NEURAL CORRELATES OF INNER SPEAKING, IMITATING AND HEARING: AN FMRI STUDY

Lœvenbruck H.[1], Grandchamp R.[1], Rapin L.[1], Perrone-Bertolotti M.[1], Pichat C.[1], Haldin, C.[1], Cousin E. [1], Lachaux J.P.[2], Dohen M.[3], Perrier P.[3], Garnier M.[3], Baciu M.[1]

[1]Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, Grenoble, France, [2]INSERM U1028-CNRS UMR5292, Brain Dynamics and Cognition Team, Lyon Neuroscience Research Center, Lyon-Bron, France, [3]Univ. Grenoble Alpes, Grenoble INP, CNRS, GIPSA-lab, Grenoble, France

{helene.loevenbruck/romain.grandchamp/marcela.perrone-bertolotti/cedric.pichat/celise.haldin/monica.baciu}@univ-grenoble-alpes.fr, jp.lachaux@inserm.fr, {marion.dohen/pascal.perrier/maeva.garnier}@gipsa-lab.grenoble-inp.fr

ABSTRACT

The neural correlates of inner language are still ill-defined. Several varieties need further examination: deliberate monologal or dialogical inner speech and unbidden inner language during mind wandering. Using fMRI, we probed varieties of inner speech along dialogicality and intentionality in 24 healthy participants, during five tasks: speech perception, monological self voice inner speech, monological other voice inner speech, dialogical other voice inner speech, and mind wandering with verbal episodes. Results are interpreted within a predictive control account in which intentional inner speech is viewed as deriving from multisensory goals, transformed into motor commands. Efference copies of the inhibited motor commands are converted into predicted percepts (inner voice). Based on the comparison between the five tasks we report the neural correlates of dialogicality and intentionality in inner speech.

Keywords: Neurolinguistics, Inner production, Perception, Mind wandering, Predictive Control

1. INTRODUCTION

Inner speech can be defined as “the subjective experience of language in the absence of overt and audible articulation” [1]. It plays a critical role in reasoning, executive function, memory and self-awareness [7] and has been shown to vary along several axes. On the dialogicality axis, it can be monological or dialogical, reflecting the experience of communication [2]. We can use our own voice, we can covertly imitate someone speaking or we can imagine hearing someone. On the intentionality axis, we sometimes deliberately engage in inner speech. Other times, we find ourselves unwilfully using inner language, what is called verbal mind wandering [9].

The neural structures that mediate this manifold activity are still ill-defined (for reviews, see [4,5,9]). Inner speech has been viewed as the mental simulation of overt speech, which implies that speech production processes are at play, although production is inhibited [8]. Inner speech is associated with the perception of voices which suggests that perception processes could also be involved. The shifts in perspective (from self-speaking to other-imagining) which occur in inner dialogue deserve further description [2]. Finally, the relations between wilful inner speech and verbal mind wandering need deeper examination [6]. We report the results of an fMRI study in which we examined varieties of inner speech along the dialogicality and intentionality axes and compared them with speech perception.

2. METHODS

2.1. Participants

Twenty-four healthy right-handed (10 men; mean age = 29.5 years, SD =10.04; 14 women, mean age = 28.07 years, SD = 8.14) native speakers of French were included. Each participant gave informed written consent and received 30 €. The study was approved by the local ethics committee (38RC14.304, ID-RCB: 2014-A01403-44).

2.2. Tasks

Participants were first introduced to an avatar with a saliently high-pitched voice, who gave them instructions and provided thorough training to make sure they produced inner speech without mouthing. In the first four tasks, each trial started with the visual presentation of a word and a picture of the corresponding object. For example, the word “ball”, with a picture of a ball framed within a clock was presented for 2s, after which the clock rotated and the participant performed the task, which lasted for 4s. Each trial was repeated several times in each condition (see 2.3). In the fifth task, one word-image pair was presented for 2s, then a clock was displayed for 30s. **In the Inner Speech Self (MS) condition**, participants had to mentally generate definitions of the objects. **In the Inner Speech Other (MO)**

condition, participants had to mentally generate definitions, imitating the voice of the avatar. **In the Imagined Speech (DO) condition**, participants had to imagine that the avatar was addressing them, starting with "Here is a typical image of a…" and simply naming the objects without defining them (to reduce cognitive load). **In the "Speech Perception" (SP) condition**, participants had to listen to definitions played in the earphones. **In the "Verbal Mind Wandering" (VMW) condition**, after the initial 2s word-image pair, participants were asked to fixate a stylized clock rotating for 30s. They were instructed to monitor spontaneously occurring thoughts. At the end of the trial, they reported the periods with verbal thoughts, by selecting time portions on the clock, using a joystick. Stimulus presentation and response collection were controlled with Presentation (http://www.neurobs.com).

2.3. fMRI protocol

A block design paradigm was used. In the MS, MO, DO and SP conditions, trials were separated by a fixation cross displayed for 2s. At the beginning of each block, an instruction screen was displayed for 6s while the instructions were played in the earphones. Each block was composed of 5 trials of the same condition. A fixation cross was displayed for 8s before and after each block. When a participant was doing the task for the first time in the run, the block started with three training trials. The sequence of conditions was pseudo-random across participants, with DO after MO to reduce confusion. Each participant completed 3 sequences of blocks in each fMRI run. Two runs were recorded. This resulted in 30 test trials (6 blocks of 5 trials) plus 3 training trials per condition per participant. For each participant, the sequence of conditions was the same in the two runs.

2.4. Pre- and Post-experiment questionnaires

One day before the experiment, participants filled in the Edinburgh Handedness Inventory and a mental imagery questionnaire. After the experiment, they filled in a recall questionnaire with a list of 60 words, for which they checked whether they had generated a definition in the scanner (20 words were distractors). This aimed at testing their attention during the tasks. Participants also filled in subjective questionnaires to report how well they performed the tasks and to describe their thought contents during VMW.

2.5. fMRI acquisition

Experiments were performed using a whole-body 3T scanner (Philips Medical Systems, Eugene, OR) with a 32-channel head coil at the University Hospital in Grenoble. The manufacturer-provided gradient-echo/T2*weighted EPI method was used. Forty-two adjacent axial slices parallel to the bi-commissural plane were acquired in non-interleaved mode. Slice thickness was 3 mm. The in-plane voxel size was 3×3 mm (240×240 mm field of view with a 80×80 pixel data matrix). The main sequence parameters were: TR = 2.5s, TE = 30ms, flip angle= 82°. Two fMRI runs were acquired. During the break between the two runs, a T1-weighted high-resolution 3D anatomical volume was acquired, with a 3D T1 TFE sequence (field of view = 256×224×175mm; resolution: .89x .89x1.37mm; acquisition matrix: 192×137×128 pixels; reconstruction matrix: 288×288×128 pixels). Subjects' gazes were monitored with an eyetracker.

2.6. fMRI data analysis

Data analysis was performed using SPM12 software (Wellcome Department of Imaging Neuroscience, UK). Individual scans were time-corrected, realigned, normalized to the MNI space and spatially smoothed by an 8-mm FWHM Gaussian kernel. Times-series for each voxel were high-pass filtered (1/512 Hz) to remove low-frequency noise and signal drift. The fMRI signal was analysed using single-participant general linear model. For each participant, five conditions of interest (MS, MO, DO, SP, VMW) were modelled as regressors. Movement parameters derived from realignment corrections were included as factors of no interest. The run number was added as an additional factor. For the first-level analysis, five contrasts corresponding to each regressor of interest vs. implicit baseline were computed. For the second level, (i) one-sample T-tests were performed to obtain simple effects of conditions, (ii) conjunction analyses were carried out between each inner speech condition and SP and (iii) one-way within-subject ANOVA was performed to obtain differential effects between conditions. In all analyses (except for the contrast between MS and MO), significant voxel clusters on individual *t*-maps were identified with Family Wise Error (FWE) correction at $p < .05$. Location of cluster maxima was determined using Automated Anatomical Labeling (AAL) map [11].

3. RESULTS

Mean accuracy score in the word recall task for all participants reached 84.42% ± 16.63. This high score together with the eyetracker monitoring suggest that participants were focused on the tasks. The minimal head movement registered in the scanner confirms that they did not articulate. The subjective questionnaires indicated that the VMW condition contained verbal episodes.

FMRI contrasts between each condition and the baseline are presented in Table 1, all p<.05, FWE correction.

To assess whether production-perception processes were recruited in all varieties of wilful inner speech, MS, MO and DO were confronted to SP. Activations in MS and SP are shown separately (Fig. 1a-b) and for MS, MO and DO in conjunction with SP (Fig. 1c).

To study the varieties of wilful inner speech along the dialogicality axis, MS was compared with MO (changing voice, Fig. 2a-b) and MO was compared with DO (shifting from monologue to dialogue, Fig. 2c-d).

Finally, to explore the intentionality axis, activations in the VMW condition, which according to participants contained verbal episodes, were plotted in Fig. 3 (to be compared with MS in Fig. 1a).

Table 1: Peak activated clusters in each condition vs. baseline (in bold) and their sub-clusters

Condition				MNI Coordinates		
	Region Label (aal)	Extent	t-value	x	y	z
MS vs BL	**Frontal_Inf_Oper_L**	**2113**	**14.12**	**-48**	**11**	**5**
	Frontal_Inf_Tri_L		14.08	-36	26	-1
	Putamen_L		11.66	-18	11	-1
	Frontal_Sup_Medial_L	**771**	**12.84**	**-3**	**26**	**41**
	Supp_Motor_Area_L		11.33	-6	17	62
	Supp_Motor_Area_R		10.13	6	8	62
	Occipital_Mid_R	**116**	**9.97**	**36**	**-82**	**14**
	Occipital_Sup_R		8.07	18	-94	20
	Frontal_Mid_2_L	**37**	**7.75**	**-30**	**53**	**14**
	Temporal_Mid_L	**38**	**7.73**	**-51**	**-40**	**2**
	Frontal_Sup_2_L	**25**	**7.22**	**-9**	**53**	**35**
	Frontal_Sup_Medial_L		6.26	-9	44	41
	Occipital_Mid_L	**50**	**6.89**	**-39**	**-67**	**-1**
	Temporal_Inf_L		6.45	-45	-52	-16
	Calcarine_L	**21**	**6.81**	**0**	**-82**	**-4**
	Precentral_R	**5**	**6.79**	**54**	**2**	**44**
	SupraMarginal_L	**2**	**6.58**	**-45**	**-43**	**32**
	Cerebellum_6_R	**14**	**6.50**	**36**	**-64**	**-25**
	Fusiform_L	**5**	**6.34**	**-30**	**-46**	**-19**
	Temporal_Pole_Sup_R	**4**	**6.12**	**54**	**14**	**-4**
	Hippocampus_L	**1**	**6.00**	**-18**	**-40**	**14**
	Insula_R	**1**	**6.00**	**39**	**17**	**2**
MO vs BL	**Supp_Motor_Area_L**	**661**	**11.52**	**-9**	**17**	**47**
	Supp_Motor_Area_L		10.28	-9	5	62
	Frontal_Inf_Orb_2_L	**717**	**11.51**	**-45**	**20**	**-7**
	Frontal_Inf_Oper_L		11.12	-51	11	5
	Occipital_Mid_R	**29**	**9.74**	**30**	**-85**	**17**
	Putamen_L	**93**	**8.54**	**-18**	**11**	**2**
	Precentral_L	**78**	**8.49**	**-48**	**-4**	**50**
	Hippocampus_L	**25**	**7.84**	**-15**	**-16**	**-19**
	Precentral_R	**12**	**7.67**	**54**	**-1**	**44**
	Frontal_Mid_2_L	**11**	**7.25**	**-30**	**50**	**11**
	Insula_R	**77**	**7.20**	**36**	**17**	**2**
	Putamen_R	**16**	**7.07**	**24**	**5**	**2**
	Caudate_R	**3**	**6.90**	**18**	**23**	**5**
	Temporal_Inf_R	**9**	**6.83**	**48**	**-67**	**-28**
	Cerebellum_6_R	**6**	**6.52**	**36**	**-58**	**-28**
	Precentral_R	**1**	**6.00**	**63**	**8**	**17**
DO vs BL	**Occipital_Mid_R**	**230**	**10.86**	**33**	**-82**	**11**
	Cuneus_R		9	15	-94	20
	Temporal_Mid_R		8.71	48	-70	2
	Supp_Motor_Area_L	**503**	**10.54**	**0**	**11**	**59**
	Supp_Motor_Area_L		10.52	-6	2	65
	Frontal_Inf_Tri_L	**432**	**10**	**-42**	**32**	**20**
	Frontal_Inf_Oper_L		9.84	-51	11	2
	Frontal_Inf_Orb_2_L		9.78	-42	20	-7
	Precentral_L	**64**	**8.51**	**-48**	**-7**	**47**
	Precentral_R	**29**	**8.31**	**54**	**2**	**44**
	Insula_L	**36**	**7.53**	**48**	**8**	**-1**
	Lingual_L	**13**	**7.16**	**0**	**-79**	**-7**
	Postcentral_L	**14**	**7.01**	**-60**	**2**	**20**
	Occipital_Mid_L	**18**	**6.76**	**-39**	**-70**	**2**
	Rolandic_Oper_R	**4**	**6.54**	**60**	**8**	**14**
	Frontal_Mid_2_L	**1**	**6.25**	**-36**	**50**	**23**
	Occipital_Sup_L	**1**	**5.99**	**-9**	**-97**	**8**
SP vs BL	**Temporal_Sup_L**	**784**	**15.27**	**-63**	**-22**	**5**
	Temporal_Sup_L		11.03	-45	-22	5
	Temporal_Sup_R	**491**	**13.32**	**63**	**-10**	**-1**
	Temporal_Sup_R		12.90	63	-28	8
	Frontal_Inf_Tri_L	**345**	**8.85**	**-51**	**35**	**14**
	Frontal_Inf_Orb_2_L		8.40	-45	23	-7
	Occipital_Mid_R	**12**	**8.14**	**39**	**-82**	**14**
	Precentral_L	**27**	**7.69**	**-51**	**-7**	**47**
	Temporal_Inf_R	**17**	**7.51**	**45**	**-61**	**-7**
	Supp_Motor_Area_L	**25**	**7.26**	**-9**	**8**	**62**
	Lingual_L	**12**	**7.25**	**0**	**-79**	**-4**
	Frontal_Sup_2_L	**29**	**7.22**	**-12**	**29**	**50**
	Supp_Motor_Area_L		6.43	-6	17	56
	Precentral_R	**4**	**6.91**	**54**	**2**	**44**
	Temporal_Inf_L	**9**	**6.80**	**-45**	**-43**	**-13**
	Hippocampus_L	**20**	**6.74**	**-21**	**-16**	**-19**
	Frontal_Sup_Medial_L	**5**	**6.58**	**-9**	**47**	**41**
	Fusiform_L	**2**	**6.13**	**-33**	**-46**	**-19**
	Precentral_L	**1**	**5.99**	**-42**	**2**	**53**
VMW vs BL	**Parietal_Sup_R**	**161**	**10.91**	**21**	**-58**	**56**
	Frontal_Sup_Medial_L	**305**	**10.90**	**-6**	**29**	**35**
	Supp_Motor_Area_L		7.41	-9	14	56
	Frontal_Sup_2_L		6.89	-18	17	65
	Frontal_Mid_2_L	**186**	**10.13**	**-30**	**50**	**14**
	Frontal_Sup_2_L		7.97	-24	44	35
	Parietal_Inf_R	**97**	**8.86**	**42**	**-37**	**47**
	Temporal_Inf_R	**107**	**8.82**	**51**	**-64**	**-4**
	Occipital_Mid_R		8.58	36	-82	17
	Parietal_Sup_L	**37**	**8.29**	**-18**	**-67**	**59**
	Parietal_Inf_L	**100**	**8.20**	**-51**	**-55**	**41**
	Frontal_Inf_Oper_R	**111**	**7.97**	**57**	**17**	**5**
	Insula_R		7.80	36	14	-1
	Frontal_Inf_Orb_2_R		6.26	48	20	-7
	Frontal_Mid_2_R	**66**	**7.90**	**30**	**50**	**26**
	Supp_Motor_Area_R	**20**	**7.26**	**15**	**20**	**62**
	Frontal_Inf_Orb_2_L	**148**	**7.20**	**-42**	**17**	**-7**
	Insula_L		7.17	-33	17	2
	Occipital_Mid_L	**4**	**6.82**	**-36**	**-73**	**5**
	Cerebellum_Crus1_L	**2**	**6.40**	**-33**	**-58**	**-34**
	Frontal_Sup_2_R	**1**	**6.10**	**24**	**14**	**65**
	Frontal_Inf_Tri_R	**1**	**6.07**	**48**	**35**	**-1**
	Frontal_Sup_2_R	**1**	**5.97**	**27**	**47**	**11**
	Frontal_Mid_2_R	**2**	**5.97**	**33**	**50**	**14**

Figure 1: Tasks vs baseline (a) SP (b) MS (c) Conjunctions MS & SP, MO & SP, DO & SP.

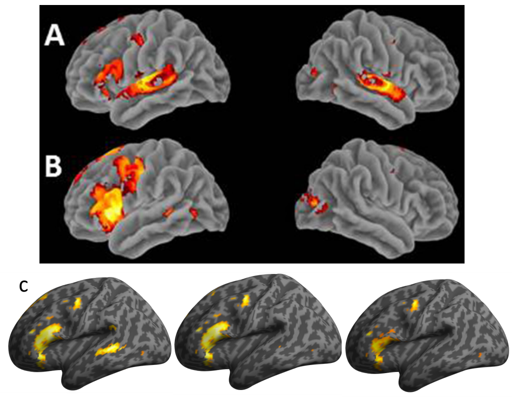


Figure 2: (a) MS vs MO, (b) MO vs MS, (c) MO vs DO, (d) DO vs MO.

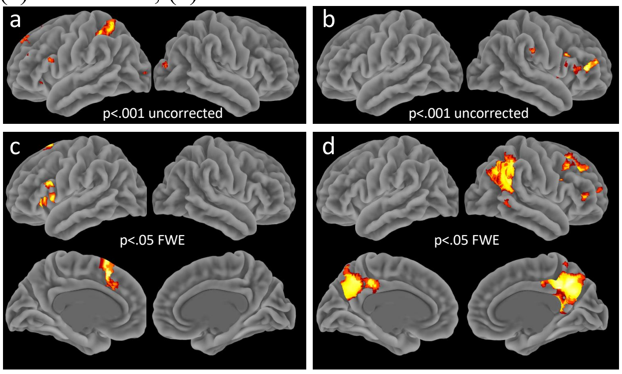
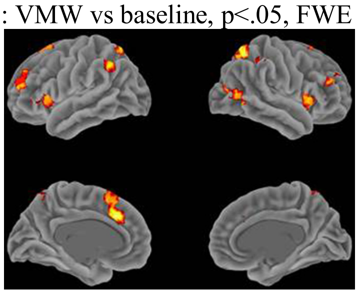


Figure 3: VMW vs baseline, p<.05, FWE



4. DISCUSSION

Occipital activation in all conditions was related to the visual processing at the beginning of each trial. In line with previous works [9], compared with baseline, SP recruited bilateral superior temporal gyri (STG), left supramarginal gyrus (SMG), left inferior frontal gyrus (IFG), left superior frontal gyrus (SFG), bilateral premotor (PM) cortex, left supplementary Motor Area (SMA), left motor cortex, left hippocampus (Fig. 1a) and MS revealed greater left hemisphere activation in the IFG, middle frontal gyrus (MFG), SFG, SMG, posterior middle temporal gyrus (MTG), hippocampus, together with bilateral SMA, bilateral PM cortex, and right cerebellum (Fig. 1b). The MS/SP conjunction confirmed that left IFG, SFG, MTG, SMA, SMG, hippocampus, bilateral PM cortex, and occipital posterior MTG were recruited by both conditions. This supports the claim that wilful inner speech involves the inhibited production of motor commands, generated in frontal regions [8]. Efference copies of the commands would be processed by cerebellar internal models, giving rise to a sensory prediction, the inner voice, in STG/MTG.

The conjunctions between SP and the other two inner speaking conditions (MO and DO) revealed a weaker temporal activation (Fig. 1c). In MO and DO, internal models are less expert than in MS, and presumably generate more precarious auditory predictions, which could explain the lesser auditory cortex activation, compatible with the participant's subjective experience of a fainter voice percept.

Along the dialogicality axis, covertly using someone else's voice (MO) vs one's own voice (MS) resulted in a lateralisation shift. Greater left IFG and parietal activation was observed in MS vs MO and greater right IFG and parietal activation in MO vs MS (Fig. 2a-b, uncorrected). Perspective shifting, from self-speaking with other's voice (MO) to other-imagining (DO), yielded greater activation in left IFG and SMA in MO vs DO and greater right IFG and medial FG, right superior and inferior parietal lobules and bilateral precuneus in DO vs MO (Fig. 2c-d). Parietal and precuneus activations are consistent with previous studies on perspective taking [2,3].

Along the intentionality axis, VMW was mainly associated with activations in bilateral IFG, medial prefrontal cortex, superior and inferior parietal gyri, precuneus, occipital regions, and left caudate, thalamus and cerebellum (Fig. 3), consistent with studies on mind wandering and the default mode [10].

5. CONCLUSION

This fMRI study examined varieties of inner speech along dialogicality and intentionality axes.

It was found that inner speech engages speech production processes in left inferior frontal and premotor regions, as well as perception processes in temporal regions, supporting a predictive control model of inner speech [8].

Along the dialogicality axis, covertly using someone else's voice instead of one's own resulted in a shift of activations to the right hemisphere. Changing perspective, from self-speaking to other-imagining, yielded activations in precuneus and parietal lobules.

Finally, along the intentionality axis, unwilful inner speech was associated with less activation in temporal regions than wilful inner speech, presumably reflecting the subjective evanescence of VMW. Further analyses are in progress to better describe this last condition.

ACKNOWLEDGEMENT

This research was funded by the INNERSPEECH project ANR-13-BSH2-0003-01. The IRMaGe MRI facility was partly funded by ANR-11-INBS-0006. We are grateful to L. Fadiga, Y. Klein, L. Lamalle, I. Troprès, A. Vilain, T. Woodward for suggestions. We thank F. Gautheron and A. Steinhilber for their help.

REFERENCES

[1] Alderson-Day, B., & Fernyhough, C. (2015). Inner Speech: Development, Cognitive Functions, Phenomenology, and Neurobiology. *Psych. Bulletin*, *141*(5), 931–965.

[2] Alderson-Day, B., Weis, S., McCarthy-Jones, S., Moseley, P., Smailes, D., & Fernyhough, C. (2016). The Brain's Conversation with Itself: Neural Substrates of Dialogic Inner Speech. *Social Cog. and Aff. Neurosc.*

[3] Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. *Nature Reviews Neuroscience*, *13*(8), 556–571.

[4] Christoffels, I. K., Formisano, E., & Schiller, N. O. (2007). Neural correlates of verbal feedback processing: An fMRI study employing overt speech. *Human Brain Mapping,* 28(9), 868–879.

[5] Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J. C., & Warburton, E. A. (n.d.). The neural correlates of inner speech defined by voxel-based lesion–symptom mapping. Brain, 134(10), 3071–3082.

[6] Hurlburt, R. T., Alderson-Day, B., Kühn, S., & Fernyhough, C. (2016). Exploring the Ecological Validity of Thinking on Demand: Neural Correlates of Elicited vs. Spontaneously Occurring Inner Speech. *PLoS ONE*, *11*(2), e0147932

[7] Langland-Hassan, P., & Vicente, A. (2018). *Inner Speech: New Voices.* Oxford University Press, UK.

[8] Lœvenbruck, H., Grandchamp, R., Lucile, R., Nalborczyk, L., Dohen, M., Perrier, P., … Perrone-Bertolotti, M. (2018). A cognitive neuroscience view of inner language: to predict and to hear, see, feel. In *Inner Speech: New Voices* (pp. 131–167). Oxford Univ. P.

[9] Perrone-Bertolotti, M., Rapin, L., Lachaux, J.-P., Baciu, M., & Loevenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research*, *261*(0), 220–239.

[10] Raichle, M. E. (2015). The brain's default mode network. *Annual Review of Neuroscience*, (38), 433-47.

[11] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*(1), 273–289.