIR</AnalysisSoftware>
<Loc>
  <Type>Loc</Type>
  <FunctionalArea>Left frontal eye field</FunctionalArea>
  <Brodmann></Brodmann>
  <Score>4.82</Score>
  <CoordReported>0.000000 -0.002000 0.036000</CoordReported>
  <Reported>-0.050000</Reported>
  <Reported>0.000000</Reported>
  <Reported>0.000000</Reported>
  <Coord>0.050000 -0.002000 0.036000</Coord>
  <Sig>-0.002000</Sig>
  <Sig>0.036000</Sig>
</Loc>
<Loc>
  <Type>Loc</Type>
  <FunctionalArea>Right frontal eye field</FunctionalArea>
  <Brodmann></Brodmann>
  <Score>4.52</Score>
  <CoordReported>0.040000 -0.000000 0.048000</CoordReported>
  <Reported>0.040000</Reported>
```

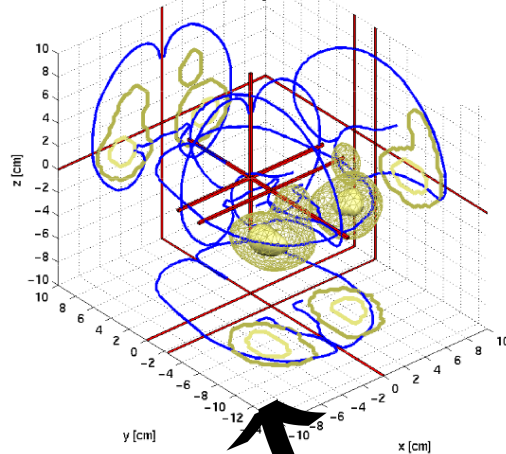
3D visualization



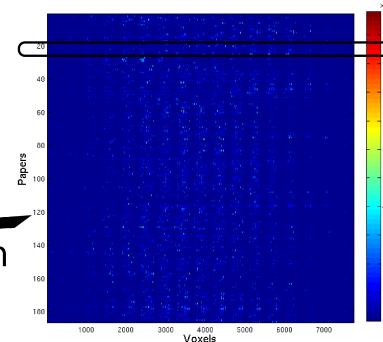
Kernel density



NMF component



Matrix of kernel densities



One row

Non-negative matrix factorization

Brede Database neuroanatomy taxonomy



Hierarchy of **brain regions**.

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

Fields recorded: Canonical name, variation in names, abbreviations, links to NeuroNames and other databases.

Graph constructed with GraphViz (Gansner and North, 2000).

Functional segregation in brain regions

For a brain region = 1 to 313 brain regions:

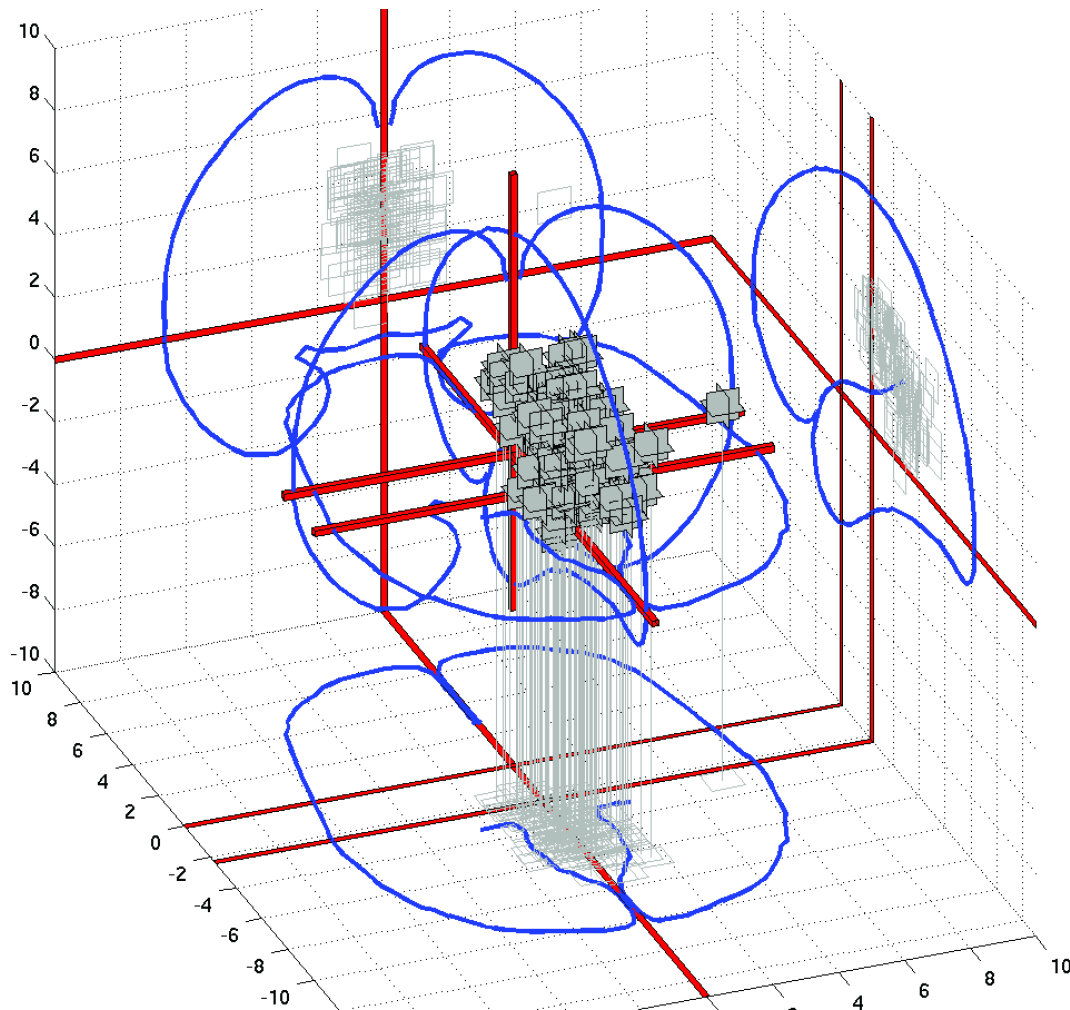
- Step 1: Get all coordinates for the specific area, build a density model, exclude coordinates that are outliers
- Step 2: Determine themes of the brain area with text mining on abstracts that contain coordinates within the brain area
- Step 3: Determine whether specific themes are spatially clustered in the brain area by testing whether two sets of coordinates are separated.

end

Step 4: Intertwine results from all brain regions

(Nielsen et al., 2006; Nielsen et al., 2005).

Step 1: Identify coordinates



Simple SQL-like command in Matlab to find locations

Corner cube visualization of 116 “posterior cingulate” coordinates found

An outlier: “Right postcentral gyrus/posterior cingulate gyrus” from (Jernigan et al., 1998).

Build kernel density estimate of the coordinates.

Step 2: Bag-of words matrix

	'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	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮

For the further analysis: Include all papers that contain one or more of coordinates found.

Representation of the abstracts of the papers in a bag-of-words matrix: (abstract \times words)-matrix $\equiv \mathbf{X}(N \times P)$.

Exclude a large list of word: Anatomical, "stop words", ...

Step 2: Non-negative matrix factorization

Non-negative matrix factorization (NMF) decomposes a non-negative data matrix $\mathbf{X}(N \times P)$ (Lee and Seung, 1999)

$$\mathbf{X} = \mathbf{WH} + \mathbf{U}, \quad (1)$$

where $\mathbf{W}(N \times K)$ and $\mathbf{H}(K \times P)$ are also non-negative matrices.

“Euclidean” cost function for

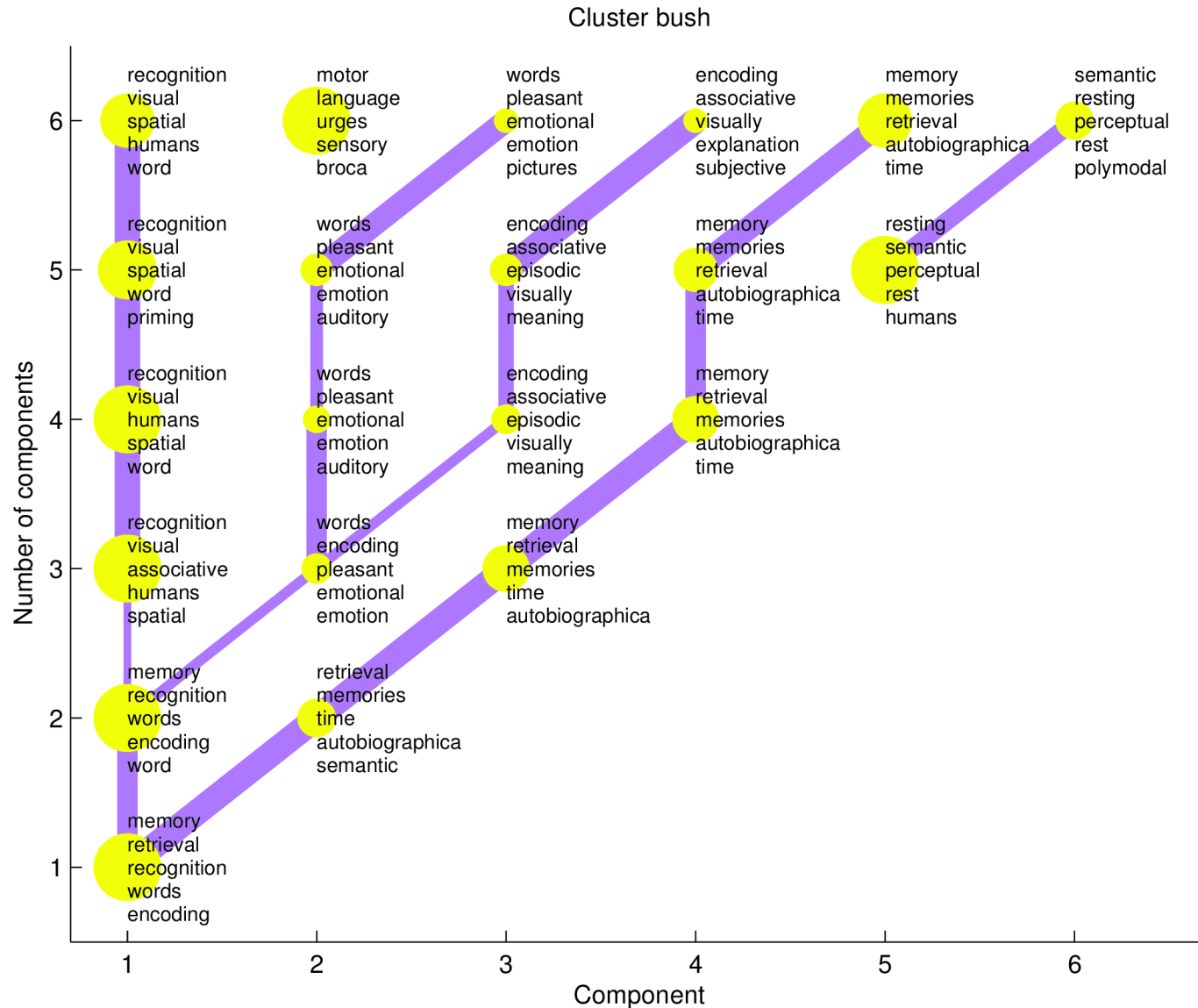
$$E_{\text{“eucl”}} = \|\mathbf{X} - \mathbf{WH}\|_F^2 \quad (2)$$

Iterative algorithm (Lee and Seung, 2001)

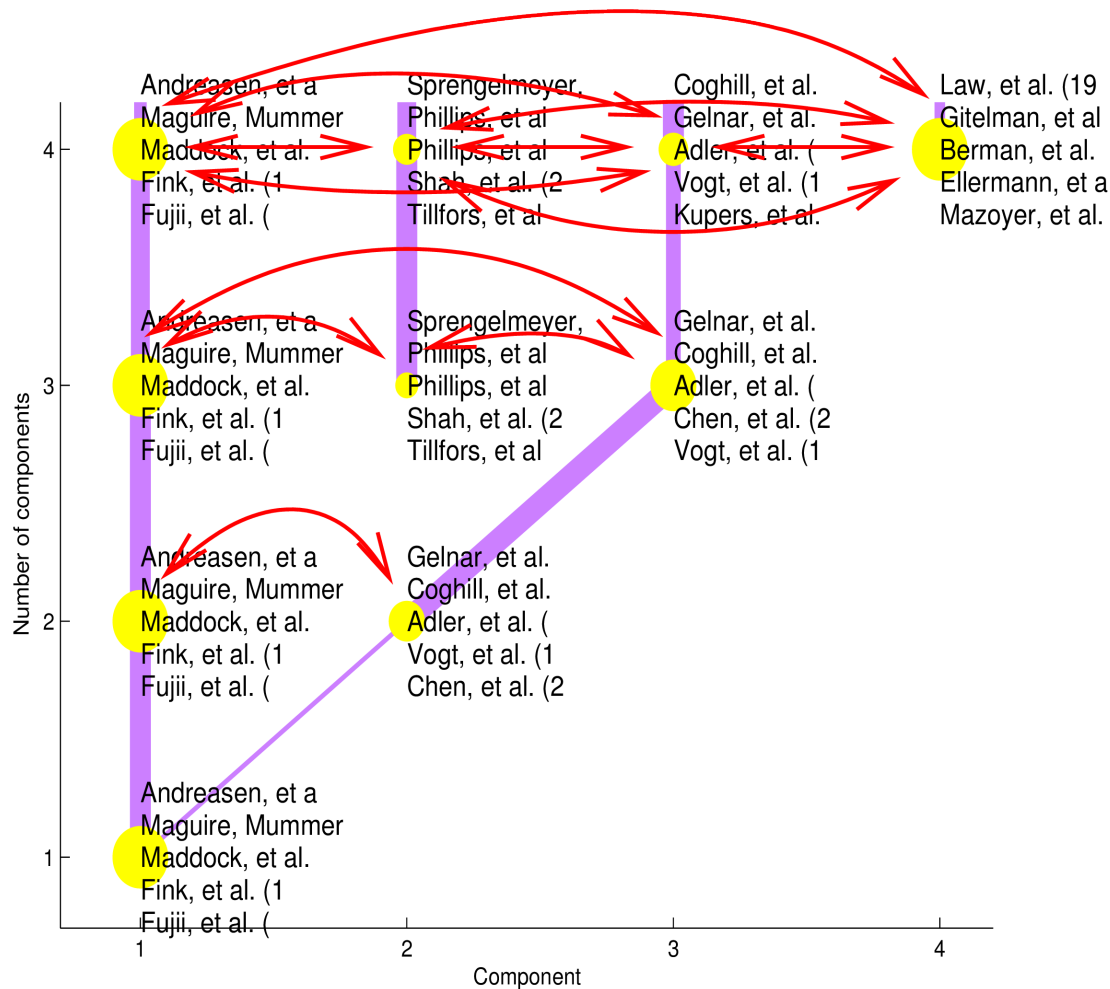
$$\mathbf{H}_{kp} \leftarrow \mathbf{H}_{kp} \frac{(\mathbf{W}^\top \mathbf{X})_{kp}}{(\mathbf{W}^\top \mathbf{WH})_{kp}} \quad (3)$$

$$\mathbf{W}_{nk} \leftarrow \mathbf{W}_{nk} \frac{(\mathbf{XH}^\top)_{nk}}{(\mathbf{WHH}^\top)_{nk}}. \quad (4)$$

Step 2: 'Medial temporal lobe' NMF result



Step 3: Test spatial distribution

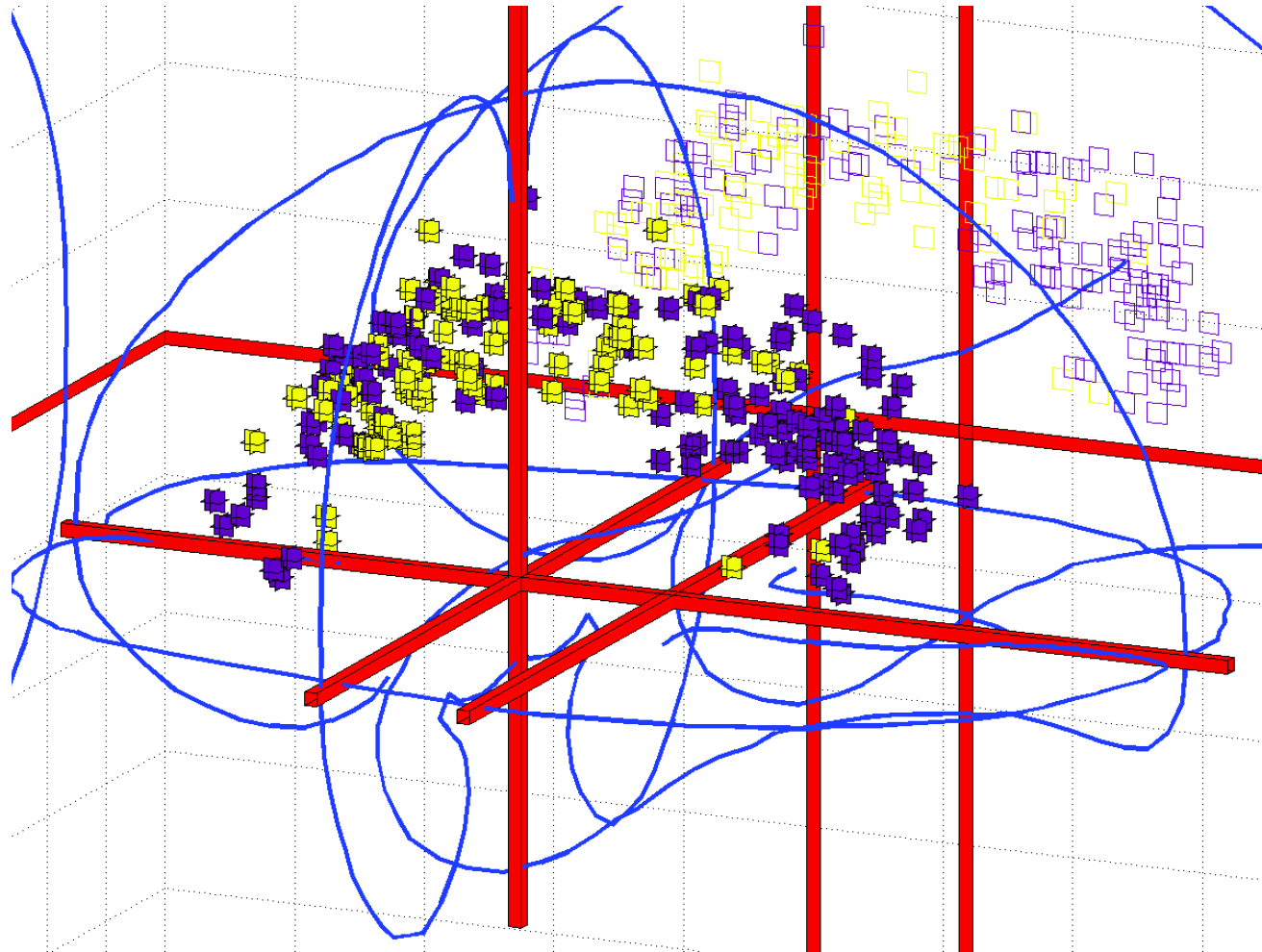


Extract locations from grouped papers.

Test if the spatial distribution of locations for a group is different from the distribution from an other group with Hotelling's T^2 or convex hull pelling permutation test.

All possible tests within a level of non-negative matrix factorization are performed.

Step 4: 'Cingulate gyrus'



Coordinates associated with topics on: pain, painful vs. memory, retrieval

Combining text and coordinate directly

Construct a functional parcellation of the brain based on combined text and coordinate analysis (Nielsen et al., 2004; Nielsen, 2009).

Brede Database with neuroimaging papers.

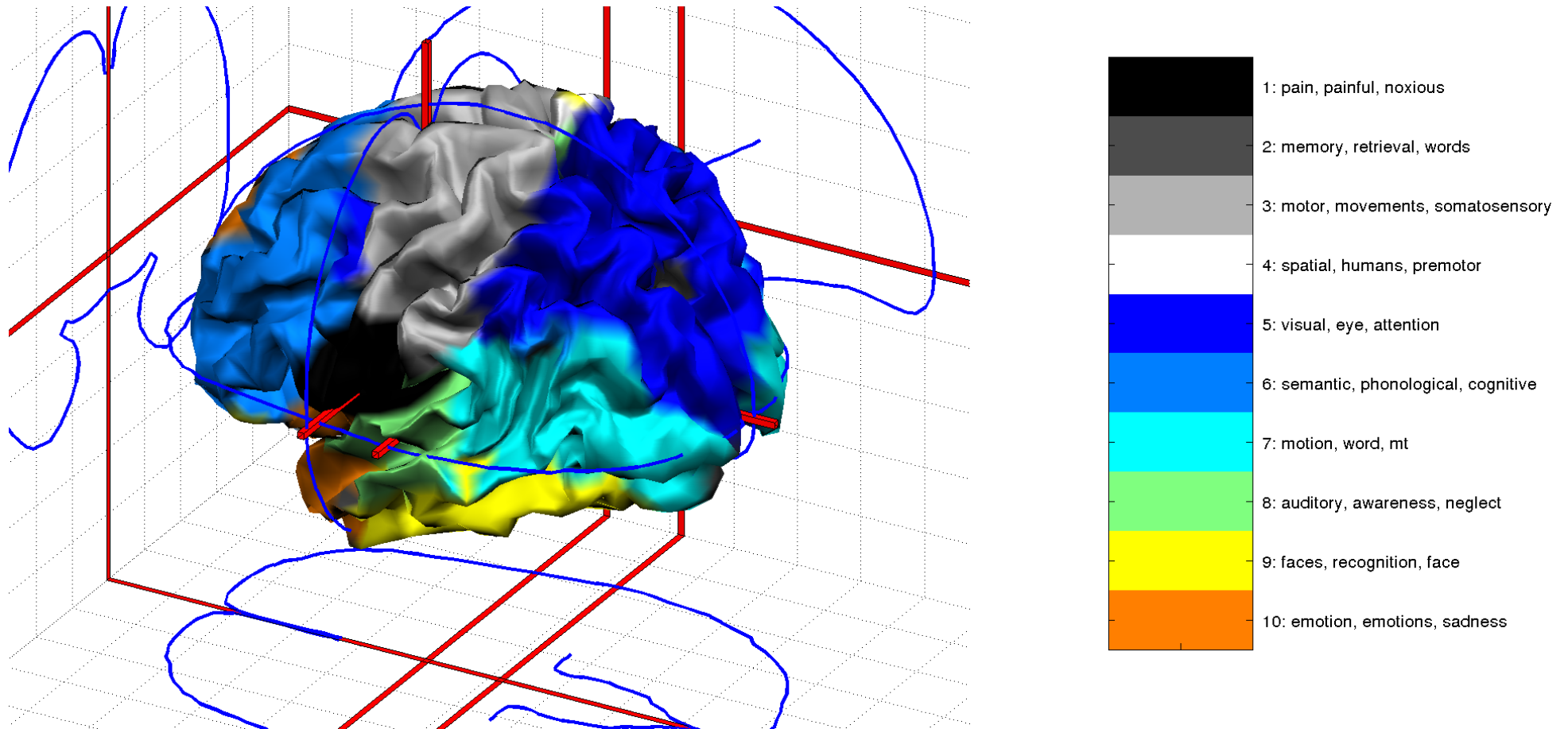
Bag-of-words matrix from abstracts, scaling, and stop words elimination getting a (paper \times words)-matrix, \mathbf{X}_1

Kernel density estimates from coordinates contained in the papers getting a (papers \times vertices)-matrix or (papers \times voxels), \mathbf{Y} .

Product matrix: $\mathbf{Z}_1 = \mathbf{Y}'\mathbf{X}_1$ getting a (vertices \times labels)-matrix.

Approximative non-negative matrix factorization: $\mathbf{WH} \approx \mathbf{Z}_1$ getting a (vertices \times topics)-matrix, \mathbf{W} and a (topic \times words)-matrix, \mathbf{H} .

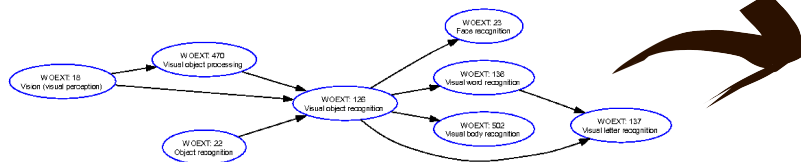
Functional parcellation of the cortex



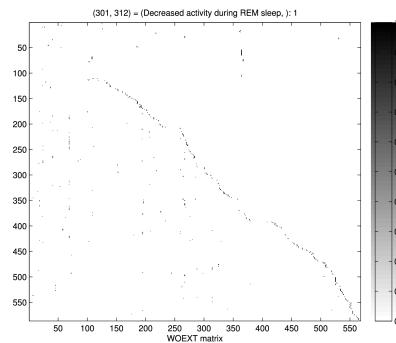
Winner-take-all function on \mathbf{W} (the surface) and \mathbf{H} (the legend).

Brain function ontology

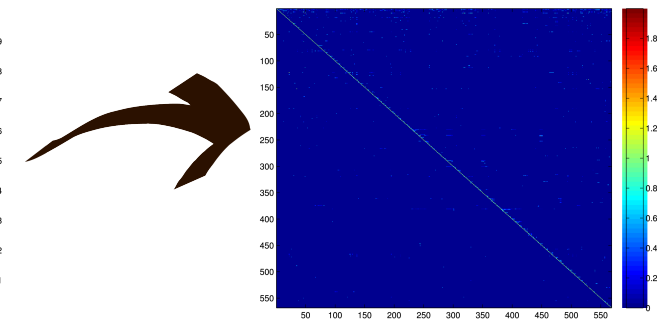
Brain function ontology



Adjacency matrix



"Recursed" adjacency matrix



Experiment label ('external components', brain function ontology, cognitive ontology) organized in a directed graph (Nielsen, 2005).

This graph can be converted to an adjacency matrix.

An experiment (i.e. contrast) in Brede Database may be labeled with an item from the ontology.

Functional parcellation with cognitive ontology

Brede Database with neuroimaging paper.

Adjacency matrix of experiment label ('external components', brain function ontology) graph (C) propagated $D = \sum_i (\lambda C)^i$.

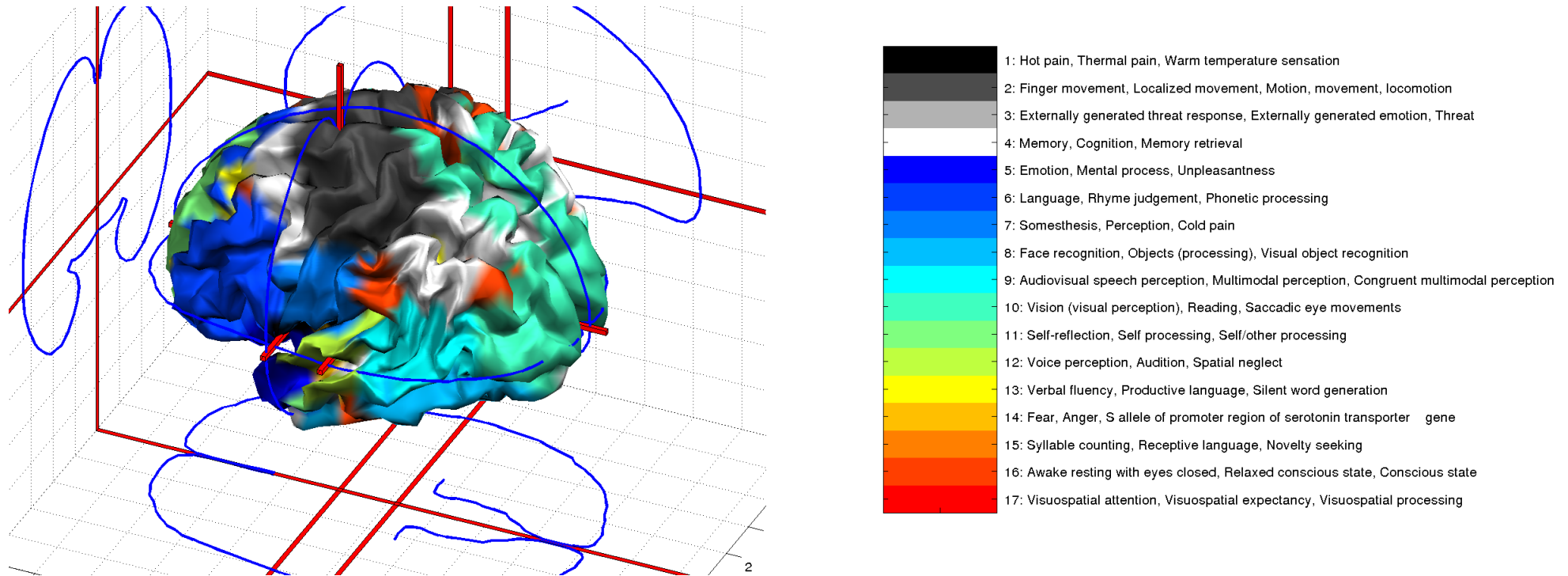
Propagated experiment label (experiments \times labels)-matrix $X_2 = TD$, where the (experiments \times experiment labels)-matrix T represents the (manual) label of each experiment.

Kernel density estimates from coordinates contained in the experiment getting a (experiments \times vertices)-matrix or (experiments \times voxel)-matrix (Y).

Product matrix: $Z_2 = Y'X_2$ getting a (vertices \times experiment labels)-matrix.

Approximative non-negative matrix factorization: $WH \approx Z_2$ getting a (vertices \times topics)-matrix (W) and a (topic \times words)-matrix (H)

Brain atlas constructed from experiment labels



Winner-take-all function on \mathbf{W} (the surface) and \mathbf{H} (the legend with items from the experiment labels/ontology).

Brede tools

Presently developing a Python Brede Package with features such as Features: Handling of data from Brede Wiki, Neurosynth, word lists, data sets, surfaces, . . . : <https://github.com/fnielsen/brede>

Brede Database: <http://neuro.compute.dtu.dk/services/brededatabase/>

Brede Toolbox (matlab): <http://neuro.compute.dtu.dk/software/brede/>

Brede Wiki: <http://neuro.compute.dtu.dk/wiki/>.

References

- Bowden, D. M. and Martin, R. F. (1995). NeuroNames brain hierarchy. *NeuroImage*, 2(1):63–84. ISSN 1053-8119.
- Fox, P. T. and Lancaster, J. L. (1994). Neuroscience on the net. *Science*, 266(5187):994–996.
- Gansner, E. R. and North, S. C. (2000). [An open graph visualization system and its applications to software engineering](#). *Software — Practice and Experience*, 30(11):1203–1234. ISSN 00380644.
- Hamilton, A. F. (2009). Lost in localization: a minimal middle way. *NeuroImage*, 48(1):8–10.
- Jernigan, T. L., Ostergaard, A. L., Law, I., Svarer, C., Gerlach, C., and Paulson, O. B. (1998). Brain activation during word identification and word recognition. *NeuroImage*, 8(1):93–105. WOBIB: [35](#).
- Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791.
- Lee, D. D. and Seung, H. S. (2001). [Algorithms for non-negative matrix factorization](#). In Leen, T. K., Dietterich, T. G., and Tresp, V., editors, *Advances in Neural Information Processing Systems 13: Proceedings of the 2000 Conference*, pages 556–562, Cambridge, Massachusetts. MIT Press.
- Mai, J. K., Assheuer, J., and Paxinos, G. (1997). *Atlas of the Human Brain*. Academic Press, San Diego, California. ISBN 0124653618.
- Nielsen, F. Å. (2003). [The Brede database: a small database for functional neuroimaging](#). *NeuroImage*, 19(2). Presented at the 9th International Conference on Functional Mapping of the Human Brain, June 19–22, 2003, New York, NY.
- 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. *Coordinate-based meta-analysis on data from a neuroinformatics database across many different brain functions*.

- 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., Balslev, D., and Hansen, L. K. (2006). [Data mining a functional neuroimaging database for functional segregation in brain regions](#). In Olsen, S. I., editor, *Den 15. Danske Konference i Mønstergenkendelse og Billedanalyse*, Copenhagen, Denmark. The Department of Computer Science, University of Copenhagen.
- Nielsen, F. Å. and Hansen, L. K. (1999). [Modeling of BrainMap data](#). Online at http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/4711/pdf/imm4711.pdf.
- Nielsen, F. Å. and Hansen, L. K. (2000). [Functional volumes modeling using kernel density estimation](#). Online at http://www.imm.dtu.dk/pubdb/views/edoc_download.php/4688/pdf/imm4688.pdf. *Description of kernel density modeling of Talairach coordinates. Examples are shown with modeling the spatial distribution of Talairach coordinates from the BrainMap database based on the database labels.*
- Nielsen, F. Å. and Hansen, L. K. (2002a). [Automatic anatomical labeling of Talairach coordinates and generation of volumes of interest via the BrainMap database](#). *NeuroImage*, 16(2). Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan.
- Nielsen, F. Å. and Hansen, L. K. (2002b). [Modeling of activation data in the BrainMap™ database: Detection of outliers](#). *Human Brain Mapping*, 15(3):146–156. DOI: [10.1002/hbm.10012](#).
- Nielsen, F. Å., Hansen, L. K., and Balslev, D. (2004). [Mining for associations between text and brain activation in a functional neuroimaging database](#). *Neuroinformatics*, 2(4):369–380.
- 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](#).
- Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. (2011). [Large-scale automated synthesis of human functional neuroimaging data](#). *Nature Methods*, 8(8):665–670. DOI: [10.1038/NMETH.1635](#). *Describes the Neurosynth system for collected, extracting and analysis of stereotaxic coordinates in human brain mapping publication.*