THE SPEECH ACTION REPOSITORY:
EVIDENCE FROM A SINGLE CASE NEUROIMAGING STUDY

Cornelia Eckers [1], *Bernd J. Kröger* [1], *and Stefan Heim* [2, 3, 4, 5]

[1]*Department of Phoniatrics, Pedaudiology, and Communication Disorders,*
RWTH Aachen University
[2]*Section Structural Functional Brain Mapping, Department of Psychiatry,*
Psychotherapy and Psychosomatics, Medical School, RWTH Aachen University
[3]*Research Centre Jülich, Institute of Neuroscience and Medicine (INM-1),*
Research Centre Jülich
[4]*Section Clinical and Cognitive Neurosciences, Department of Neurology, Medical School,*
RWTH Aachen University
[5]*JARA – Translational Brain Medicine, Jülich and Aachen*

(ceckers, bkroeger, sheim)@ukaachen.de

Abstract: The speech-action-repository (SAR) is a neurofunctional and neurocomputational model of syllable processing. The model is capable of storing sensorimotor representations of high-frequent syllables by a supramodal hub and its connections to unimodal sensorimotor state maps. In order to support the notion of the SAR, a functional imaging study was conducted. Within the fMRI-experiment a single case subject responded overtly and covertly to different visually and auditory presented homogeneous and heterogeneous syllable blocks. An auditory-visual conjunction analysis for determining this distributed supramodal hub, revealed an activation network, comprising bilateral precentral gyrus, left inferior frontal gyrus (area 44), left supplementary motor area, and bilateral superior temporal gyrus. The analysis of main effect of syllable priming, i.e. heterogeneous vs. homogeneous syllable-blocks, in order to detect neural activation relating to access of the state maps, revealed an activation pattern, distributed over the frontal, temporal, and parietal lobe. These results are compatible in agreement with the notion of the SAR. Thus, the present study provides evidence for the neural representation of the SAR, which is in line with the Kröger model.

1 Introduction

A model of syllable processing, called speech-action-repository (SAR), was suggested by Kröger et al. (2011) [19, 18]. This hypothetical neurofunctional and neurocomputational model of syllable processing is capable of storing and activating sensorimotor representations of high frequent syllables during syllable production as well as perception. This is possible by two different types of maps which are part of the model, i.e. a self-organizing higher level phonetic map and higher level sensorimotor state maps. The phonetic map is implemented as supramodal long-term memory hub (see Fig. 1), which stores the associations, i.e. neural patterns of motor and sensory states for related high frequent syllables. This phonetic map is interconnected by bidirectional links to the state maps, i.e. the sensory and motor short-term memory state map (Fig. 1). An isolated activation within the phonetic map leads to simultaneous co-activation of the different state maps by syllable-specific stored link weights.

This approach of a higher level supramodal hub and its associations to higher level unimodal sensorimotor state maps is different to other speech processing models, like the DIVA model of [15] and the model of [20, 21], which also deal with syllable processing. A cortical hub, like in our approach, is a region with high degree or high centrality which is crucial for ensuring that overall path lengths across a network are short [4]. Based on this model and the findings to cortical hubs, it is hypothesized that the SAR or syllable processing is cortical distributed, but centralized by one or more hubs. The assumption of more than one hub (not shown in Fig. 1) is based on anatomically aspects and the structure of the SAR. As it can be seen in Figure 1, the model maps are connected by neural mappings, which are dense between the phonetic map and each state map. Because it is assumed, that the state maps are cortical distributed, that means in frontal, parietal as well as temporal regions [16], activation of syllable representations forced a lot of energy. This would be less, if there are two or more hubs cortical represented, which are simple connected to each other in neuron-by-neuron connections. Thus, state maps, which are located nearby hub are connected to it and the hubs among themselves are connected with each other in order to realize a more efficient and economic activation process.

In order to support the assumption of a SAR comprising a supramodal hub (phonetic map) as well as unimodal state maps (phonemic state map, auditory state map, somatosensory state map, and motor plan state map) data by a functional magnetic resonance tomography (fMRI) in a single subject were analyzed using two different analyses: (1) a conjunction analysis was conducted in order to find the neural representation of the supramodal hub(s), (2) an analysis of the main effect of syllable priming contrasting heterogeneous and homogeneous syllable blocks (see below) was conducted in order to find neural representations of different state maps.

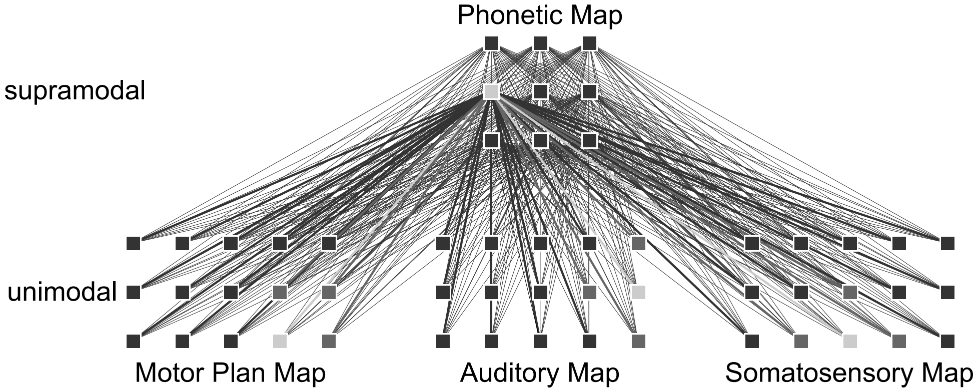

Figure 1 - Dense interconnections of supramodal neural phonetic map with unimodal neural state maps: Syllable activation within the P-Map leads to co-activation of every neuron within the state maps with different activation levels; light = full activated; dark = zero activated

2 Methods

In order to support the assumption of one or more supramodal hubs (phonetic maps) connecting unimodal state maps (e.g. auditory state map) during syllable processing, a single case fMRI experiment was created. The subject was male, 27 years old, native German speaker, right handed, and was free from any psychiatrical and neurological diseases. The reaction paradigm comprised four different conditions, controlling input modalities, i.e. visual and auditory presentation mode, as well as output modalities, i.e. overt and covert speech response. The four conditions arised out of each combination of an input and an output modality. Each task lasted about 17 minutes. A sparse scanning procedure was used that allowed the subject to produce

Table 1 - Schematic representation of the processes taking place during the four different conditions. Condition 1 = read, 2 = repeat, 3 = silent read, 4 = silent repeat. Syllable processing is the process which all conditions have in common (right column). This principle is used in the conjunction analysis.

	Condition				
Process	1	2	3	4	(1∩2∩3∩4)
Auditive perception	x		x		
Visual perception		x		x	
Syllable processing	**x**	**x**	**x**	**x**	**x**
Overt production	x	x			
Covert production			x	x	

utterances in relative silence and avoids movement-related artifacts. Each of these different conditions consisted of homogeneous syllable blocks, including one of ten possible CV syllables with [b] or [?] (glottal stop) in every combination with [aː], [eː], [iː], [oː], and [uː] and heterogeneous syllable blocks, including every [b]- or [?]-combination two times. It was assumed that during homogeneous blocks the hemodynamic response (BOLD-effect) decreases within the SAR (priming effect). A smiley appeared after each stimulus cueing the subject to speak now. There were ten different homogeneous blocks and two different heterogeneous blocks in each condition. Furthermore each of these blocks was repeated including a target [ɛː] or [bɛː] randomly presented in order to hold concentration. Totally there were 20 homogeneous blocks and 4 heterogeneous blocks randomly presented to the subject during each task. Each block lasted 40 seconds (s), including 10 stimuli (each 1 s), 10 smileys (each 1 s), including pauses to the next stimulus (1 s), and if appropriate a target with smiley and pause (3 s), and further a 7 s pause to the following block. The subject had to react with a button press when he saw or heard a target. Blocks without a target included a 3 s pause randomly inserted in the block instead.
The subject participated four times in the same experiment. These repeated measurements were realized in order to investigate whether reproducible activation patterns are observed.

3 Data analysis

The experiment and data acquisition took place within a Siemens Magnetom Trio 3T Scanner. We obtained T2* weighted functional images (time echo (TE) = 40 ms, time repetition (TR) = 3000 ms, flip angle = 90°, 39 slices, field of view (FOV) = 192 mm) using Echo Planar Imaging (EPI) acquisition. Each functional sequence consisted of thirty-nine 1.9 mm thick axial slices, positioned to image around the perisylvian fissure of the brain. After the experiment a T1 weighted anatomical volume was obtained using magnetization-prepared rapid acquisition with gradient echo (MP-RAGE) sequence of about 9 min 50 s (TE = 3.03 ms, TR = 2300 ms, FOV = 256 mm, slice thickness 1mm, 176 slices, flip angle = 9°). A total of 1352 scans were acquired for each subject. Functional MRI data were preprocessed using standard SPM8 procedures, i.e. realignment (timing and motion correction), normalization ton MNI space (a brain template) and smoothing (8 mm FWHM Gauss Kernel) [3], on Matlab 7.10 platform (MathWorks Inc., Natick, MA). Preprocessed data were statistical evaluated using a block-design analysis and using on a least-squares estimation with the general linear model for serially auto-correlated observations [7, 9, 10, 25]. The design matrix was generated with a synthetic haemodynamic response function [8, 17]. The δ-functions of the stimulus onsets for each condition (read, repeat, silent-read, silent-repeat) were convolved with the canonical haemodynamic response with a distribution of 33 s [17]. Each condition was contrasted against the implicit (resting) baseline, yielding the beta estimates for each condition. These data were analyzed subsequently by calculating the logical conjunction of visual and auditory conditions during overt as well as covert speech at a level of $p < 0.001$ in order to find the neural representation of the higher level supramodal hub(s) (phonetic map). A second analysis calculated main effects of syllable priming including contrasted heterogeneous and homogeneous blocks at a level of $p < 0.001$, in order find neural representations of the higher level unimodal state maps of the SAR. The SPM8 Anatomy Toolbox was used to identify the cytoarchitectonic localization of the effects and to compare the two different analyses. Last, a calculation of the effects of interest (EOI) in the resulted activated regions was made in order to investigate the reproducibility of the current findings in this single subject.

4 Results

The auditory-visual conjunction analysis, i.e. supramodal hub (phonetic map) detection analysis, revealed an activation network comprising bilateral precentral gyrus (PrCG), left inferior frontal gyrus (IFG, area 44), and left supplementary motor area (SMA) as well as bilateral superior temporal gyrus (STG, see Tab. 2).

The analysis of main effect of syllable priming, in order to detect the state maps, revealed an activation pattern of frontal areas, i.e. bilateral IFG (area 44/45), left PrCG, as well as left SMA, and parietal areas, i.e. left superior parietal lobe (SPL), inferior parietal cortex (IPC) as well as area 2, and temporal areas, i.e. bilateral STG (Tab. 3 and Fig. 3). Activations in IFG, PrCG, and STG were indeed found in both analyses, but are located in different parts (see Fig. 3).

The analysis of reproducibility by effects of interest (EOI) of the left hemisphere within the resulted activated regions of supramodal syllable processing, i.e. (distributed) phonetic map, during every condition at every time, revealed a consistent but variable activation pattern. Activation during overt response is stronger than during covert response (see Tab. 4 and as example Fig. 2).

Table 2 - Activated regions during supramodal syllable processing (voxel size $\geq$20). References to cytoarchitectonic maps: area 17/18: [1]; area 6: [11]; area 1: [12, 13]); area PFm: [5]; area 44: [2]. Cluster overlap with cytoarchitectonic areas is listed if it exceeds 10%.

Cluster Size (Voxels)	Local Maximum in Macroanatomical Structure	x	y	z	Percent of Cluster Volume in Cytoarchitectonic Area	
Cluster 1 (5113)	Area 17	-6	-93	-2	14.9	left Area 17
					13.3	left Area 18
					10.8	right Area 17
Cluster 2 (308)	left PrCG	-50	-10	53	56.3	left Area 6
					11.4	left Area 1
Cluster 3 (105)	left STG	-60	-47	15		
Cluster 4 (126)	right PrCG	57	-8	47	61.8	right Area 6
Cluster 5 (88)	right STG	52	-41	12	15.3	right IPC (PFm)
Cluster 6 (51)	left IFG (Area 44)	-62	4	13	48.5	left Area 44
Cluster 7 (40)	left SMA	-5	-13	71	97.8	left Area 6

Table 3 - Activated regions during syllable priming (voxel size ≥220). References to cytoarchitectonic maps: area hIP3/7A: [24, 23]; areas 44/45: [2]; areas PF/PFt: [5]; area 2:[14]; area 6: [11]. Cluster overlap with cytoarchitectonic areas is listed if it exceeds 10%

Cluster Size (Voxels)	**Local Maximum in Macroanatomical Structure**	**x**	**y**	**z**	**Percent of Cluster Volume in Cytoarchitectonic Area**	
Cluster 1 (1284)	left SPL	-30	-66	59	37.9	left SPL (7A)
					10.1	left hIP3
Cluster 2 (783)	left IFG	-45	18	4	24.6	left Area 44
Cluster 3 (773)	left SPT	-64	-22	-4	14.5	left IPC (PF)
Cluster 4 (586)	left IFG	-46	23	29	35.8	left Area 45
					12.2	left Area 44
Cluster 5 (579)	left Area 2	-53	-31	43	31.3	left Area 2
					25.5	left IPC (PF)
					22.8	left IPC (PFt)
Cluster 6 (417)	right IFG	40	22	9	13.2	right Area 45
Cluster 7 (359)	right SPT					
Cluster 8 (264)	left SMA	0	13	48	32.2	left Area 6
					10.9	right Area 6
Cluster 9 (254)	left PrCG	-43	3	31	18.1	left Area 44
Cluster 10 (225)	right IFG	54	17	39		

Table 4 - Effect of Interest for left hemisphere regions which are activated during supramodal syllable processing, i.e. phonetic map(s)

		left PrCG		**left STG**		**left IFG (area 44)**		**left SMA**	
Condition and Session No.		***ß-est.***	***σ***	***ß-est.***	***σ***	***ß-est.***	***σ***	***ß-est.***	***σ***
READ	T1	6,87	1,58	4,96	1,46	6,84	1,38	3,39	1,61
	T2	6,40	1,32	3,71	1,22	8,74	1,16	2,02	1,35
	T3	9,52	1,30	3,84	1,20	9,46	1,14	5,28	1,32
	T4	11,31	1,40	4,68	1,29	11,14	1,23	8,32	1,43
REPEAT	T1	5,65	1,52	5,58	1,41	8,56	1,33	2,24	1,55
	T2	9,16	1,41	2,47	1,30	10,09	1,24	3,26	1,44
	T3	6,63	1,50	1,83	1,39	11,73	1,32	4,90	1,53
	T4	12,37	1,38	3,75	1,28	13,72	1,21	7,80	1,41
S_READ	T1	4,39	1,55	4,85	1,44	5,36	1,36	5,76	1,58
	T2	4,00	1,34	2,83	1,24	3,29	1,17	2,75	1,36
	T3	3,51	1,27	2,36	1,17	3,55	1,11	3,69	1,29
	T4	1,95	1,49	3,01	1,38	1,01	1,31	2,14	1,52
S_REPEAT	T1	2,72	2,06	2,01	1,91	2,41	1,81	0,75	2,10
	T2	4,16	1,55	3,88	1,44	4,22	1,36	4,32	1,58
	T3	1,46	1,53	4,60	1,42	2,82	1,35	1,76	1,56
	T4	6,16	1,52	5,24	1,40	2,99	1,33	5,03	1,55

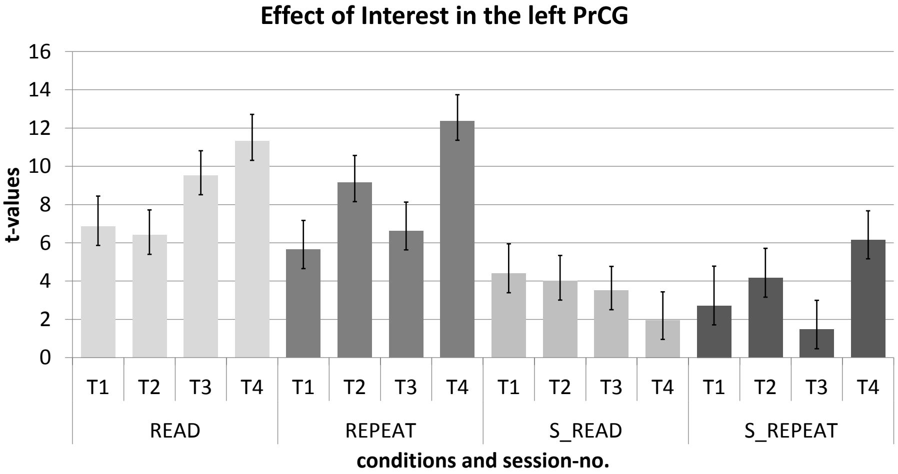

Figure 2 - Effect of Interest for each conditions at each session (T1, T2, T3, T4) of the left PrCG

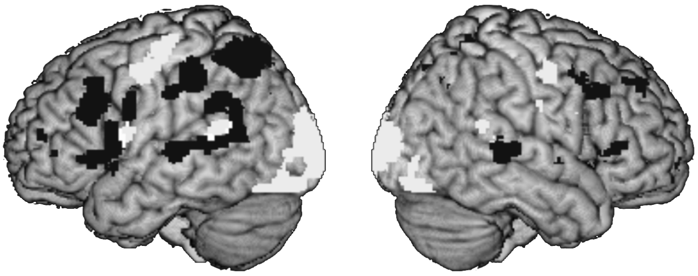

Figure 3 - Supramodal syllable processing (light, $p \leq 0.001$) might represent distributed phonetic maps and syllable priming (dark, $p \leq 0.05$) might represent access to sensorimotor state maps of a single subject.

5 Discussion

In order to support the SAR and its hypothesis of a syllable storage, i.e. a supramodal hub (phonetic map) and its connections to unimodal sensorimotor state maps (auditory, somatosensory and motor plan state map), a single subject fMRI experiment was performed. A conjunction analysis of the conditions, which were used in this study, revealed activation within bilateral PrCG, left IFG (area 44) and SMA, as well as bilateral STG. These regions are in line with the SAR and its notion of distributed supramodal hubs. That means, that there are multiple hubs simply connected to each other. These regions were partially shown in syllable preparation in previous fMRI studies of speech production as well [6, 22]. This might support our findings although previous studies did not investigate supramodal cortical activation.

The main effect of syllable priming analysis revealed activation within frontal areas, i.e. bilateral IFG (area 44/45), left PrCG, as well as left SMA, and parietal areas, i.e. left SPL, IPC as well as area 2, and temporal areas, i.e. bilateral STG. These results are compatible with the notion of a SAR, in which different regions are assumed to represent short-term memory motor as well as sensory state maps. More precisely the parietal regions might represent somatosensory aspects [16], and frontal areas represent motor aspects, which were found in speech production studies as well [6, 22]. Furthermore the temporal areas might represent auditory aspects.

In order to investigate reproducibilty of the findings the effect of interest was analyzed in left sided activated regions of supramodal syllable processing. It could be shown, that the resulted

regions are stable but variable in statistical strength over time. Overt conditions showed stronger activation than covert conditions. This might be due to the task, that thinking does need less effort on the one hand and was more boring to the subject on the other hand.
The topic of this paper will be further investigated in a bigger sized sample. Using detection tensor imaging might help to determine anatomical or structural relation of the supramodal hubs to each other and with the state maps.

References

[1] AMUNTS, K., A. MALIKOVIC, H. MOHLBERG, T. SCHORMANN and K. ZILLES: *Brodmann's areas 17 and 18 brought into stereotaxic space-where and how variable?*. Neuroimage, 11:66–84, 2000.

[2] AMUNTS, K., A. SCHLEICHER, U. BURGEL, H. MOHLBERG, H. UYLINGS and K. ZILLES: *Broca's region revisited: Cytoarchitecture and intersubject variability.* J. Comp. Neurol., 412:319–341, 1999.

[3] ASHBURNER, J., C. CHEN, G. FLANDIN, R. HENSON, S. KIEBEL, J. KILNER, V. LITVAK, R. MORAN, W. PENNY, K. STEPHAN, C. HUTTON, V. GLAUCHE, J. MATTOUT and C. PHILLIPS: *The SPM8 Manual.* Functional Imaging Laboratory: Wellcome Trust Centre for Neuroimaging, London, 2008.

[4] BULLMORE, E. and O. SPORNS: *Complex brain networks: graph theoretical analysis of structural and functional systems.* Nature Reviews Neuroscience, 10:186–198, 2009.

[5] CASPERS, S., S. GEYER, A. SCHLEICHER, H. MOHLBERG, K. AMUNTS and K. ZILLES: *The human inferior parietal cortex: cytoarchitectonic parcellation and interindividual.* Neuroimage, 33:430–448, 2006.

[6] EICKHOFF, S. B., S. HEIM, K. ZILLES and K. AMUNTS: *A systems perspective on the effective connectivity of overt speech production.* pyiosophical Transactions of the Royal Society, 376:2399–2421, 2009.

[7] FRISTON, K. J.: *Statistical parametric mapping.* In THATCHER, R. W., M. HALLET, T. ZEFFIRO, E. R. JOHN and M. HUERTA (eds.): *Functional imaging*, San Diego, 1994. Academic Press.

[8] FRISTON, K. J., P. FLETCHER, O. JOSEPHS, A. HOLMES, M. D. RUGG and R. TURNER: *Event-related fMRI: Characterizing differential responses.* Neuroimage, 7:30–40, 1998.

[9] FRISTON, K. J., A. P. HOLMES, J.-B. POLINE, B. J. GRASBY, C. R. WILLIAMS, R. S. J. FRACKOWIAK and R. TURNER: *Analysis of fMRI time-series revisited.* Neuroimage, 2:45–53, 1995a.

[10] FRISTON, K. J., A. P. HOLMES, J.-B. POLINE, B. J. GRASBY, C. R. WILLIAMS, R. S. J. FRACKOWIAK and R. TURNER: *Statistical parametric maps in functional imaging: A general linear approach.* Human Brain Mapping, 2:189–210, 1995b.

[11] GEYER, S.: *The Microstructural Border Between the Motor and the Cognitive Domain in the Human Cerebral Cortex.* Springer, Wien, 2003.

[12] GEYER, S., A. SCHLEICHER and K. ZILLES: *Areas 3a, 3b, and 1 of Human Primary Somatosensory Cortex: 1. Microstructural Organization and Interindividual Variabilty.* NeuroImage, 10:63–83, 1999.

[13] GEYER, S., A. SCHLEICHER and K. ZILLES: *Areas 3a, 3b, and 1 of Human Primary Somatosensory Cortex: 2. Spatial Normalization to Standard Anatomical Space.* NeuroImage, 11:684–696, 2000.

[14] GREFKES, C., S. GEYER, T. SCHORMANN, P. ROLAND and K. ZILLES: *Human somatosensory area 2: observer-independent cytoarchitectonic mapping, interindividual variability, and population map.* Neuroimage, 14:617–631, 2001.

[15] GUENTHER, F. H.: *Cortical interactions underlying the production of speech sounds.* Communication Disorders, 39:350–365, 2006.

[16] GUENTHER, F. H., S. S. GOSH and J. A. TOURVILLE: *Neural modeling and imaging of the cortical interactions underlying syllable production.* Brain and Lanugage, 96:280–301, 2006.

[17] JOSEPHS, O., R. TURNER and K. FRISTON: *Event-relatedfMRI.* Human Brain Mapping, 5:243–248, 1997.

[18] KRÖGER, B. J., P. BIRKHOLZ and C. NEUSCHAEFER-RUBE: *Towards an Articulation-Based Developmental Robotics Approach for Word Processing in Face-to-Face Communication.* PALADYN Journal of Behavioral Robotics, 2:82–93, 2011.

[19] KRÖGER, B. J., J. KANNAMPUZHA and C. NEUSCHAEFER-RUBE: *Towards a neurocomputational model of speech production and perception.* Speech Communication, 51:793–809, 2009.

[20] LEVELT, W. J. M.: *Speaking: From intention to articulation.* MIT Press, Cambridge, MA, 1992.

[21] LEVELT.W.J.M. and L. WHEELDON: *Do speakers have access to a mental syllabary?.* Cognition, 50:239–269, 1994.

[22] RIECKER, A., J. KASSUBEK, K. GRÖSCHEL, W. GRODD and H. ACKERMANN: *fMRI reveals two distinct cerebral networks subserving speech motor control.* Neurology, 64:700–706, 2005.

[23] SCHEPERJANS, F., S. EICKHOFF, H. L., H. MOHLBERG, K. HERMANN, K. AMUNTS and K. ZILLES: *Probalistic Maps, Morphometry, and Variability of Cytoarchitectonic Areas in the Human Superior Parietal Cortex.* Cerebral Cortex, 18:2141–2157, 2008.

[24] SCHEPERJANS, F., K. HERMANN, S. EICKHOFF and K. AMUNTS: *Observer-Independent Cytoarchitectonic Mapping of the Human Superior Parietal Cortex.* Cerebral Cortex, 18:846–867, 2008.

[25] WORSLEY, K. J. and K. J. FRISTON: *Analysis of fMRI time-series revisited–again.* Neuroimage, 2:359–365, 1995.